



## Evaluating Driver Education Programs



*Prepared by*

Kathryn M. Clinton  
Larry Lonero

*Northport Associates*

182 Bagot Street  
Cobourg, Ontario  
Canada  
K9A 3G2  
905-377-8883  
[www.northportassociates.com](http://www.northportassociates.com)

*Prepared for*



607 14th Street, NW  
Suite 201  
Washington, DC 20005  
800-993-7222  
[www.aaafoundation.org](http://www.aaafoundation.org)

*October 2006*

**Comprehensive Guidelines**

---

This report was funded by the AAA Foundation for Traffic Safety in Washington, D.C. Founded in 1947, the AAA Foundation is a not-for-profit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, from individual members, from AAA-affiliated insurance companies, as well as from other organizations or sources.

This publication is distributed by the AAA Foundation for Traffic Safety at no charge, as a public service. It may not be resold or used for commercial purposes without the explicit permission of the Foundation. It may, however, be copied in whole or in part and distributed for free via any medium, provided the AAA Foundation is given appropriate credit as the source of the material.

The opinions, findings, conclusions, and recommendations expressed in this publication are those of the authors and are not necessarily those of the AAA Foundation for Traffic Safety, any individuals who peer-reviewed this report, or advisory group members. The AAA Foundation for Traffic Safety assumes no liability for the use or misuse of any information, opinions, findings, conclusions, or recommendations contained in this report.

If trade or manufacturer's names are mentioned, it is only because they are considered essential to the object of this report and their mention should not be construed as an endorsement. The AAA Foundation for Traffic Safety does not endorse products or manufacturers.

©2006, AAA Foundation for Traffic Safety



# Contents

**Glossary of Acronyms** > 5

**List of Tables and Figures** > 7

**Acknowledgements** > 9

**Preface** > 11

**Executive Summary** > 13

**1. Introduction** > 23

- Overview of the *Guidelines*. . . . . 24
- How to Read the *Guidelines*. . . . . 25
- Target Audiences. . . . . 25
- Driver Education Programs in the U.S. and Canada. . . . . 26
- Approaches to Driver Education Program Evaluation. . . . . 29

**2. Evaluation Model and Framework** > 31

- What is Program Evaluation? . . . . . 32
- Why is Evaluation Important? Lessons from Related Fields . . . . . 33
- What is Good Evaluation? . . . . . 35
- Depicting Programs through Logic Models . . . . . 40
- Developing an Evaluation Model and Framework for Driver Education . . . . . 43
- Driver Education Evaluation Model . . . . . 43
- Integrated Driver Education Evaluation Framework . . . . . 45

**3. Driver Education Evaluation Guidelines** > 55

- Applying the Framework—How to Use Evaluation Effectively. . . . . 55
- The Evaluation Team . . . . . 55
- Guiding Questions for Driver Education Program Evaluation . . . . . 56
- Steps for Evaluating Driver Education Programs . . . . . 57
- STEP 1:** Focus the Evaluation. . . . . 59
- STEP 2:** Select Evaluation Methods . . . . . 77
- STEP 3:** Develop Data Collection Plan and Select Data Collection Tools. . . . . 101
- STEP 4:** Gather, Analyze, and Summarize Data . . . . . 119
- STEP 5:** Interpret and Act Upon the Evaluation Findings. . . . . 129

**4. Basic Driver Education Program Evaluation Scenario** > 137

- Program Logic Model . . . . . 138
- Evaluation Decisions . . . . . 140
- Evaluation Design . . . . . 145
- Evaluation Quality. . . . . 149

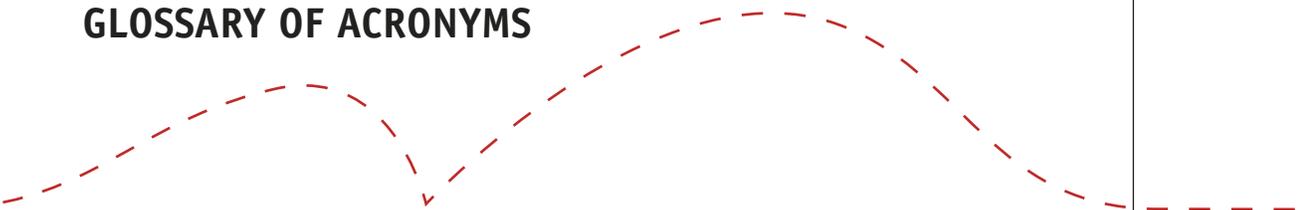
(continued)

*(Table of Contents—continued)*

<b>5. Advanced Driver Education Program Evaluation Scenario</b>	<b>➤ 151</b>
Program Logic Model . . . . .	153
Evaluation Decisions . . . . .	153
Evaluation Design . . . . .	157
Evaluation Quality . . . . .	164
<b>6. Conclusion</b>	<b>➤ 165</b>
<b>References and Evaluation Resources</b>	<b>➤ 167</b>
References . . . . .	167
Evaluation Resources . . . . .	173
<b>APPENDIX A: Review of the Literature: Driver Education Evaluation—Past and Present</b>	<b>. . . . .179</b>
<b>APPENDIX B: Glossary of Terms</b>	<b>. . . . .213</b>
<b>APPENDIX C: Program Evaluation Standards Definitions</b>	<b>. . . . .231</b>
<b>APPENDIX D: Overview of Evaluation Steps</b>	<b>. . . . .235</b>
<b>APPENDIX E: Worksheets</b>	<b>. . . . .239</b>
<b>APPENDIX F: Data Collection Tools: Basic Evaluation Scenario</b>	<b>. . . . .251</b>
<b>APPENDIX G: Data Collection Tools: Advanced Evaluation Scenario</b>	<b>. . . . .281</b>
<b>APPENDIX H: Attitude and Risk-Taking Questionnaires</b>	<b>. . . . .313</b>
<b>APPENDIX I: Benchmark Program Standards</b>	<b>. . . . .317</b>
<b>APPENDIX J: Hiring an Evaluator</b>	<b>. . . . .329</b>



## GLOSSARY OF ACRONYMS



<b>AAA</b>	Formerly American Automobile Association
<b>AAAFTS</b>	AAA Foundation for Traffic Safety
<b>ADTSEA</b>	American Driver and Traffic Safety Education Association
<b>CAA</b>	Canadian Automobile Association
<b>CIPP</b>	Context, Input, Process, and Product Evaluation
<b>CPM</b>	Critical Path Method
<b>DSAA</b>	Driving School Association of the Americas
<b>GDL</b>	Graduated Driver Licensing Program
<b>GLP</b>	Graduated Licensing Program
<b>MADD</b>	Mothers Against Drunk Driving
<b>MPI</b>	Manitoba Public Insurance Corporation
<b>NHTSA</b>	National Highway Traffic Safety Administration
<b>NIDB</b>	National Institute for Driver Behavior
<b>NIMH</b>	National Institute for Mental Health
<b>NTSB</b>	National Transportation Safety Board
<b>OECD</b>	Organisation for Economic Cooperation and Development
<b>PERT</b>	Program Evaluation and Review Technique
<b>R&amp;D</b>	Research and Development
<b>RCT</b>	Randomized controlled trial
<b>RSEA</b>	Road Safety Educators' Association
<b>SADD</b>	Students Against Drunk Driving
<b>SES</b>	Socioeconomic status
<b>TIRF</b>	Traffic Injury Research Foundation
<b>TRB</b>	Transportation Research Board
<b>TTI</b>	Texas Transportation Institute
<b>UMTRI</b>	University of Michigan Transportation Research Institute
<b>UNC-HSRC</b>	University of North Carolina Highway Safety Research Center



# LIST OF TABLES AND FIGURES

## TABLES

<b>1</b>	Program Evaluation Standards	39
<b>2</b>	Evaluation Types and Applicability to Program Areas	47
<b>3</b>	Examples of Evaluation Targets and Success Indicators	48
<b>4</b>	Integrated Driver Education Evaluation Framework	51
<b>5</b>	Organizing Program Information	63
<b>6</b>	Example of a Driver Education Program Logic Model	67
<b>7</b>	Sample Evaluation Questions Checklist	71
<b>8</b>	Driver Education Evaluation Targets	73
<b>9</b>	Suggested Activities for Evaluation Levels	83
<b>10</b>	Comparing Qualitative and Quantitative Research	84
<b>11</b>	Examples of Qualitative and Quantitative Methods	85
<b>12</b>	Research Methods Checklist	86
<b>13</b>	Data Comparison Methods for Driver Education Evaluations	88
<b>14</b>	Ethics and Protection of Human Rights Checklist	98
<b>15</b>	Data Types	102
<b>16</b>	Definitions of Data Generation and Data Gathering Methods	103
<b>17</b>	Identifying Data Sources for Driver Education Evaluations	104
<b>18</b>	Examples of Indicators for Program Success	108
<b>19</b>	Data Collection Tools Checklist	112

<b>20</b>	Questions to Help Assess the Quality of Data Collection Tools	116
<b>21</b>	Data Collection Logistics Plan	121
<b>22</b>	Examples of Data Analysis Procedures for Qualitative and Quantitative Data	125
<b>23</b>	Evaluation Decisions for a Medium-Size Driver Education Program	140
<b>24</b>	Sample Evaluation Design for a Medium-Size Driver Education Program	146
<b>25</b>	Evaluation Decisions for a Large Driver Education Program	154
<b>26</b>	Sample Evaluation Design for a Large Driver Education Program	158

## FIGURES

<b>1</b>	Components of Good Evaluation Theory	36
<b>2</b>	Planning and Evaluation Cycle	41
<b>3</b>	Generic Program Logic Model	42
<b>4</b>	Driver Education Evaluation Hierarchy	43
<b>5</b>	Driver Education Evaluation Model	44
<b>6</b>	Types of Stakeholders	61
<b>7</b>	Hypothetical Program Logic Model for a Medium-Size Driver Education Program	139
<b>8</b>	Hypothetical Program Logic Model for a Large Driver Education Program	152



## ACKNOWLEDGEMENTS

*Evaluating Driver Education Programs: Comprehensive Guidelines* was developed with input from a small group of experts who served as project team members. Included were Mr. Dan Mayhew and Dr. Douglas Beirness from the Traffic Injury Research Foundation, Mr. Val Pezoldt from the Transportation Research Center of Texas Transportation Institute, Dr. Douglas Black of Northport Associates, Dr. Erik Olsen from the National Institute of Child Health and Human Development, and Mr. John Brock of General Dynamics Information Technology. The *Guidelines* greatly benefited from the astute insight and thoughtful comment provided on project direction and drafts of the *Guidelines* by these team members.

The *Guidelines* also benefited from the assistance of an active Advisory Group comprised of driver education and evaluation experts from across North America. A 2-day consultative workshop was attended by 18 advisory group members. Critical aspects of evaluation were brainstormed and debated, and consensus was sought on key issues including guideline users, content, methodology, and format. An additional group of international driver education experts was involved through an online discussion board, which provided an avenue for ongoing dialogue about the evaluation of driver education programs throughout the project and also an opportunity to review and comment on the literature review and discussion paper prepared and distributed as part of the *Guidelines* consultation.

The authors would like to acknowledge the considerable contribution of the Advisory Group members, who provided expert input into developing these *Guidelines*, and many of whom provided insightful comments on a draft of the *Guidelines*. This input was invaluable in conceptualizing and organizing the *Guidelines* and the two companion documents.

The Advisory Group membership included:

Mr. Walter J. Barta, Alberta Motor Association

Dr. Raymond Bingham, University of Michigan Transportation Research Center

Mr. Eric Chapman, California Department of Motor Vehicles

Professor Mary Chipman, University of Toronto

Dr. Richard Compton, National Highway Traffic Safety Administration

Dr. Susan Ferguson, Insurance Institute for Highway Safety

Mr. John Harvey, Association of State Supervisors of Safety and Driver Education,  
and Oregon Department of Transportation

Mr. Andrew Krajewski, Maryland Motor Vehicle Administration  
Mr. Scott Masten, University of North Carolina Highway Safety Research Center  
Dr. James McKnight, Transportation Research Associates  
Dr. James Nichols, Nichols and Associates  
Dr. Judith Ottoson, Georgia State University  
Mr. Ray Peck, R.C. Peck & Associates  
Ms. Debbie Prudhomme, Driving School Association of the Americas  
Dr. Allen Robinson, American Driver and Traffic Safety Education Association  
Dr. Michael Scriven, Western Michigan University  
Dr. Bruce Simons-Morton, National Institute of Child Health and Human Development  
Dr. William Van Tassel, AAA National Headquarters

The detailed review provided by Dr. James Hedlund is also acknowledged. Lastly, the authors would like to extend their appreciation to the AAAFTS and BMW of North America who supported this project and to the Foundation staff who provided insights and guidance during the development of the *Guidelines* and companion documents.

## PREFACE

*Evaluating Driver Education Programs: Comprehensive Guidelines*, sponsored by the AAA Foundation for Traffic Safety and BMW of North America, provides a detailed background for planning and conducting effective evaluation of beginner driver education, and for integrating evaluation into program development and policy. The *Guidelines* cover a range of evaluations from simple to complex, and are written primarily for program evaluators, researchers, and other technical audiences. Actual tools, such as surveys, focus group guides, and log books that can be used or adapted for evaluating beginner driver education programs are included.

There are two companion documents. *Evaluating Driver Education Programs: Management Overview* provides a general introduction to evaluation and research methods relevant to the whole range of driver education evaluation, including safety impacts. The *Management Overview* is intended for driving school owners, driver educators, program managers, administrators, and others with limited background in research methods. It provides a general introduction to the art and science of program evaluation, with a specific focus on how program evaluation concepts and methods can be applied to driver education evaluation.

*Evaluating Driver Education Programs: How-To Guide* is a hands-on manual about the formative types of evaluation that program developers and managers can use to improve the quality of their products and processes. Formative evaluation helps develop, improve, or “form” a program by assessing its content and products, and its processes and delivery. The *How-to Guide* provides step-by-step guidance for actually conducting a basic evaluation aimed at improving a beginner driver education program. The *Guide* was developed especially for driving school operators and owners, program developers, and managers.

The *Guidelines*, *Management Overview*, and *How-to Guide* are intended to meet the needs of different people in the driver education field and to support better, more focused evaluations. These three documents provide a set of tools that can be used to carefully and rigorously evaluate beginner driver education programs. It is hoped that their use will result in a growing body of evaluation information that can be built upon, leading to better driver education programs and, ultimately, safer young drivers.

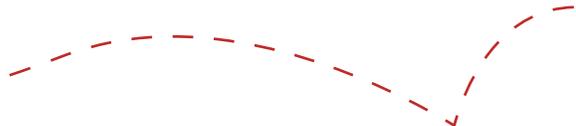
The three documents and related evaluation resources are also available on the website of the AAA Foundation for Traffic Safety, [www.aaafoundation.org](http://www.aaafoundation.org).



## EXECUTIVE SUMMARY

The development of *Evaluating Driver Education Programs: Comprehensive Guidelines*, as well as the two companion documents, involved extensive review and consultation processes, which assessed past evaluations of beginner driver education programs, as well as expert opinions and examples of program evaluation guidelines from other fields.

The *Guidelines* provide a road map to better understand why evaluation is important and how to conduct and interpret different types of driver education evaluations. The document includes a comprehensive and detailed examination of the full scope of beginner driver education evaluation. Important concepts and principles of evaluation research are explained, and an evaluation model and framework for beginner driver education are introduced. The evaluation guidelines that follow from this framework provide a systematic, stepped process for evaluating existing or new beginner driver education programs. Appendices include a review of past driver education evaluations, as well as a glossary of terms, worksheets, data collection tools, program standards, and guidance for hiring evaluators.



*. . . a systematic, stepped process for evaluating existing or new beginner driver education programs.*

The target audiences for this document and its two companion documents include: 1) evaluators who are asked to help in driver education evaluations; 2) program managers and administrators who may be encouraged to undertake systematic evaluation of driver education programs; and 3) traffic safety researchers who want to evaluate driver education programs. Most programs and regulatory bodies do not have professional evaluation staff, and in order to make comprehensive evaluation practical for field use, the *Guidelines* are intended to be “hands-on” and allow for customization to meet individual program needs and capabilities. Other audiences including regulators, insurers, state and provincial administrators, and consumer protection officials will find the *Guidelines* and companion documents helpful in understanding the importance of comprehensive and consistent evaluation of driver education programs.

Beginner driver education evaluation is different from general driver safety research and also has some unique needs compared to other types of education evaluation. The *Guidelines* provide a comprehensive approach to program evaluation specifically targeted to beginner driver education. They have been designed to expand the scope and quality of evaluations and to aid understanding of evaluation among those who are not specialist evaluators.

## Goals and Levels of Driver Education Evaluation

Evaluation research can serve two basic goals in the development of driver education: 1) to improve programs; and 2) to demonstrate the impacts of programs. The first goal is to find ways to improve program processes and products. The second goal is to determine what the program's outcomes are—whether the students have learned the specific knowledge and skills taught, and whether the program has an effect on crashes. The answers that a program evaluation provides depend on which goals are being evaluated.

A critical step is to determine the goals for the evaluation and the level of resources, time, and effort to be committed. Using information on the evaluation's purpose, goals, and objectives, as well as available financial and human resources, a level of evaluation can be identified that provides the best fit between these feasibility criteria and the evaluation to be implemented. Evaluations are categorized into four levels to accommodate different evaluation goals and the full range of driver education programs.

Briefly, Level 1 requires the fewest resources and includes planning activities that provide the foundation for implementing an extensive evaluation at a time when fewer constraints exist. Level 2 extends the evaluation tools used for improving a program and adds limited assessment of student knowledge and skill outcomes. This level can be considered for more intensive evaluations where some ability to manage data is available. Level 3 includes the ongoing formative program evaluation and improvement in Level 2, and expands the focus to outcome evaluations, quantitative methods, and benchmarking organization quality. Level 4 is the broadest and requires substantial resources and expertise to evaluate student skill and knowledge outcomes, safety impacts, and socioeconomic analyses.

These levels are explained in further detail in Chapter 1. Potential evaluators are guided on selecting a level and specific activities within their resource and evaluation capabilities. Looking beyond immediate evaluation capabilities, and establishing longer-term evaluation goals are also important, however. Evaluation should become a progressive and integral part of program implementation and improvement.

## Past Evaluations

The *Guidelines* present a state-of-the-art review of the driver education evaluation literature in Appendix A. This review summarizes earlier reviews of driver education evaluations and

examines recent or particularly important individual evaluations. Its main focus is identifying strengths and weaknesses of the highly diverse driver education evaluation literature. Unlike earlier reviews, the purpose is not to determine whether driver education has worked in the past, but rather to see how evaluation research can be improved and used to help driver education work better in the future. An overview is presented here to provide the reader with a brief background to the rationale that directed the guideline development process.

Driver education has long been mandated to address the tragically high risk of young novice drivers. It is also a popular and convenient means of achieving independent mobility, which is important to both young people and their parents. Driver education has strong “face validity” as a safety measure because parents think it makes their teenagers better, safer drivers. Objective, scientific evaluation has been applied to beginner driver education, mostly to the relatively standard high school programs of 30 classroom hours and 6 hours driving.

Evaluation reviews have concluded that safety effectiveness of driver education is not supported by the majority of evaluation studies. Past evaluations, however, have left many questions about the effectiveness of driver education partially or completely unanswered. Key areas where driver education evaluation has been found lacking include:

- *Program logic or theory*, that is the program rationale and model that explain how a program should meet its goals, or why we think it should work. The description of a program’s rationale is called a logic model and is often presented in a flowchart or table showing the relationships among important program components such as assumptions, goals and objectives, and activities and outcomes. There has been little evaluation of the theory or logic underlying various driver education programs.
- *Formative evaluation*, that is the application of evaluation tools to improve the content and delivery of a program. Little formative evaluation of intermediate effects has occurred, so how well driver education students achieve, retain, and use desired skills and knowledge is unclear. Driver education courses vary greatly in quality. There has been limited evaluation of program differences as well as the quality, comprehensibility, and usability of curriculum products and processes.
- *Methodological soundness*, that is the appropriateness of research design and methods used in evaluations. In the relatively small number of existing evaluations, problems of scope, design, and sampling limit unequivocal conclusions about the ultimate value of driver education at present and how its impact might be improved in the future.

## Driver Education Evaluation Studies

Driver education evaluation studies take three basic forms:

- 1) **Experimental studies**, such as randomized controlled trials that involve randomly assigning drivers to various training conditions and comparing subsequent crash rates and other measures;
- 2) **Quasi-experimental studies** that observe differences between self-selected groups of driver education students and those who learn to drive in other ways; and
- 3) **Ecological studies** that measure impacts on crashes of differences or changes in requirements or support for formal driver education.

The best known evaluation is the now-aging DeKalb County, Georgia, study, which has been considered the most extensive and rigorous evaluation of beginner driver education. Even this study, however, had serious limitations. The DeKalb evaluation involved randomly assigning 16,000 U.S. high school student volunteers to three groups: 1) intensive training, 2) minimal training, or 3) no formal driver education. The results failed to show a dramatic, long-term benefit of a special course, and reactions to the results had profound effects on driver education. Its conclusions are still controversial after more than 20 years. Most other evaluation studies have been more limited in scope and scale. With some exceptions, the experimental evaluations typically found no statistically significant effects of driver education on crash records. Among more recent evaluation studies, several quasi-experimental and ecological studies have been conducted. Two large-scale ecological evaluations showed positive effects of driver education.

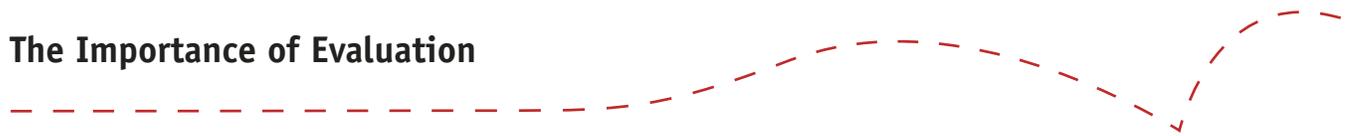
Overall, scientific evaluation of driver education has been quite limited. The safety impacts of beginner driver education are particularly hard to evaluate because suitable comparison groups are hard to establish. Many of the earliest evaluations compared groups of young drivers who received different forms of driver education but also differed in other ways that might affect their driving record and other outcome measures. Important extraneous or confounding differences include location of residence, income, or other important socioeconomic factors. Even when they can be established, equivalent comparison groups are hard to maintain over time, since assigned or selected groups can have different dropout rates. In evaluations of crash rates, very large groups of drivers must be studied to detect moderate differences in rates.

When considering safety effects, the proper measure of crash experience is critically important and also rather controversial. How crashes are measured can provide quite different

results, whether through self-report, government records, or insurance records. Crash experience can also be expressed in different rates, depending on the denominator chosen, and this also has important implications. Crashes per licensed driver can give very different results from crashes per capita. Crashes per mile traveled gives different results from population-based measures. Rate questions are not minor technical issues, since they lie at the core of fundamental evaluation questions such as, “Is the proper success criterion for driver education safer mobility or a safer youth population?” Different crash rate measures would be used to measure results according to these different criteria.

More systematic and comprehensive evaluation is essential to the future of driver education. Past studies have demonstrated that assumptions about what is effective in reducing young driver crash risk are not always well founded; therefore, objectively evaluating both existing and new programs is critical. While evaluation is important to improving the effectiveness and efficiency of driver education, recognizing its limitations is also important. Evaluation of driver education, like driver education itself, is evolving and still far from its ultimate conclusion. The *Guidelines* are intended to accelerate development of more and better evaluation, which in turn should support development of more effective driver education.

## The Importance of Evaluation



There are many important reasons to evaluate programs. These include: identifying program strengths, weaknesses, and ways to improve programs and measure progress; collecting evidence on program effectiveness and impacts; determining or strengthening program accountability; sharing what works and doesn't with others; and influencing policy makers, partners, sponsors, and funding agencies.

*More systematic and comprehensive evaluation is essential to the future of driver education.*

Program evaluation is critical for developing more effective programs in the future, but its past limitations need to be recognized and corrected. Good evaluation is designed and organized using a comprehensive evaluation framework. This framework should be: 1) based on sound evaluation theory; 2) structured using relevant evaluation models; and 3) developed using benchmark program evaluation standards to ensure the quality of all aspects of the evaluation.

Program evaluation standards are used throughout the evaluation process in the *Guidelines* as benchmarks for checking evaluation quality. In 1994, the Joint Committee on Standards

for Educational Evaluation developed the standards used in these *Guidelines*. They have been approved by the American National Standards Institute (ANSI) and widely adopted in many fields, including education, public health, injury prevention, and human services. The Joint Committee grouped the standards into four categories:

1. *Utility*—The evaluation will serve the information needs of the intended users.
2. *Feasibility*—The evaluation will be realistic, prudent, diplomatic, and frugal.
3. *Propriety*—The evaluation will be conducted legally, ethically, and with regard for the welfare of those involved and affected by its results.
4. *Accuracy*—The evaluation will reveal technically adequate information about the worth or merit of the program being evaluated.

Ongoing evaluation is critical, not only for determining changes in crash rates, the ultimate criterion of success, but also to assess and improve programs. A comprehensive approach to evaluation addresses driver education's theory of effect or logic model, its products and processes, student learning outcomes, and the program's quality of operation and management. The scope of driver education program evaluation should be as comprehensive as practically possible. Many opportunities exist for program improvement by conducting evaluations across the program's life cycle, beginning with the program's theoretical or logical basis, progressing through program development and implementation, and finally reckoning costs and benefits. Evaluation should be an integral part of this process.

## **An Integrated Evaluation Framework for Driver Education**

A broad, comprehensive evaluation approach is required for driver education, one that realistically incorporates several key aspects of evaluation. These include program description, rationale and logic, context, products, and processes; the entire continuum of outcomes and impacts; all relevant delivery organizations, including businesses and governments; concerned stakeholders, such as consumers, providers, insurers, and regulators; and a range of applicable evaluation methods and appropriate measures. The driver education evaluation framework developed for the *Guidelines* addresses the full range of program evaluation needs—program theory, context, standards, products, processes, outcomes, and impacts.

There are two types of evaluation—formative and summative. Formative evaluation involves evaluation processes intended to improve program components, such as content and

teaching methods. Summative evaluation includes evaluation processes intended to assess the impacts of the program, such as student learning achievement and crash reduction. The framework includes both evaluation types. Within each type, many specific evaluation targets or aspects of a program can be evaluated. Targets range from user needs and stakeholder expectations, to management processes, instructional products and processes, learning and behavioral outcomes, and safety impacts.

The framework also provides a wide range of data gathering and research methods related to the evaluation targets. Examples of possible evaluation methods include pilot testing and content analysis; qualitative research, such as focus groups and interviews; standardization, such as benchmarking, certification, and auditing; instrumented vehicle observation; surveys; record studies and modeling; ecological studies; longitudinal studies; quasi-experiments; and randomized controlled experimental trials.

## **Step-by-Step Evaluation Guidelines for Driver Education Programs**

These *Guidelines* present comprehensive step-by-step procedures for designing and carrying out an evaluation. The *Guidelines* can assist researchers and program administrators determine the type and scale of evaluation to undertake in their specific circumstances, and then how to implement the evaluation. Guidance is also provided for creating a suitable team to design and manage the evaluation. The *Guidelines* are organized into five key evaluation steps:

1. Focus the evaluation.
2. Select the evaluation methods.
3. Identify the data collection plan and tools.
4. Gather, analyze, and summarize the data.
5. Interpret and act upon the evaluation findings.

Each step begins with a brief description and an explanation of why it is important. The activities required to complete the step are then explained in detail. Examples and worksheet samples are provided, and the program evaluation standards are integrated into each step. Even though these steps are ordered in a logical sequence, the order can be modified to fit the specific circumstances of individual programs.

Research design guidance on how to structure valid comparisons, control potential biases, and determine sample sizes is also provided. These are all key issues that have often

been inadequately managed in past driver education evaluations. The practical aspects of choosing the kinds of data to collect and the logistics of data collection, as well as identifying appropriate data handling and analysis procedures, are addressed. Interpreting, reporting, and using evaluation results are discussed, including the most important task of determining which changes to implement as a result of the evaluation and developing an action plan. Guidance on hiring external help is also provided.

## **Driver Education Evaluation Scenarios**

Two evaluation scenarios for different types of driver education programs complete the *Guidelines*. The first provides a template for conducting a basic program evaluation of a hypothetical medium-size driver education program, essentially a Level 2 evaluation. The scenario is a concrete example of how evaluators might use the evaluation model and framework to conduct an evaluation project to improve a modest program. It includes a logic model to frame the hypothetical evaluation; a research design; and examples of several evaluation tools, actual survey questions, and checklists that can be used by similar driver education programs. This scenario represents an evaluation level that can be used to assess the program's products and processes, and to measure a range of learning outcomes and survey reports of the students.

The second scenario builds on the first and offers examples of tools for undertaking a more advanced evaluation of a larger driver education program. This scenario illustrates how to use the evaluation model and framework in a more comprehensive approach, including measuring safety impacts. The hypothetical program is a large, highly developed program. This scenario expands the range of potential evaluations, in which all outcomes and impacts can be evaluated, including program impacts on students' subsequent crashes.

Both scenarios are accompanied by appendices, which provide examples of actual tools to assist evaluation teams who wish to consider either a medium or large-scale evaluation.

## **Conclusion**

The evaluation structure for driver education builds on the concepts of evaluation theory and logic models, evaluation models, and program evaluation standards. This structure is based on a composite driver education evaluation model and framework. The *Guidelines* will help driver education researchers and managers determine which type and scale of evaluation fits their specific circumstances. They can then conduct a program evaluation based on sound and astute decisions about what the evaluation intends to achieve and how it will aid program

improvement and impact. These stepped evaluation actions can be used to improve driver education program evaluations to a higher standard than in the past.

The *Guidelines* are expected to promote more consistent and competent application of evaluation principles and methods. Once adopted and implemented on a regular basis, they will establish new standards for driver education evaluation, resulting in improvement in both driver education evaluation and driver education programs. While evaluation is important to improving the effectiveness and efficiency of driver education, recognizing its limitations is also important. This recognition has been lacking in the past and has led to unfortunate policy decisions. Evaluation of driver education, like driver education itself, is evolving and still far from its ultimate conclusion. These *Guidelines* provide the understanding and step-by-step guidance to increase the potential of evaluation to help driver education programs improve their products and processes and become more effective in assisting youth become safer drivers.





# 1. Introduction

**D**river education programs seek to teach novice drivers the skills and knowledge necessary to achieve independent mobility and perform as safe and efficient drivers. Driver education evaluation aims to *improve* driver education programs, or to demonstrate or *prove* the safety benefits of improved programs. While quite a number of program evaluations have been carried out over the years, not much compelling evidence exists yet that young people who complete driver education programs subsequently drive more safely or have fewer crashes than those who receive less formal driver instruction. As a result, the safety effects of beginner driver education have been highly controversial. Many questions need to be considered. Do driver education programs enhance or detract from safety? Do some types of driver education programs lead to better educational and safety outcomes than others? Can we identify which components of driver education programs do and do not work? Do programs meet their learning objectives? How can driver education programs be improved in order to yield safer young drivers?

While answering these questions would be desirable, most have been left partially or completely unanswered. Methodological weaknesses have plagued many evaluations. Some have neglected to assess learning outcomes, have used comparisons between groups that were not really comparable, or have used sample sizes too small to reliably find moderate effects. Such common inadequacies have led to different interpretations and controversy over the meaning of past driver education program evaluations.

These *Guidelines* are intended to lead to better, more focused evaluations by providing a set of practical tools that can be used to rigorously examine beginner driver education programs. It is hoped that their use will result in a growing body of evaluation data that can be built upon, leading to better driver education programs and ultimately improved safety outcomes.



## Overview of the *Guidelines*

**Chapter 1** presents an overview of how to use the *Guidelines* and discusses two important factors: target audiences and the rationale for developing comprehensive evaluation for driver education.

**Chapter 2** presents an overview of important concepts in evaluation theory and practice. This provides the reader with the evaluation context within which to understand the *Guidelines*. This chapter also introduces driver education evaluation concepts and structures and describes the evaluation model and framework within which the *Guidelines* have been developed.

**Chapter 3** presents comprehensive guidance for evaluation of driver education programs through a series of steps that can be followed sequentially. Alternative evaluation approaches, based on evaluation purpose and resource criteria are identified. This chapter also includes examples, worksheet samples, and program evaluation standards.

**Chapter 4** provides a template for conducting a basic program evaluation of a hypothetical medium-size driver education program. The chapter, and related appendices, include a logic model to frame the hypothetical evaluation; a research design; and an outline of several evaluation tools, actual survey questions, and checklists that similar driver education programs can use. This scenario represents a level of evaluation that can be used to assess a program's products and processes, and to measure a range of student learning outcomes.

**Chapter 5** adds to the template presented in Chapter 4 and offers examples of tools for undertaking a more advanced evaluation of a much larger driver education program. In this case, the hypothetical program is based in a state or province in North America. This scenario involves more comprehensive evaluation, in which the full array of outcomes and impacts can be evaluated, including any program impact on crashes.

**Chapter 6** which concludes the *Guidelines*, is followed by a section on **Evaluation Resources** that can be referred to for specialized evaluation methodology and analysis information, and includes examples of other evaluation guidelines.

**Appendices** include a review of the driver education evaluation literature, a glossary of terms, program evaluation standards' definitions, and worksheets that can be copied for use in driver education evaluations. Questionnaire formats and items that can be used in surveys of teens, parents, and instructors, as well as benchmarking standards and guidance on how to hire an evaluator, are also provided.

## How to Read the *Guidelines*

Because the review of past evaluations and a comprehensive perspective on driver education evaluation are both essential components of the *Guidelines*, this document is necessarily detailed. This detail will help different audiences understand the importance of evaluation and how to implement more effective evaluations. Some readers will be interested in reading the entire document to appreciate the background, as well as understand what is actually involved in different aspects of driver education evaluation. Others will want to be directed to the sections that are most relevant to their needs.

As a quick guide, readers who are new to evaluation but are not going to be directly involved in a driver education program evaluation may want to read Chapter 2 and perhaps skim Chapters 4 and 5 to get an idea of why evaluation is important and what is actually involved in a medium- or large-scale driver education evaluation.

Readers interested in the details of how to conduct an evaluation of a driver education program should peruse Chapter 2 but spend more time working through Chapters 3-5.

Readers wishing to learn enough about evaluating driver education to be able to hire an external evaluator should read Chapter 2 and skim Chapters 4 and 5 to see how key aspects of different-sized evaluations can be conducted. They should then look carefully at Appendix J, which provides more detailed information on how to hire an evaluator. The evaluation resources on page 173 may also be of assistance.

Readers with evaluation or research expertise may be interested in learning more about the history of evaluation experience in driver education. Appendix A introduces readers to the driver education evaluation literature. More detailed information about how to implement an evaluation of a driver education program is found in Chapters 2 and 3.

## Target Audiences

These *Guidelines*, as well as the companion documents, are written for three primary audiences: 1) evaluators who are asked to help with driver education evaluations; 2) driver education managers and administrators who may be encouraged to undertake or manage systematic evaluations; and 3) traffic safety researchers who want to evaluate driver education programs.

While evaluation of beginner driver education is a form of driver safety research, it has special requirements that are different from more general research approaches. Driver education evaluation also has requirements that are different from other types of education evaluation. The *Guidelines* will help evaluators perform relatively simple as well as more comprehensive evaluations of beginner driver education, which can be compared and built upon over time. They are designed to provide appropriate tools for both researchers who wish to expand the scope and improve the replicability of driver education evaluations, and non-research professionals in the business of teaching or overseeing driver education.

Also important among the target audiences are driver education managers and administrators. Most programs, and even government regulatory bodies, do not have professional evaluation staff. This situation is not unique to driver education—it is also widely reflected in the public health, human services, and education fields. Making a comprehensive evaluation framework practical for field use requires a hands-on set of steps or guidelines. These steps allow the evaluation framework to be customized to meet individual program needs and capabilities, while maintaining consistency of approach and techniques.

Other audiences including regulators, insurers, state and provincial administrators, and consumer protection officials will find the *Guidelines* and the supplementary documents helpful in understanding the importance of comprehensive and consistent evaluation of driver education programs.

For all who use the *Guidelines*, the driver education evaluation framework provides a common starting point. The scope of individual program evaluations, however, will vary depending on several factors, including program size, the rationale for undertaking the evaluation, the evaluation's goals and objectives, staff and financial resources, and the skill base of those undertaking the evaluation.

## **Driver Education Programs in the U.S. and Canada**

Beginner, pre-licensing driver education has been a traditional social response to young people's needs for mobility and the tragically high rates of injuries and deaths among young novice drivers. Typically, driver education has been a formal course of study, delivered by a paid instructor, and has included classroom instruction and in-car training. In some programs, the classroom and in-car components are closely coordinated, but in

others, classroom instruction and in-car training are delivered quite separately, and new drivers may receive one or the other but not necessarily both. A few jurisdictions recognize a formalized course of instruction delivered by parents.

Driver education has been widely available in public secondary schools, although this type of driver education has declined substantially. High school driver education programs typically have 30 hours in the classroom and 6 hours of driving. For the most part, they cover legal requirements, vehicle handling, and efforts to motivate beginners to fear the consequences of crashes. Commercial driver educators often use similar program standards and may operate within a secondary school under contract. Driver education has become a substantial but highly fragmented industry. Commercial and other interests encourage student participation. Driver education is widely supported by insurance premium discounts and driver licensing provisions that provide incentives to complete formal instruction.

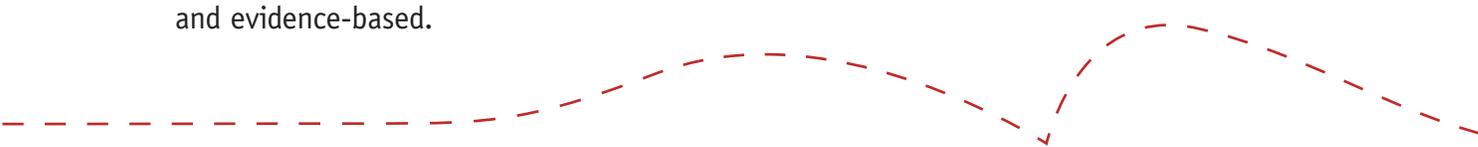
*Driver education has become a substantial but highly fragmented industry.*

Traditionally, driver education meant instruction only before the new driver was licensed to drive independently. A less common form of driver instruction occurs after drivers are licensed to drive independently. In a few jurisdictions, such as Finland and Michigan, new drivers are required to take a second stage of training after they have been licensed drivers for a short period of time.

An interesting recent development within driver education is the entry of well-funded private corporations, along with the more traditional government, private foundations, and not-for-profit associations. In the United States, for example, MetLife Casualty, an automotive insurer, took an active role in supporting Top Driver, which has made efforts to consolidate the private driving school industry, in partnership initially with General Motors and then with equity participation from Ford Motor Company. In the 1990s, Ford took a majority ownership in Young Drivers, a long-established international driver education provider based in Canada. AAA, which has a longstanding record of producing and delivering driver education and training programs, has developed extensive new curriculum approaches and

material. AAA's *Licensed to Learn*, for example, is a very comprehensive risk-based approach that incorporated new technology and interactivity when it was introduced in 2001. More recently, AAA has upgraded several of its programs, including *Teaching Your Teens to Drive* and *Driver Improvement Programs*. These programs are offered or taught directly by many local AAA Clubs throughout the United States. AAA has also embarked on a new initiative that could lead to a national network of AAA-approved driving schools.

Trends also exist in instructional methods and program delivery. Traditionally, all driver education activities involved face-to-face interaction between instructor and learner, although classroom instruction was often supported with film and video, and sometimes with simulators. More recently, self-instruction, computer-based instruction, simulation, and web-based instruction have become prevalent. While these changes have produced profound changes in the technological, business arrangement, regulatory, and driver licensing environments in which driver education operates, whether they will improve the safety effectiveness of driver education is still unclear. Most changes are directed to delivery efficiency, and they are largely technology driven and entrepreneurial, rather than systematic and evidence-based.



To a much greater extent than in the past, driver education is much less a single entity than it has been historically. Some high school driver education programs involve thousands of students each year, while some jurisdictions' programs are small and may only teach a small minority of new drivers. Most commercial driving schools are relatively small, many having only a single location. A few driving schools have many locations and teach thousands of students each year. Web-based programs may also teach many thousands of students. Driver education costs to students can range from free to hundreds of dollars. Operating input costs can vary greatly, from well under \$100 per student to several hundred dollars. School operating standards can range from none to strict centralized control and ISO certification, and instructor qualifications range from very low levels to highly qualified professional teachers.

The growing diversity of driver education programs is an indication of vigorous development and may lead to greater potential effectiveness in the future. While this diversity also increases the need for more and better evaluation, it also complicates it. As the following chapters will show in more detail, the larger public and private programs have

potential for a comprehensive range of evaluations. Evaluation methods can improve these programs, and it should be possible to demonstrate the safety impacts of larger programs where such impacts occur.

Smaller programs have good potential for formative types of evaluations and for evaluating student learning outcomes. As shown next, use of appropriate approaches to evaluation can help make these programs as good as possible in terms of instructional and operational effectiveness.

## Approaches to Driver Education Program Evaluation

One of the critical points in the evaluation process is to determine how comprehensive an evaluation to undertake. An evaluation team has to make a decision fairly early in the evaluation process about the goals of the evaluation and the level of resources, time, and effort that will be committed. Using this information, a level of evaluation can be identified that provides the best fit between these feasibility criteria and the evaluation to be implemented.

The *Guidelines* use four evaluation levels, representing a full range of evaluation options. These are general levels of effort that evaluations can undertake according to program needs, size, and resources. They can help to determine the scope of an evaluation. Detailed information on the levels and examples of related evaluation activities are found in Chapter 3, Step 2, and summarized here.

**Level 1** requires the fewest resources. It includes program and evaluation planning activities that provide the foundation for moving on to a more extensive evaluation at a time when fewer constraints exist. These activities, described in detail in Chapter 3, Step 1, include: describing the program; setting program goals and objectives; developing a logic model that shows how the program is expected to meet its objectives; and identifying evaluation objectives, questions, and targets. Benchmarking the program against industry standards, or surveying customers to determine satisfaction levels, can be undertaken. Examining instructor qualifications, the uniformity of instructional delivery, and facility management can also be part of this level. A well-managed local program provider or school authority can aspire to this level, consisting primarily of formative evaluation and qualitative evaluation methods.

**Level 2** builds upon evaluation efforts of Level 1 and moves toward a more comprehensive evaluation process. Level 2 includes Level 1's planning activities and adds the assessment of student knowledge and skill outcomes. Level 2 can be considered if more resources than required by Level 1 are available, or by Level 1 teams who have completed most of their activities and are prepared to undertake a more active evaluation process.

**Level 3** includes the ongoing formative program evaluation and improvement of Level 2, and expands the focus to outcome evaluations, quantitative methods, and benchmarking organization quality. This level is more likely to be undertaken by organizations with more resources to allocate to ongoing program evaluation and improvement, such as major program providers, large materials suppliers, industry associations, and smaller state or provincial governments.

**Level 4** the most advanced, requires substantial resources and expertise, most likely available to national and larger state or provincial governments, or large research organizations with a secure funding base. This level involves comprehensive outcome evaluation including safety impact evaluation and socioeconomic analyses.

Having identified all available options, evaluation teams can effectively assess their present capabilities to evaluate their programs. This will help them determine the most appropriate level of evaluation to undertake. Although the range of evaluation levels is presented as a sequential and progressive process, each evaluation team will choose priority activities within the resource and evaluation capabilities of the program. Evaluators, however, need to look beyond immediate program evaluation capabilities and establish longer-term goals. In this way, the evaluation becomes a progressive and integral part of program implementation and improvement.



## 2. Evaluation Model and Framework

This chapter presents an overview of important evaluation concepts and terminology. Evaluation knowledge, theory, and models provide the foundation for developing a composite driver education evaluation model. This model, in turn, is used to create an integrated evaluation framework for driver education programs. This is the framework used to structure the evaluation guidelines. Program managers and administrators who want to understand more about evaluation concepts and the structure used to develop the *Guidelines* will benefit from this chapter.

### Why Evaluate?

Evaluation is an essential part of the life cycle of good programs. It is as important as careful planning, development, and delivery. Knowing what a program is accomplishing and how it is doing relative to its plan are essential. There are many important reasons to evaluate programs, including:

- Identifying program strengths and weaknesses;
- Reflecting on and measuring progress;
- Identifying ways to improve programs;
- Making decisions about how to change programs;
- Collecting evidence on program effectiveness and impact;
- Assessing program efficiency;
- Determining or strengthening program accountability;
- Sharing what works and doesn't with other program managers, researchers, and evaluators; and
- Influencing policy makers, partners, sponsors, and funding agencies.

The purpose of each evaluation will, of course, be determined by the program's goals, needs, and resources.

## What is Program Evaluation?

Evaluation means knowing, in concrete, specific terms, what the goals and objectives of a program are, and how effective and efficient the program is in achieving them, in objective, scientific terms. It is important to approach evaluation with an understanding of some of the most frequently used evaluation terms. The most important definitions are presented here, and a glossary of terms is found in Appendix B.

A *program* is a series of activities supported by resources and intended to achieve specific outcomes among particular target groups (Porteous, Sheldrick, and Stewart 1997). There are many definitions of *program evaluation*. Those following help to understand the wide scope of evaluation purposes and the range of possible evaluation activities:

**A** *program is a series of activities supported by resources and intended to achieve specific outcomes among particular target groups.*

- The systematic determination of the quality or value of something (Scriven, in Davidson 2004).
- The systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming (Patton 1997).
- An adaptation of social research methods to studying social interventions so that sound judgments can be drawn about the social problems addressed, and *the design, implementation, impact, and efficiency* (emphasis added) of programs that address those problems (Rossi, Lipsey, and Freeman 2004).

Program evaluation should be an integral part of program planning and implementation, and thus, the first tasks in an evaluation are also program planning tasks. As both processes have a common goal, a successful program, the overlap is expected.

*Stakeholders and target groups* include individuals and groups, both internal and external to the program, who have an interest in the program and its evaluation. These are the people involved in or affected by the evaluation. Target groups are the people affected by

the program, that is, students, their parents, and staff. Other stakeholders include community members and organizations, decision makers, partners, sponsors, and funding agencies.

## Why is Evaluation Important? Lessons from Related Fields

Education, public health, injury prevention, and human services are all fields related to driver education that have extensive program evaluation experience. Educational evaluation emphasizes evaluating a program's internal functions suggesting that evaluation's most important purpose is to improve, rather than prove or determine, the value of the program (Stufflebeam 2003; W.K. Kellogg Foundation 1998). Typically, if students achieve learning objectives, an education program is considered successful. Road safety evaluation more often focuses on ultimate public health impacts—saving lives and reducing injury. Arguably, the lack of integration of health and education perspectives in driver education evaluation has contributed to the controversy surrounding driver education and its failure to develop fully to meet its safety goals.

In the education and human services evaluation fields, there are extensive taxonomies of evaluation models. Some models, however, are believed to have limited use in program evaluation; for example, Stufflebeam (2001) considers experimental models to have only limited utility. Because the experimental models have been dominant in driver education evaluation, however, they are particularly important for the *Guidelines*.

**E***ducation, public health, injury prevention, and human services are all fields related to driver education that have extensive program evaluation experience.*

Stufflebeam (2001) suggests that experiments may provide less information than is needed for guiding program improvement. This concern can presumably be mitigated by supplementing experiments with a broader range of measures, including knowledge gain, skill performance, and multiple impact measures. Quasi-experimental studies also attempt to overcome the difficulties with experimental designs, but these, too, must be supplemented with additional measures. Skepticism of experimental models is tempered by recognition of the approach's virtues, principally that it may permit firmer conclusions

regarding causal connections between programs and consequences. Experimental methods have strong credibility from their long use in medicine and agricultural research, and their more recent prominence in experimental psychology, which provides the training for many safety researchers.

Health promotion, the behavioral component of public health, also offers guidance in program evaluation. It is one of the leading disciplines in the comprehensive planning of individual and organizational behavior change. While the behavioral focus in public health is relatively recent, the intervention and evaluation literature is extensive. It covers all types of health-related behaviors including disease prevention, diet and nutrition, fitness and active living, injury prevention, maternal/newborn care, smoking prevention and cessation, and alcohol and other substance abuse. Recent work in the field has attempted to bring coherence to the very large volume of diverse health promotion experiences and to focus on understanding the successes and failures in influencing health behaviors.

Public health, like driver education, also has to confront the issues of defining successful health outcomes and impacts. Is it enough to change health-related knowledge, attitudes, or even behavior? Or must we also look for some ultimate measures of impact, such as reduced incidence of disease, measures of increased well being and longevity, or even a positive cost/benefit analysis before deciding that a program is effective?

One of the most comprehensive health promotion models, the PRECEDE/PROCEED model (Green and Kreuter 1999), addresses planning and evaluation needs in health promotion and health education and provides guidance for safety programming. It outlines a series of phases in planning, implementing, and evaluating programs. The framework takes into account the multiple factors that shape health-related behavior and guides the identification of specific subsets of these factors as targets for intervention.

Education, public health, injury prevention, and human services have embraced comprehensive evaluation models, frameworks, and prescriptive guidelines to advance the effectiveness of interventions and their evaluation. In these fields, logic models are routinely used to guide the development of evaluation. Toolkits, handbooks, and checklists have also been developed to provide detailed guidance for conducting program evaluations in these fields. Several examples are listed in the Evaluation Resources section found on page 173.

A strong research basis supports the need for re-inventing driver education (Lonero et al. 1995; National Transportation Safety Board 2005). The challenge, however, is that a

broad program integrating motivational, social, family, community, and regulatory influences is probably required. Taking the lead from other fields, comprehensive approaches to program development and evaluation are needed for driver education to become an effective multi-faceted influence program.

## What is Good Evaluation?

A good evaluation is designed and organized using a comprehensive evaluation framework. This framework should be: 1) based on evaluation theory; 2) structured using relevant evaluation models; and 3) developed using benchmark program evaluation standards to ensure the quality of all aspects of the evaluation.

### *Evaluation Theory*

Why should we use theory in objective, scientific evaluation? Theories are “summaries of formal or informal observations, presented in a systematic, structured way, that help explain, predict, describe or manage behavior” (Goldman and Schmalz 2001, 277). A theory describes the linkages among the various parts of a program and its goals and objectives. This is the starting point for evaluation. The components of good evaluation theory provide the foundation for developing effective evaluations. Evaluation involves four core questions that are related to the key components of good evaluation theory (Ottoson, 2006).

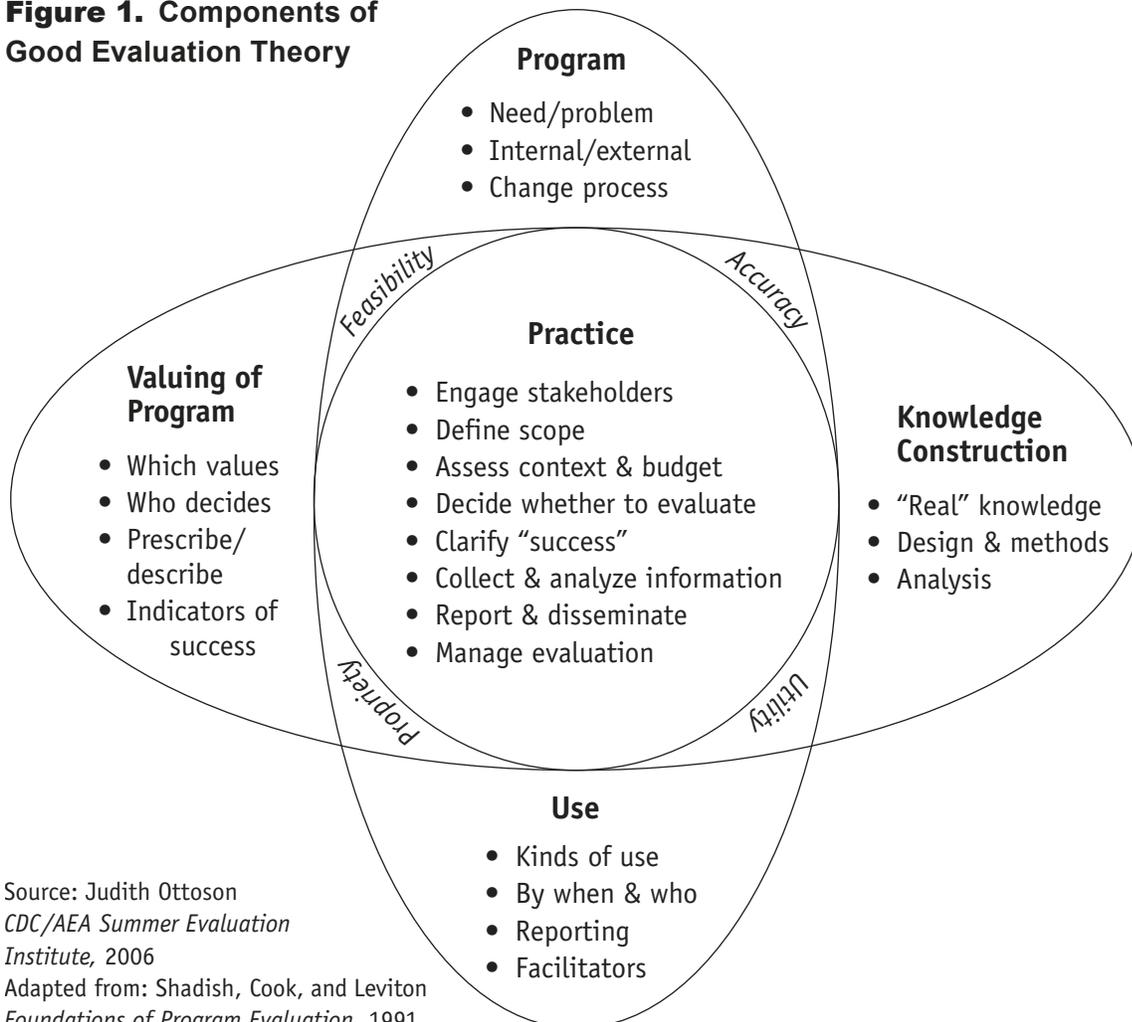
- 1. What is the program?** To answer this question, relevant program information should be identified and described specifically. This information should include the needs that the program addresses, the full scope of the program (from policy to activities), detailed program structure, the external context, and the processes by which the program is expected to achieve its objectives.
- 2. What counts as evidence of the program’s value or merit?** Answering this question includes identifying the program’s values, the people who decide these values, the ways to make the values transparent and available to stakeholders and target groups, and indicators and standards of success.
- 3. How is knowledge constructed or built about the worth or value of the program?** To answer this question, we must determine whether the knowledge about the program that will be gained from the evaluation is special, and what

knowledge is credible, meaningful, and important (i.e., “real”) to program staff, users, and other stakeholders.

- 4. How are the evaluation findings used?** Determining how evaluation findings will be used is also important. This includes who will use the findings and when, who will be responsible for facilitating the use of the findings, who will facilitate their use, and which communication formats will be used for disseminating the findings.

These four questions represent the components of good evaluation: program description, the valuing of the program, knowledge construction, and use of evaluation findings, as seen in Figure 1. These components intersect and influence the evaluation actions, represented in the center circle of the diagram under “Practice,” that make up program evaluation. Note how the program evaluation standards of feasibility, accuracy, utility, and propriety are integrated into this diagram (see pages 37-39 for detailed explanation).

**Figure 1. Components of Good Evaluation Theory**



Source: Judith Ottoson  
 CDC/AEA Summer Evaluation  
 Institute, 2006

Adapted from: Shadish, Cook, and Leviton  
 Foundations of Program Evaluation, 1991

## *Evaluation Models*

Evaluation specialists have developed numerous evaluation models. Four broad groups are presented to illustrate the range of possible evaluation approaches (Trochim 2001).

- Scientific-experimental models emphasize methodological rigor and objectivity, and include experimental and quasi-experimental evaluations, objectives-based education research, economics-based evaluations including cost/benefit and cost effectiveness analysis, and theory-driven evaluation.
- Management-oriented systems models are predominantly used in business and government but can be applied in many organizational settings. Examples are PERT (Program Evaluation and Review Technique), CPM (Critical Path Method), and CIPP (Context, Input, Process, and Product). They emphasize evaluation comprehensiveness within the larger organizational context.
- Qualitative/anthropological models emphasize the importance of subjectivity, observation, and human interpretation in evaluation. Included are naturalistic and qualitative evaluation approaches, such as grounded theory (the systematic generation of theory from research data).
- Participant-oriented models focus on the importance of participants to the evaluation process, and include client-centered, stakeholder, and consumer-oriented approaches.

Each type of evaluation model brings unique and valuable perspectives to the evaluation process. While not all models are useful in every evaluation, an optimal framework integrates the relevant aspects of all four categories. Although comprehensive evaluations are most desirable when feasible, even small-scale evaluations, limited in scope perhaps by resources, time, or logistical constraints, should be undertaken rather than opting for no evaluation.

## *Program Evaluation Standards*

Ensuring the adequacy and quality of the evaluation itself is as critical as evaluating the program. Program evaluation standards are used throughout the evaluation process as

benchmarks against which to check evaluation quality. The standards used in these *Guidelines* were developed in 1994 by the Joint Committee on Standards for Educational Evaluation. They have been approved by the American National Standards Institute (ANSI) and have been widely adopted in many fields including education, public health, injury prevention, and human services.

### Joint Committee Standards

The Joint Committee grouped the standards into the following four categories:

1. *Utility*—Is the evaluation useful? Utility standards ensure that an evaluation will serve the information needs of the intended users.
2. *Feasibility*—Is the evaluation viable and practical? Feasibility standards ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.
3. *Propriety*—Is the evaluation ethical? Propriety standards ensure that an evaluation will be conducted legally, ethically, and with regard for the welfare of those involved in the evaluation, as well as those affected by its results.
4. *Accuracy*—Is the evaluation correct? Accuracy standards help ensure that an evaluation will reveal technically adequate information about the features that determine the worth or merit of the program being evaluated.

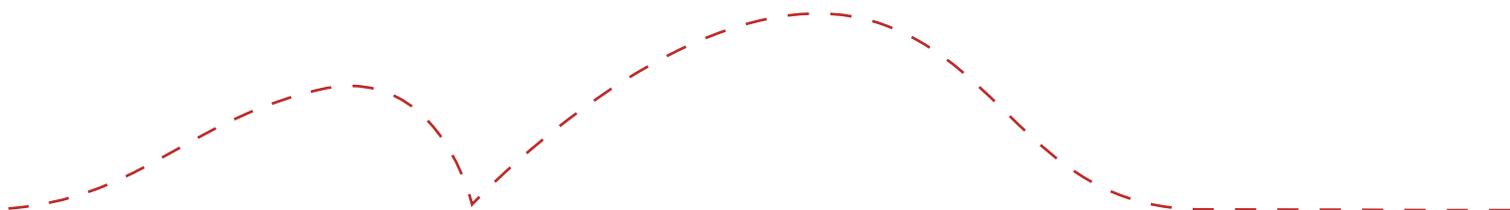
These standards provide guidance in conducting sound and fair evaluations, and are especially important to use when deciding among evaluation options. The standards can help avoid an unbalanced evaluation, such as one that is accurate and feasible but not useful, or one that is useful and accurate but too costly or time-consuming. In addition, the standards can be applied while planning an evaluation and throughout its implementation (Centers for Disease Control and Prevention 1999).

As Table 1 indicates, within each of the four categories are several specific standards that help assess an evaluation's quality. Detailed definitions of the standards are found in Appendix C.

**Table 1. Program Evaluation Standards**

Utility Standards		Feasibility Standards		Propriety Standards		Accuracy Standards	
U1	Stakeholder identification	F1	Practical procedures	P1	Service orientation	A1	Program documentation
U2	Evaluator credibility	F2	Political viability	P2	Formal agreements	A2	Context analysis
U3	Information scope and selection	F3	Cost effectiveness	P3	Rights of human subjects	A3	Described purposes and procedures
U4	Values identification			P4	Human interactions	A4	Defensible information sources
U5	Report clarity			P5	Complete, fair assessment	A5	Valid information
U6	Report timeliness and dissemination			P6	Disclosure of findings	A6	Reliable information
U7	Evaluation impact			P7	Conflict of interest	A7	Systematic information
				P8	Fiscal responsibility	A8	Analysis of quantitative information
						A9	Analysis of qualitative information
						A10	Justified conclusions
						A11	Impartial reporting
						A12	Metaevaluation

*Within each of the four categories are several specific standards that help assess an evaluation's quality. Detailed definitions of the standards are found in Appendix C.*



## Depicting Programs through Logic Models

When planning a program evaluation, it is important to identify and clearly articulate the theory or relationships upon which the program is based, and the goals and objectives of the program. A program theory makes explicit the outcomes that the program is intended to achieve, and the strategies and interventions it uses to get there. In other words, program theory is a statement of how a program is supposed to work. A program theory typically consists of an explanation about: 1) how the program's objectives are related to its needs; 2) how the program brings about change in outcomes; and 3) how the program is supposed to operate.

A tool, called the logic model, is routinely used in many fields (such as injury prevention, education, public health, management, and evaluation) to depict a program's theory and assumptions, as well as document program planning and implementation details. Logic models can also guide the development of evaluation processes and activities. They are typically represented in a flowchart, table, or diagram that displays the relationships between program goals, assumptions, objectives, activities, target and stakeholder groups, and outcomes.

### How a Logic Model Contributes to Evaluation

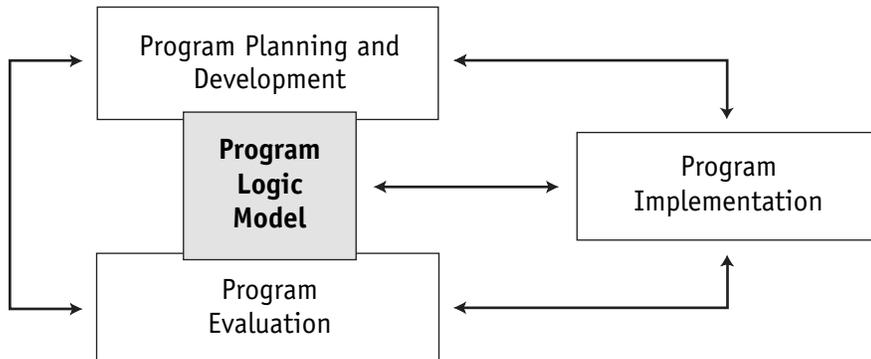
Specifically, a logic model contributes to evaluation planning by:

- Summarizing the key program components;
- Explaining the program's rationale;
- Helping explain how the program's activities contribute to its intended goals and outcomes;
- Assisting in the identification of important evaluation questions; and
- Helping program staff and stakeholders understand and comment on the evaluation plan.

Sources: Porteous, Sheldrick, and Stewart 1997; The Health Communication Unit 2006.

Figure 2 shows, in simplified terms, the relationships between the logic model and program planning, implementation, and evaluation. Note the cyclical and interdependent processes represented in the diagram.

**Figure 2. Planning and Evaluation Cycle**

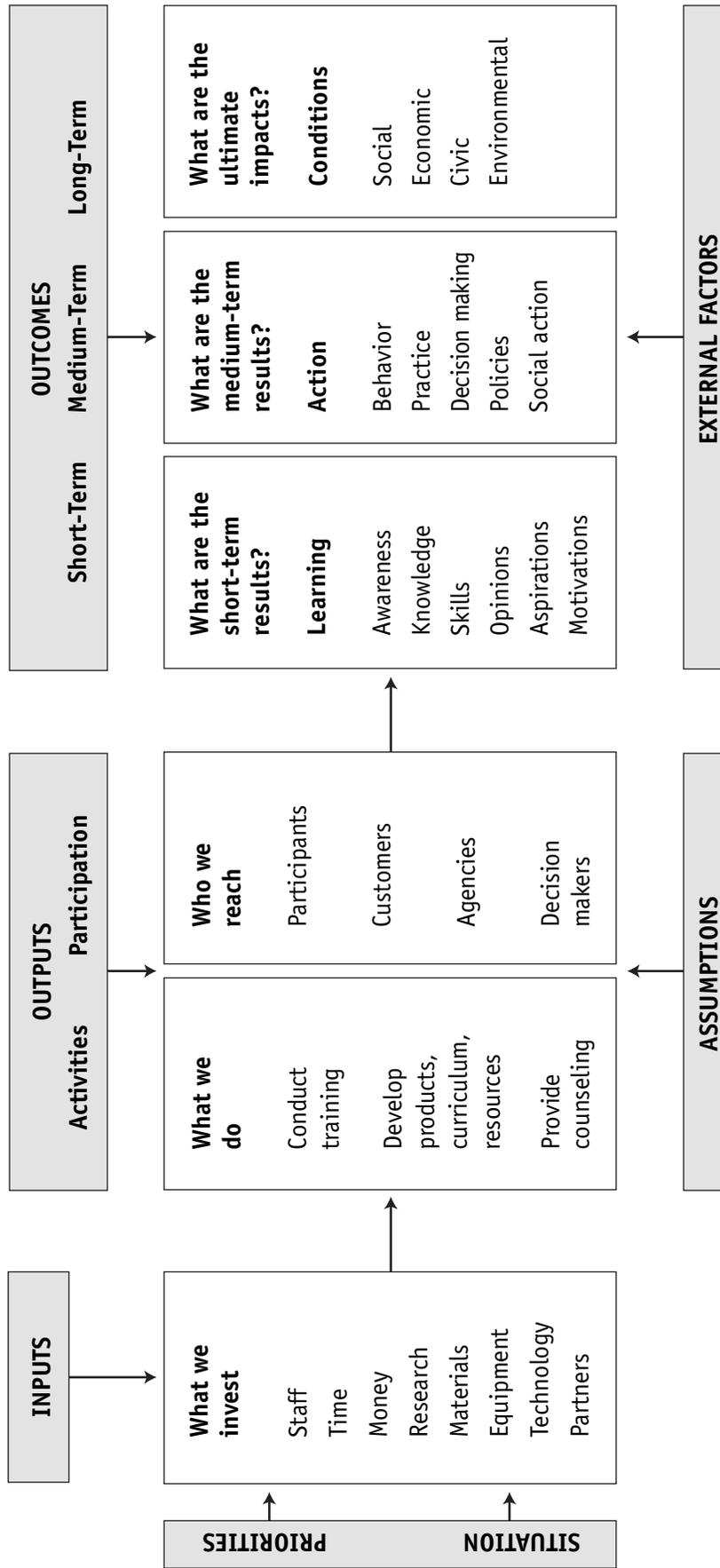


Logic models help identify assumptions underlying the program and possible gaps in program theory. They are especially helpful when a specific theory of how a program is supposed to work does not exist at all. They then identify critical program components and focus the evaluation on important activities, outcomes, and linkages. Logic models can also provide a common understanding of what the program is trying to achieve and how its components fit together. In addition, they can help identify which stakeholders to involve in the evaluation.

*. . . building a logic model is an excellent way to ascertain whether the logical connections do, in fact, exist and to identify where these connections are missing.*

Further, a logic model can be a useful way to demonstrate the connections between the program's objectives and activities to people outside the organization, such as sponsors, policy makers, and the media. Finally, building a logic model is an excellent way to ascertain whether the logical connections do, in fact, exist among program components and to identify where these connections are missing. A generic program logic model is presented in Figure 3, and an example of a driver education program logic model is found in Table 6 on page 67.

**Figure 3. Generic Program Logic Model**

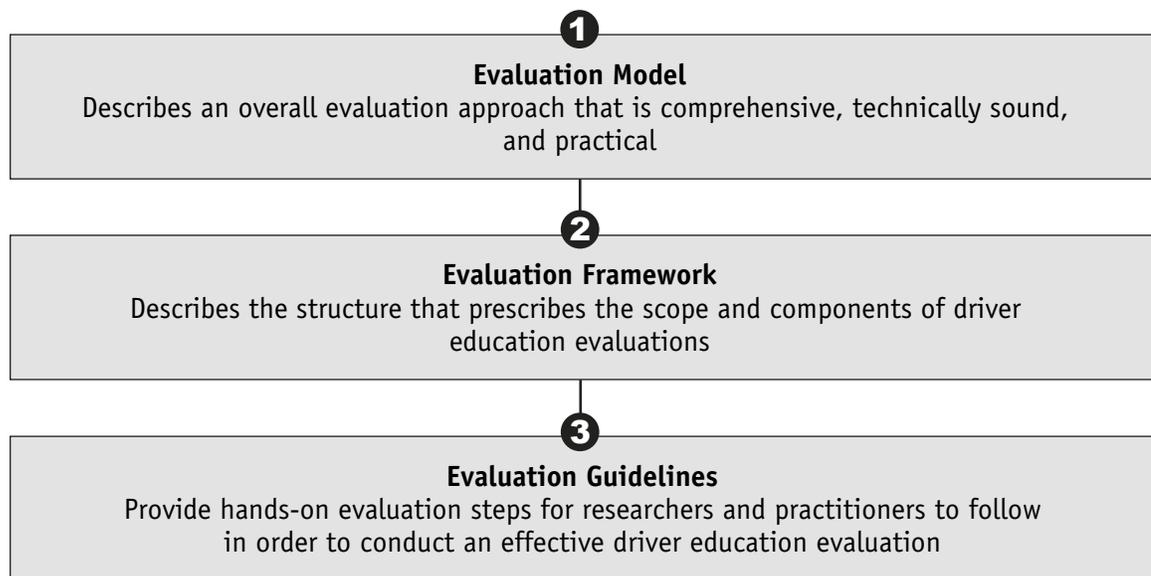


Adapted from *Enhancing Program Performance with Logic Models*, Taylor-Powell, Jones, and Henert 2002.

# Developing an Evaluation Model and Framework for Driver Education

Using the foundations of evaluation concepts, theory, and models, a three-level hierarchy has been developed to provide an overall structure for these *Guidelines*. The hierarchy begins at the most general level with a composite evaluation model that provides a broad scope for evaluating driver education programs. An evaluation framework that flows from the model specifies the scope and components specifically for driver education evaluations. The evaluation guidelines then provide hands-on, step-by-step actions and guidance to carry out an effective driver education evaluation. This hierarchy is illustrated in Figure 4.

**Figure 4. Driver Education Evaluation Hierarchy**



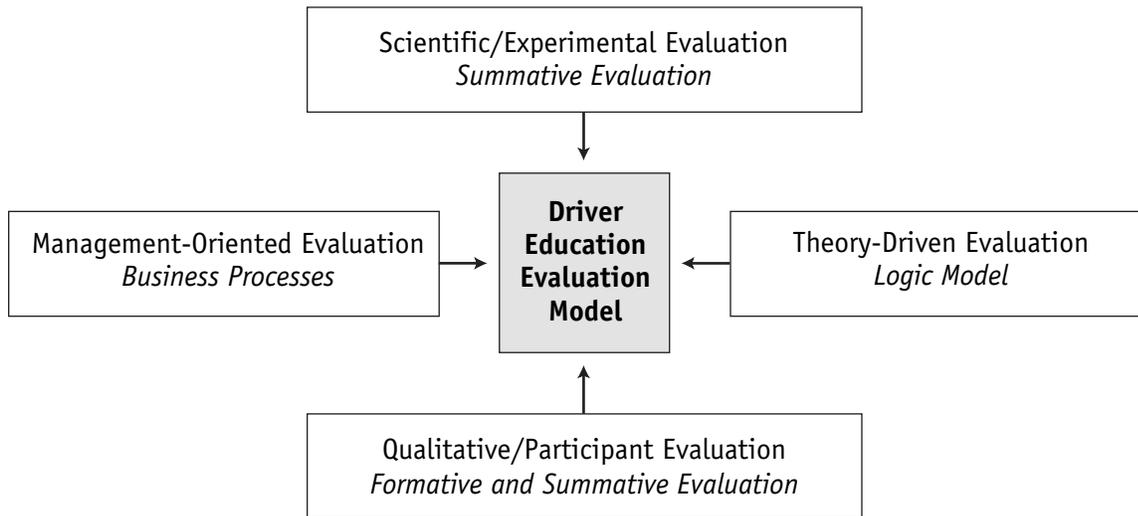
## Driver Education Evaluation Model

A broad, comprehensive evaluation approach is required for driver education, one that realistically incorporates:

- The full scope of driver education program description, logic, context, products, and processes;
- The entire continuum of outcomes and impacts;
- All relevant delivery organizations, including businesses and governments;
- Concerned stakeholders, including consumers, providers, and regulators; and
- A range of applicable evaluation methods and appropriate measures.

The four groups of evaluation models (described on page 37) provide a basis for a driver education evaluation model. As shown in Figure 5, the model incorporates theory-driven evaluation, management-oriented evaluation, scientific/experimental evaluation, and qualitative/participant evaluation into its overall structure.

**Figure 5. Driver Education Evaluation Model**



The rationale for this composite model is as follows. First, theory-driven evaluation has been added to Trochim’s categories of models to ensure that the evaluation is grounded in a theory or logic model, which provides a good idea of why and how the program is expected to achieve its objectives. Intermediate outcomes, such as knowledge, skills, attitudes, intentions, values, or behaviors that can be measured, should be included in the program’s logic model. This has been an area of weakness in driver education evaluations in the past.

Second, there is a need to encompass driver education evaluation within different organizational settings. The composite model recognizes the importance of management and allows for evaluation of organizational structure and processes. Program modification on an ongoing basis is well accepted in modern management and public administration practice. Continuous improvement increases the potential for achieving desired longer-term outcomes and impacts. Quality management is as important to effective driver education as to any other industry or program.

Lastly, to accommodate a broader spectrum of evaluation types, different combinations of experimental, management-oriented, and qualitative evaluations can be implemented

depending on the circumstances of the program. This integrated approach accommodates evaluation needs across program components, organizational levels, and diverse stakeholders.

## Integrated Driver Education Evaluation Framework

The driver education evaluation model provides the basis for a comprehensive, integrated, and adaptable evaluation framework for driver education. The framework sets the boundaries and parameters for the content, format, and utilization of the evaluation guidelines, by:

1. Identifying the scope and context to maximize the beneficial impact of evaluation on the future development of driver education;
2. Identifying the full range of evaluation *targets, methods, data, and success indicators*; and
3. Using the Joint Committee's *Program Evaluation Standards* (pages 37-39) to determine the soundness of every evaluation.

The framework consists of several key components: program areas, evaluation types, evaluation targets and indicators, data sources, and methods. One way to understand each component is to think about the questions that need to be answered and the aspects of the program that need to be evaluated in order to answer them. These are explained next.

*One way to understand each component is to think about the questions that need to be answered and the aspects of the program that need to be evaluated in order to answer them.*



## *What aspects of the program should be evaluated?*

The scope of driver education program evaluation should be as comprehensive as practically possible. There are many opportunities for program improvement by conducting evaluation repeatedly across a program's life cycle. This cycle begins with the program's theoretical or logical basis, and progresses through program development and implementation, to the final reckoning of costs and benefits. The driver education evaluation framework addresses the evaluation needs of the full range of program areas, including:

- Program theory or logic, which explains the theoretical and logical bases of the program.
- Program context, which identifies the political, economic, and social environments that influence the program.
- Program standards, which document program principles, regulations, and governance.
- Program business processes, which document the implementation, operation, and management of the program.
- Program products and processes, which identify the content of instructional materials and delivery methods.
- Program outcomes, which identify the direct educational effects of the program on students.
- Program impacts, which identify the intended and unintended consequences of the program and potential side effects.

## *Which types of evaluation are most appropriate to undertake?*

The range of evaluation types is quite broad. Evaluation scope refers to the entire range of evaluation approaches, including:

- Evaluability assessment—the assessment of the program's theory, logic, and structure to determine whether the program is ready to be evaluated.
- Evaluation planning and preparation—structuring and organizing the evaluation.
- Formative evaluation—the evaluation processes intended to improve program components.
- Summative evaluation—the evaluation processes intended to assess the worth or value of the program.
- Meta-analysis—the systematic combination of the results of separate evaluations.
- Metaevaluation—evaluation intended to improve evaluations.

Table 2 shows the relationship of the evaluation types to program areas. For each area, the table indicates which evaluation types are applicable and can be considered for the evaluation.

**Table 2. Evaluation Types and Applicability to Program Areas**

PROGRAM AREA	EVALUATION TYPES AND PROGRAM AREA APPLICABILITY					
	Evaluability Assessment	Evaluation Planning and Preparation	Formative Evaluation	Summative Evaluation	Meta-Analysis	Meta-evaluation
Program Logic	x	x				x
Program Context	x	x	x			x
Business Processes	x	x	x			x
Program Standards	x	x	x	x	x	x
Instructional Products	x	x	x			x
Instructional Processes	x	x	x			x
Student Outcomes	x	x	x	x	x	x
Social Impacts	x	x		x	x	x

 *What are the evaluation targets and related indicators of success?*

There are many specific evaluation targets or aspects of a program that can be evaluated and numerous related success indicators or measures of effectiveness. Table 3 indicates the range of possible targets for driver education evaluation along with examples of the types of indicators that can be used to determine their effectiveness.

Learning outcomes, for example, specify what students will know or be able to do as a result of taking driver education, usually expressed as knowledge, skills, or attitudes. Mobility outcomes are the results that students will experience, as a result of passing the license test, that will provide the opportunity to drive and be independently mobile. Behavioral outcomes are the program results related to students' driving behavior, such as amount and type of driving, and violation and crash experiences.

**Table 3. Examples of Evaluation Targets and Success Indicators**

Evaluation Targets	Examples of General Success Indicators
User needs	Needs assessed; Needs linked to objectives
Program logic or theory	Program content and structure linked to needs and objectives via logic model
Evaluability	Program readiness for evaluation; Existence of a clear logic model
Stakeholder expectations	Expectations for program and evaluation addressed
Operations management	Operations meet efficiency standards
Quality control	Organizational quality standards met
Customer service	Standards met; Complaints addressed
Sustainability	Revenues increasing; Political, economic, and consumer support; Industry growth
Benchmarking and certification	Applicable benchmarks met; Certifiability requirements met
Transportability of program	Expansion potential determined
Instructional products	Learning efficacy; User friendliness; Up to date
Measurement instruments	Reliable; Valid; Practical to implement and assess
Instructional processes	Defensible educational delivery methods; Training and retraining standards met
Learning outcomes	Knowledge gain and skill acquisition meet targets
Mobility outcomes	Number of students obtaining license and driving increasing
Behavioral outcomes	Exposure to risk limited
Safety impacts	Crash and casualty rates decreasing due to training
Socioeconomic impacts	Cost savings; Employment and other economic effects acceptable
Evaluation effectiveness	Program Evaluation Standards met; Evaluation findings useful and relevant

A detailed discussion of evaluation targets and success indicators is found in Steps 1B and 3A.

## Which types of evaluation methods should be used?

Integrated into the driver education evaluation model and framework is the use, as appropriate, of a range of evaluation methods including:

1. Qualitative evaluation methods to gain in-depth understanding of client needs, expectations, and feedback, and to provide an assessment opportunity to stakeholders, staff, and consumers; and
2. Quantitative methods used in experimental and quasi-experimental evaluations, record studies, surveys, and economics-based studies to examine program outcomes and impacts.

A wide range of data-gathering and research methods can be used to address evaluation questions, including:

- Problem definition methods, such as review of research, needs assessment, task analysis, and review of epidemiological findings
- Solution definition methods, such as policy analyses, logic models, and expert panels
- Program content assessment methods, such as pilot testing and content analysis
- Qualitative research methods, such as focus groups, interviews, diaries, and participant observation
- Standardization methods, such as benchmarking, certification, and auditing
- Systematic testing, instrumented vehicles, and video observation
- Surveys using questionnaires, or log books
- Record studies and modeling
- Ecological studies
- Longitudinal studies
- Quasi-experiments
- Randomized controlled experiments
- Cost/benefit and cost effectiveness analyses
- Feasibility analysis and business planning
- Policy analysis
- Meta-analysis to systematically review diverse program effects
- Metaevaluation using, for example, methodology analysis, or peer review processes to assess evaluations

These methods are discussed in Chapter 3, Step 2, and defined in the Glossary in Appendix B.

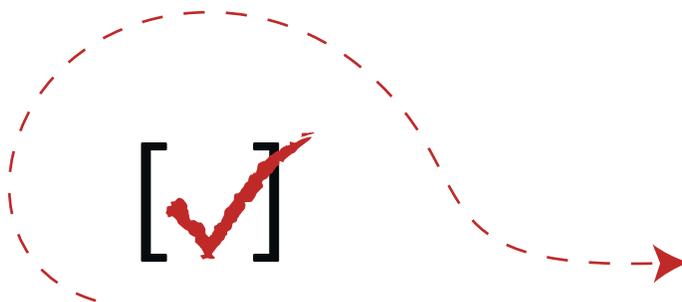
## *Where will the data be found?*

Data can be obtained from a variety of sources, depending on which methods are being used. The range of data sources for driver education evaluations includes:

- Students
- Parents
- Staff
- Management
- Stakeholders
- Experts on new drivers
- Educational research experts
- Evaluators
- Program developers
- Research on new drivers
- Program records
- Program documents
- Lawyers
- Sponsors
- Partners
- Franchisers
- Regulators
- Licensing authorities
- Insurers

Table 4 presents a summary of the key components of the driver education evaluation framework. It provides a general overview of the full range of possible evaluation activities, related targets, data sources, and methods. The table can be used as a reference to guide the evaluation process at any time during an evaluation's development and implementation. Many of the terms have been defined previously or can be found in Appendix B.

Specific procedures a driver education evaluation team can follow to develop and implement an ongoing evaluation process are detailed in the next chapter. Afterward, two hypothetical driver education evaluation scenarios are presented, one for a basic program evaluation, and the second for an advanced evaluation.



**Table 4. Integrated Driver Education Evaluation Framework**

Evaluation Type	Program Area	Evaluation Targets	Success Indicators	Data Sources	Methods	
Formative Evaluation	Program Logic	User needs	<ul style="list-style-type: none"> <li>Needs linked to objectives</li> </ul>	Students, parents, stakeholders, research on new drivers, experts on new drivers	<ul style="list-style-type: none"> <li>Document review</li> <li>Literature review</li> <li>Logic model development</li> <li>Meta-analysis</li> <li>Needs assessment</li> <li>Problem definition research</li> <li>Program content analysis</li> </ul>	
		Program logic model or theory	<ul style="list-style-type: none"> <li>Objectives linked to content</li> </ul>	Educational research experts, program developers, staff, program records and documents		
		Evaluability	<ul style="list-style-type: none"> <li>Program ready to be evaluated</li> </ul>	Evaluation experts, program developers, staff, stakeholders		
	Program Context	Stakeholder expectations	Stakeholders			<ul style="list-style-type: none"> <li>Compliance audit</li> <li>Compliance checklist</li> <li>Legal review</li> <li>Policy review</li> <li>Stakeholder analysis</li> <li>Stakeholder survey</li> </ul>
		Regulatory environment	<ul style="list-style-type: none"> <li>Constraints and obligations met</li> </ul>	Management, program records, governments		
		Contractual environment	<ul style="list-style-type: none"> <li>Obligations to sponsors, partners, franchisers clarified and met</li> </ul>	Management, lawyers, sponsors, partners, franchisers		
	Business Processes	Operations management	<ul style="list-style-type: none"> <li>Staff retained</li> <li>Staff motivated with few concerns</li> <li>Operations efficient</li> </ul>	Program records, management, staff, instructors, regulators		<ul style="list-style-type: none"> <li>Accounting audit</li> <li>Case study</li> <li>Customer satisfaction focus groups</li> <li>Customer satisfaction interviews</li> <li>Customer satisfaction survey</li> <li>Marketing analysis</li> <li>Quality control analysis</li> <li>Staff interviews</li> </ul>
		Quality control	<ul style="list-style-type: none"> <li>Quality standards met</li> <li>Program consistent</li> </ul>	Students, parents, instructors, managers, regulators		
		Marketing	<ul style="list-style-type: none"> <li>Program reach optimized</li> <li>Participation rates stable or growing</li> </ul>	Management, staff, instructors, students, parents		
		Customer service	<ul style="list-style-type: none"> <li>Standards met</li> <li>Complaints addressed</li> </ul>	Students, parents, staff, instructors, managers, regulators		
		Sustainability	<ul style="list-style-type: none"> <li>Positive cash flow, revenues increasing</li> <li>Growth acceptable</li> <li>Political and customer support</li> </ul>	Managers, regulators		

Evaluation Type	Program Area	Evaluation Targets	Success Indicators	Data Sources	Methods
Formative Evaluation (continued)	Program Standards	<ul style="list-style-type: none"> <li>Benchmarking and certification</li> <li>Transportability of program</li> </ul>	<ul style="list-style-type: none"> <li>Applicable benchmarks met</li> <li>Certifiability requirements met</li> <li>Potential for expansion in other jurisdictions established</li> </ul>	<ul style="list-style-type: none"> <li>Managers, regulators, industry members, jurisdiction representatives</li> <li>Managers, regulators</li> </ul>	<ul style="list-style-type: none"> <li>Benchmarking</li> <li>Certification</li> <li>Industry survey</li> <li>Jurisdiction survey</li> </ul>
	Instructional Products	<ul style="list-style-type: none"> <li>Curriculum materials</li> <li>Tests and measurement</li> </ul>	<ul style="list-style-type: none"> <li>Efficacious</li> <li>Up to date</li> <li>User friendly</li> <li>Reliable</li> <li>Valid</li> <li>Practical to implement and assess</li> </ul>	<ul style="list-style-type: none"> <li>Students, parents, instructors, managers, regulators</li> <li>Students, comparison groups, parents, instructors, managers, regulators</li> </ul>	<ul style="list-style-type: none"> <li>Content analysis</li> <li>Focus groups</li> <li>Instructor log books</li> <li>Instructor survey</li> <li>Interviews</li> <li>Pilot testing</li> <li>Student survey</li> </ul>
	Instructional Processes	<ul style="list-style-type: none"> <li>Instructor preparation</li> <li>Curriculum delivery and in-car practice</li> <li>Instructional facilities</li> </ul>	<ul style="list-style-type: none"> <li>Benchmarks met</li> <li>Training quality control standards met</li> <li>Re-training standards met</li> <li>Educational standards met</li> <li>Course completion and pass rates met</li> <li>Customer satisfaction</li> <li>Staff satisfaction</li> <li>Safety standards met</li> </ul>	<ul style="list-style-type: none"> <li>Instructors, managers, regulators</li> <li>Students, parents, instructors, managers, regulators</li> <li>Students, parents, instructors, managers, regulators</li> </ul>	<ul style="list-style-type: none"> <li>Benchmarking</li> <li>Case study</li> <li>Instructor log books</li> <li>Instructor survey</li> <li>Observation</li> <li>Parent survey</li> <li>Pilot testing</li> <li>Quality control analysis</li> <li>Records analysis</li> <li>Student and parent focus groups</li> <li>Student and parent interviews</li> <li>Student survey</li> </ul>

Evaluation Type	Program Area	Evaluation Targets	Success Indicators	Data Sources	Methods	
Summative Evaluation	Student Outcomes	Knowledge outcomes	<ul style="list-style-type: none"> <li>Knowledge gain targets met</li> </ul>	Students, parents, instructors, licensing authorities	<ul style="list-style-type: none"> <li>Case study</li> <li>Focus groups</li> <li>Instructor log books</li> <li>Instrumented vehicles</li> <li>Student diaries</li> <li>Student exposure log books</li> <li>Student survey</li> <li>Testing</li> <li>Video observation</li> </ul>	
		Skill outcomes	<ul style="list-style-type: none"> <li>Skill acquisition targets met</li> </ul>	Students, parents, instructors, licensing authorities		
		Motivation outcomes	<ul style="list-style-type: none"> <li>Attitude change targets met</li> </ul>	Students, parents		
		Mobility outcomes	<ul style="list-style-type: none"> <li>Targets for course pass and license rates met</li> </ul>	Licensing authorities, students, parents		
		Behavioral outcomes	<ul style="list-style-type: none"> <li>Exposure to risk limited</li> </ul>	Students, licensing authorities		
		Social Impacts	<ul style="list-style-type: none"> <li>Crash reduction impacts</li> <li>Injury reduction impacts</li> <li>Socioeconomic impacts</li> </ul>	<ul style="list-style-type: none"> <li>Targets for crash rates met</li> <li>Targets for casualty rates met</li> <li>Cost/benefit effects positive</li> <li>Other economic effects positive</li> <li>Side effects minimized</li> </ul>		<ul style="list-style-type: none"> <li>Licensing authorities, students, parents, insurers</li> <li>Licensing authorities, students, parents, insurers</li> <li>Licensing authorities, insurers, stakeholders</li> </ul>
	Meta-evaluation	Evaluation Quality	Evaluation effectiveness	<ul style="list-style-type: none"> <li>Program Evaluation Standards met</li> <li>Evaluation results useful and relevant</li> <li>Stakeholder expectations met</li> </ul>	Evaluators, managers, staff, stakeholders	<ul style="list-style-type: none"> <li>Benchmarking against Program Evaluation Standards</li> <li>Evaluation review for technical quality and usefulness of findings</li> <li>Meta-analysis</li> <li>Peer review</li> <li>Stakeholder interviews</li> </ul>





# 3. Driver Education Evaluation Guidelines

The detailed evaluation guidance presented in this chapter can be used by driver education program managers, administrators, and evaluators to design and implement effective evaluations. This information can also help illustrate evaluation requirements and aid teams who want to hire outside evaluators to assist with their evaluations.

## Applying the Framework—How to Use Evaluation Effectively

The evaluation framework provides the overall structure within which to begin to think about evaluating a driver education program. It can also help guide decisions about which evaluation types and methods are most appropriate for your program evaluation.

As you can see from the framework, the scope of driver education evaluation is fairly complex. To ensure that the framework is practical, evaluation should be included at every stage of the program's life cycle. This ranges from needs assessment and conceptualization; developing program materials and procedures; and continuing as the program is pilot tested, implemented, and proven in the field. The program will then be upgraded and eventually superseded or retired. A single evaluation of a program is rarely enough. Evaluation should be an ongoing process and an integral part of developing, implementing, and managing a driver education program. Evaluation thus becomes the impetus for continuous program improvement.

## The Evaluation Team

The evaluation team is the group of people responsible for developing and implementing the evaluation. There are a variety of people who can participate, and their roles on the evaluation team can vary.

Driver education program managers and staff, including branch and operations personnel, and classroom and in-car instructors, are key people in the evaluation. Other

stakeholders in the community might also like to help in an advisory role. These include school board and high school representatives, youth groups, AAA and CAA clubs, enforcement officials, public health officials and practitioners, and injury prevention researchers and association representatives.

If no evaluation expertise is available in-house, it is also necessary to consider hiring an outside evaluation specialist to assist with some of the more complex types of evaluation. For all evaluations, but particularly higher level ones, including outside experts, who have no vested interest in the program being evaluated, may also help raise the credibility of the evaluation.

The *Guidelines* will help the evaluation team identify which parts of the evaluation can be managed in-house and which require outside expertise. They also provide basic evaluation knowledge to assist a program manager make good decisions about the type of evaluator to bring on board. They help identify the questions to ask and resources to look for. Detailed information on when and how to hire an external evaluator is provided in Appendix J.

## Guiding Questions for Driver Education Program Evaluation

### Key Questions

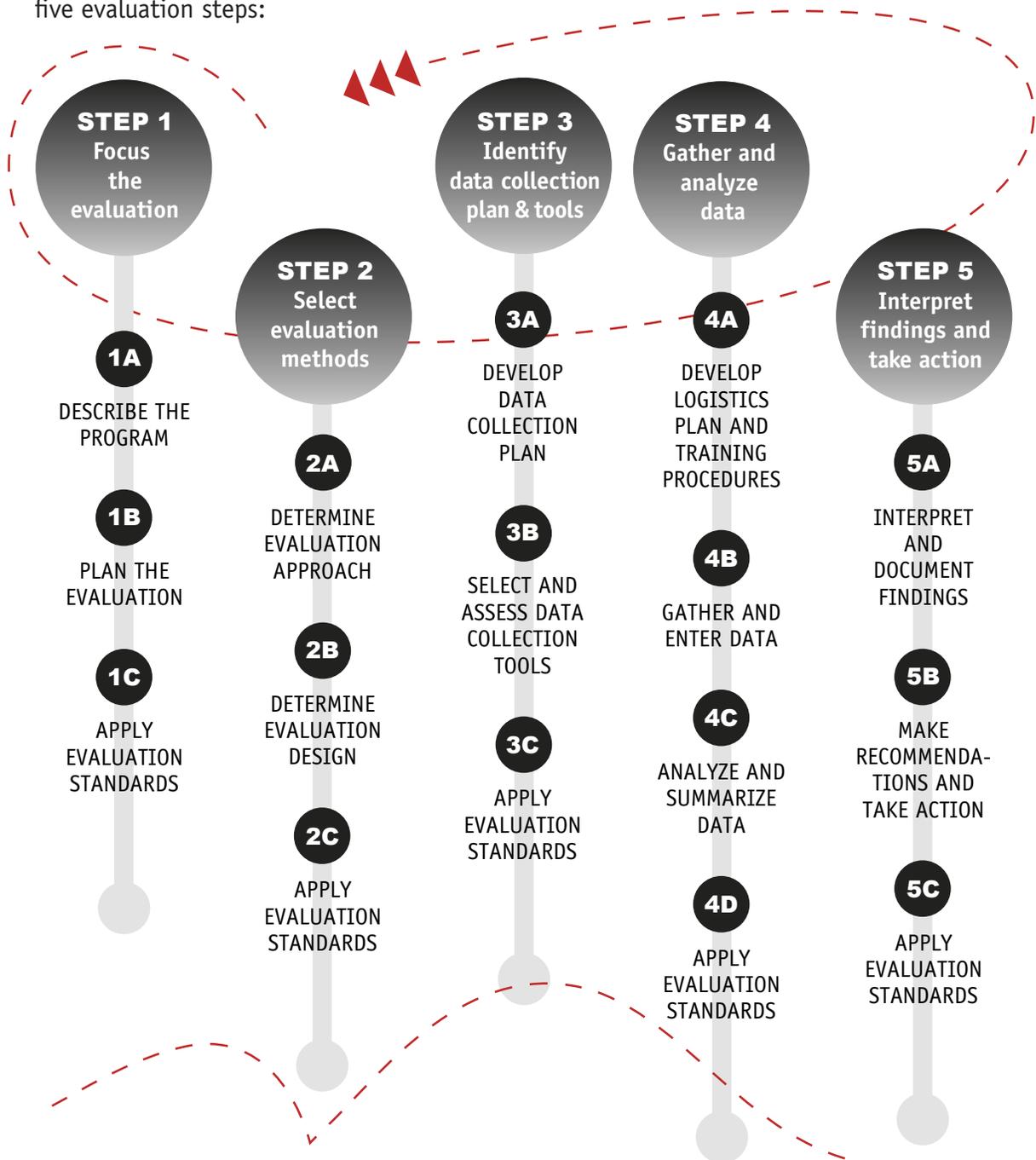
The evaluation framework helps to identify the key questions that an evaluation must answer. These are:

1. What exactly is being evaluated, and what type of evaluation is being planned?
2. What are the methods that will be used in the evaluation, and how will they be used?
3. What tools will be used to gather the evaluation information?
4. How will this information be gathered and analyzed?
5. How will the evaluation findings be interpreted, and how will they be disseminated and acted upon to ensure continuous improvement?

These questions lead to organizing the evaluation process into five major steps, which are described in detail next.

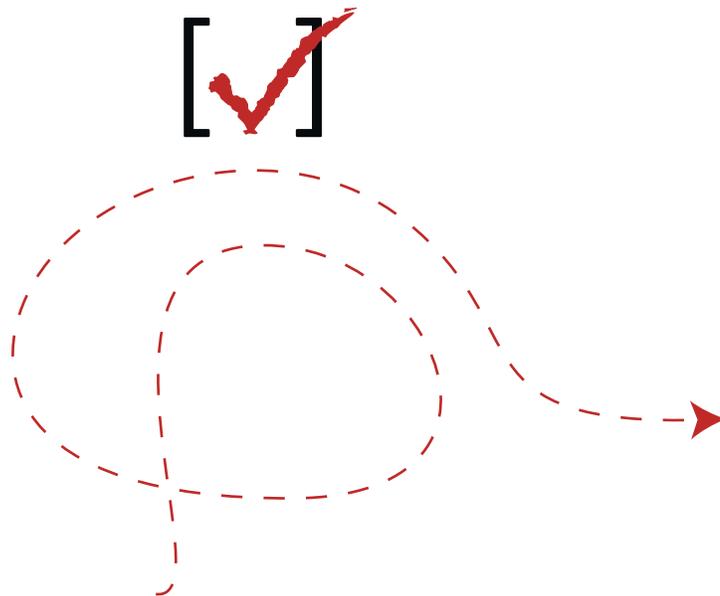
## Steps for Evaluating Driver Education Programs

This section provides a roadmap for those interested in conducting an evaluation of a driver education program for the first time, or as an expanded, more comprehensive evaluation of a program that has been previously evaluated. The steps identify the most important activities to consider as the evaluation is being planned and implemented. They can be modified or adapted to the unique circumstances of any program. They can also assist those who have little evaluation expertise or experience in consulting with professional evaluators and participating in evaluations of their own. The *Guidelines* are organized into five evaluation steps:



Even though these steps are ordered in a logical sequence, the order can be modified to fit the specific circumstances of individual programs. Each step begins with a brief description and explanation of why it is important. The activities required to complete the step are summarized in a chart and then explained in detail. Applying the relevant Program Evaluation Standards completes each step. An overview of the five steps and their related activities is found in Appendix D.

As you read through these steps, the detail may make applying the steps look harder than it will be once work begins on your actual evaluation. Just take it step by step and the process will fall into place.



# STEP 1

## FOCUS THE EVALUATION

Step 1 explains the initial detailed documentation and planning that must be undertaken to focus the evaluation. Giving careful attention to actual program details and their description is important, as well as considering key decisions that the evaluation team are charged with. These include determining the expectations of the evaluation, identifying evaluation targets, and deciding who will use the evaluation results.

As the activity chart indicates, Step 1 has three major tasks—describing the program, planning the evaluation, and applying the evaluation standards.

### Summary of Activities

<b>1A</b> DESCRIBE THE PROGRAM	<b>1B</b> PLAN THE EVALUATION	<b>1C</b> APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify stakeholders, and user and program needs</i></li> <li>➤ <i>Identify the program’s vision, goals, and objectives</i></li> <li>➤ <i>Identify and document program activities, resources, and context</i></li> <li>➤ <i>Develop a program logic model</i></li> <li>➤ <i>Assess program readiness to be evaluated</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify the purpose of the evaluation</i></li> <li>➤ <i>Identify knowledge from driver education evaluations</i></li> <li>➤ <i>Identify potential users and uses of the evaluation</i></li> <li>➤ <i>Identify key evaluation questions and targets</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## STEP 1A DESCRIBE THE PROGRAM

### Overview

Establishing a common understanding of the program’s goals, objectives, activities and outcomes at the very beginning of an evaluation process is essential. Describing the

program sets the framework for planning and executing the evaluation. The following chart summarizes the key activities in Step 1A.

<b>1A</b> <b>DESCRIBE THE PROGRAM</b>	<b>1B. PLAN THE EVALUATION</b>	<b>1C. APPLY EVALUATION STANDARDS</b>
<ul style="list-style-type: none"> <li>➤ <i>Identify stakeholders, and user and program needs</i></li> <li>➤ <i>Identify the program’s vision, goals, and objectives</i></li> <li>➤ <i>Identify and document program activities, resources, and context</i></li> <li>➤ <i>Develop a program logic model</i></li> <li>➤ <i>Assess program readiness to be evaluated</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify the purpose of the evaluation</i></li> <li>➤ <i>Identify knowledge from driver education evaluations</i></li> <li>➤ <i>Identify potential users and uses of the evaluation</i></li> <li>➤ <i>Identify key evaluation questions and targets</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

The program description summarizes the program, and explains its goals and how it accomplishes those goals. Look at who is involved in the program and who is affected. Identify user and program needs, and look at program effects and resources. If a logic model for your program doesn’t exist, create one. This will help organize program information and demonstrate the relationships between its various components. Also consider the social, political, and economic environments within which the program operates, and specify the program’s stage of development.

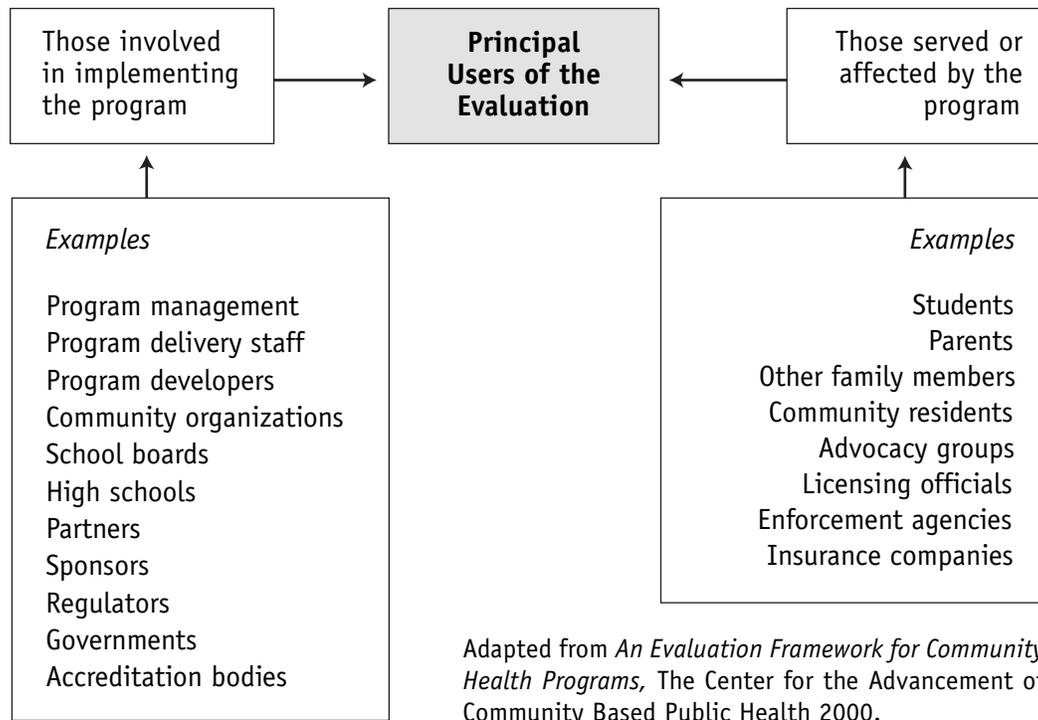
## Activities

### ➤ *Identify Stakeholders, and User and Program Needs*

It is important to know who the program’s stakeholders are. They are the individuals and groups of people who are interested in the program for varying reasons. They include the users of the program or the client group, such as driver education students, and their parents. Also included are others in the community who have an interest in the program, such as school boards, schools, insurance companies, community agencies, and potential partners or sponsors. Stakeholders can be divided into three groups: 1) those involved in implementing the program; 2) those served or affected by the program; 3) and the primary users of the evaluation.

As shown in Figure 6, the evaluation user group is typically made up of individuals from the other two groups.

**Figure 6. Types of Stakeholders**



Documenting the needs of program users and other stakeholders is also important. This ensures that program managers and evaluators are fully aware of needs currently and as they change over time. This process is known as needs assessment. Student and parent needs can be monitored through regular feedback requests incorporated into program delivery activities. A short feedback form or group discussion, for example, can be included as part of every course so that students routinely have opportunities to provide input. A similar form can be sent home or distributed at a parent meeting. Consider offering an incentive, such as a prize or discount coupon, to increase the response rate.

Student needs can also be identified through the large body of problem definition research on young novice drivers. Driver education evaluators should have a working knowledge of the needs of new drivers that driver education programs attempt to address.

Benchmark standards for program content related to student needs can also provide guidance for such an analysis. Standards are available from the American Driver and Traffic Safety Education Association (ADTSEA) (Highway Safety Center 2002), the Driving School Association of the Americas (DSAA) (Road Safety Educators' Association 2005), and the

National Institute for Driver Behavior (NIDB). Appendix I provides the DSAA standards and the web addresses for the ADTSEA and NIDB standards.

Standards are also referred to in other guideline tasks where checking your program against industry benchmarks is important (e.g., Step 1B, Evaluation Targets; Step 2A, Evaluation Approach; Step 3A, Data Collection Plan; and Step 3B, Data Collection Tools). It is important to recognize that standards, while reflecting expertise on program quality, do not necessarily identify programs that produce measurable safety impacts—these connections are yet to be established.

Other stakeholders may also provide important input. One way to obtain this information is to contact stakeholders either individually or in groups, and carry out face-to-face consultations to identify their needs related to the program.

Ascertaining the program's needs and identifying what additional information is required to accurately describe the program are also essential. Some types of program information may not be documented or tracked. Processes may need to be put in place to gather baseline program information as part of this initial activity; for example, knowing how students and parents learn about the program, what they found attractive about it, and why they decided to take your course will help describe and evaluate program marketing and promotion activities. Unless this information is routinely obtained from students while they are attending the course, it is not easy to obtain. Thinking about program needs helps identify information gaps and suggests actions to fill them.

### ➤ *Identify the Program's Vision, Goals, and Objectives*

Being clear about what the program management, staff, and stakeholders expect the effects of the program to be, as well as what is already known about the program, are also important. What are the expected outcomes for students who attend this program? What are the expected results for other stakeholders? This analysis will lead to identifying and clarifying the program's vision, goals, and objectives. Understanding fundamental aspects of the program will help determine the evaluation's purpose and expectations.

### ➤ *Identify and Document Program Activities, Resources, and Context*

The program's activities should flow directly from its goals and objectives. Consider what is being done currently and what could be done differently as program activities are documented. It may be possible to identify weak program content, activities, and delivery

processes at this stage, and these can be corrected before undertaking further evaluation.

Table 5 provides an example of a worksheet to help organize program information. In this case, a hypothetical program’s overall goal is increasing the safety of their students. A program objective of improving students’ performance in hazard perception has been agreed upon. The expectations, activities, and resources required to meet this objective have been documented as part of the basis for developing a logic model. This worksheet helps an evaluation team specify information related to each program objective. A blank worksheet is provided in Appendix E.

**Table 5. Organizing Program Information**

<i>Program Goal:</i> <b>Increase safety</b>		
<i>Objective:</i> <b>Improve students’ performance capability in hazard perception</b>		
<b>Expectations</b>	<b>Activities</b>	<b>Resources</b>
<b>Classroom Instruction</b>		
<ul style="list-style-type: none"> <li>• Lesson on hazard perception and avoidance in classroom curriculum will be up-to-date and accurate.</li> <li>• Students will be able to demonstrate knowledge about potential hazards and how to recognize them.</li> </ul>	<ul style="list-style-type: none"> <li>• Two-hour classroom session that includes video, class discussions and a role play with feedback</li> <li>• Reading assignments</li> </ul>	<ul style="list-style-type: none"> <li>• Textbook, videos</li> <li>• Worksheets</li> <li>• Role-play outline</li> </ul>
<b>In-Car Instruction</b>		
<ul style="list-style-type: none"> <li>• Hazard perception and avoidance training will be included in in-car lessons.</li> <li>• Students will be able to demonstrate hazard perception and avoidance quickly and consistently in driving practice.</li> </ul>	<ul style="list-style-type: none"> <li>• In-car training and practice in hazard perception and training</li> <li>• Commentary driving</li> <li>• Simulator practice</li> </ul>	<ul style="list-style-type: none"> <li>• Driving instructors trained in teaching hazard perception and avoidance</li> <li>• Driving instructors trained in commentary driving</li> <li>• Parents trained as coaches to reinforce hazard perception skills during practice driving sessions</li> <li>• Log books</li> <li>• Vehicles</li> <li>• Instructor mirrors</li> <li>• Simulators</li> </ul>

Adapted from  
*An Evaluation Framework for Community Health Programs*  
 The Center for Community Based Public Health 2000.

Pay particular attention to creating specific and measurable objectives. This is not always easy, but try to think of the goals and objectives in terms of the operations you would use to measure them. A knowledge objective, for example, would be thought of as the testing that measures knowledge. Clear and measurable objectives help ensure that the evaluation activities are also clear, specific, and focused on the most appropriate aspects of the program.

Assessing the resources available for the evaluation is critical to undertake early in the planning process. Otherwise there is a chance that the evaluation will be initiated but not completed because time or money run out. Consider the following at this point:

- How much money can be allocated to the evaluation?
- How much time and how many people can be assigned?
- What is the time frame for the evaluation? Is it limited by resources?
- Who can participate in the evaluation, and what skills do they bring to it?
- Are resources available from a partner or sponsor?
- Are the required supplies, equipment, and space, such as stationery, postage, audio/ videotapes, computers, software, photocopiers, phones, and meeting rooms available or able to be purchased?
- Are volunteers available to help?
- Is there a large enough sample of students and comparison drivers to reliably detect the size of program effect that is expected?
- Does the program have the resources to act upon the evaluation results?

The following list outlines the types of capacities and skills that most driver education program evaluations will require:

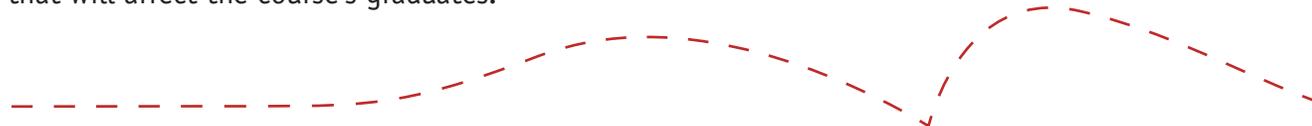
- Evaluation team leadership, management, team membership
- Literature review and secondary research
- Program evaluation design
- Data collection training and supervision
- Data collection
- Data entry
- Data analysis
- Possibly statistical expertise
- Report writing

Staff can be involved in almost all of these functions, except perhaps the more complex evaluation design and statistical analysis.

The final resource requirements for the evaluation depend on the selection of evaluation type and methods, and a more detailed final costing of the evaluation should take place after Step 2. A resources worksheet is found in Appendix E.

Describing the context within which the program operates ensures that the influences of external environments are also considered at the outset. It is important to understand what the existing and potential influences are, as well as how they impact the program. Examples include the program's history and current political environment. Also potentially relevant are geographical setting and social and economic conditions. Driver licensing changes, for example, may be taking place or planned that will affect the course's graduates.

The program's stage of development is also important to take into account. The development and implementation of an evaluation of a new program will be quite different from that of a program that has existed for many years. The initial evaluation of a program that is in the process of being developed will focus primarily on an evaluation of program processes and products (i.e., formative evaluation). Perhaps the development of the program will be far enough along to collect baseline data and evaluate the program on a pilot basis. An established and well-developed program will have existed long enough to warrant conducting evaluation of student and program outcomes (summative evaluation) as well as further cycles of formative evaluation. Formative and summative evaluations are described in detail in Step 2A on pages 78-80.



**D***escribing the context within which the program operates ensures that the influences of external environments are also considered at the outset.*

In most cases, teams will be evaluating existing driver education programs. Thus the full range of evaluation options can be considered as they proceed to plan the evaluation and determine its approach and methods.

➤ *Develop a Program Logic Model*

As explained in Chapter 2, a logic model is a useful way to organize and depict all the critical components of a program and their relationships. Where you start to create a logic model depends on the program's stage of development. An existing program will use a top-down approach, starting with goals and objectives, and working through activities, to outcomes and impacts. The team will ask "What is the program doing, and why do we think it will create the planned-for change?" The logic model for a new program will more likely take a bottom-up approach. In this case, the team will ask "What change is the program expected to bring about and what should it be doing to achieve it?" This means starting with the expected program impacts, and working backward to activities, and then goals and objectives.

To create a logic model, first determine what the program is trying to achieve. Then link the program's goals and objectives to the available resources, to the program's activities and products, and finally to the program's expected outcomes and impacts. Use all available information about your program to develop the logic model. A document review will help you collect all the important program information. Look at work plans, strategic and operational plans, manuals, training materials, organization charts, budgets, statements of goals and objectives, and any previous evaluation reports. At this point, consulting with some

**T***o create a logic model, first determine what the program is trying to achieve. Then link the program's goals and objectives to the available resources, to the program's activities and products, and finally to the program's expected outcomes and impacts.*

of the program's stakeholders may also be helpful to ensure that, from their perspectives, nothing critical has been omitted. But remember that the logic model shouldn't be too detailed or burdensome and ideally will fit on one or two pages.

Logic models can take many forms. Refer back to the generic logic model on page 42, and look at the example of a driver education program logic model in Table 6. This chart provides examples of the types of information that can be included in a logic model for a driver education program. Blank worksheets for developing a logic model are found in Appendix E.

**Table 6. Example of a Driver Education Program Logic Model**

Program Goals and Objectives	Program Processes and Activities	Outcomes and Impacts	Target Groups
<b>Goal: PROGRAM VIABILITY</b>			
<i>Objective:</i> Economic competitiveness	Operations management	Efficiency	Management, students, parents
		Financial control	Management
	Quality control	Documented quality	Management, students, parents
	Marketing	Program sales	Management, students, parents
	Customer service	Customer satisfaction	Management, students, parents
	Government relations	Regulatory compliance	Management
<b>Goal: DRIVER MOBILITY</b>			
<i>Objective:</i> Starting independent driving career	Classroom teaching	Basic knowledge	Students
	In-car practice	Basic skill	Students
		Student and parent confidence	Students, parents
<b>Goal: DRIVER SAFETY</b>			
<i>Objectives:</i> Performance capability	Knowledge teaching	Rules	Students
		Expectations	
	Skills training	Vehicle handling	Students
		Attention control	
		Hazard perception	
	Risk appreciation		
Driving choices and results	Insight training	On-road performance	Students
	Practice and habit formation	Crash reduction	Students

➤ *Assess Program Readiness to be Evaluated*

Assessing whether the program is ready to be evaluated is referred to as evaluability assessment. It is important to be sure that everything is in place before time and resources are spent even on the initial stages of a program evaluation.

In some cases, although the program is established and operating, there may not be enough documented information to immediately plan an evaluation. The information required to adequately describe the program may not be available and organized so that it is accessible and usable. This information should be obtained prior to evaluation planning. You may need to develop a logic model to accurately understand and describe the program’s goals, objectives, and expected outcomes.

Another factor to consider is the resources available to undertake or support an evaluation. When carefully assessed, resources may not be adequate; therefore, planning is needed to budget the required time and money for an evaluation some time in the future. Also, if the capability to act upon the evaluation results is not adequate, plan to correct this either prior to or in conjunction with the next task, evaluation planning.

## STEP **1B** PLAN THE EVALUATION

### Overview

Once the program is well defined and documented, it is time to focus more specifically on identifying additional information needed to plan an effective evaluation. Carefully think through this second part of Step 1 to ensure that the evaluation can meet its objectives, and that expectations about what can be achieved are realistic. Careful attention here also ensures that the evaluation experience is a positive one that all will be willing to incorporate into program activities on an ongoing basis. The chart below identifies the activities required to plan the evaluation.

1A. DESCRIBE THE PROGRAM	<b>1B</b> PLAN THE EVALUATION	1C. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify stakeholders, and user and program needs</i></li> <li>➤ <i>Identify the program’s vision, goals, and objectives</i></li> <li>➤ <i>Identify and document program activities, resources, and context</i></li> <li>➤ <i>Develop a program logic model</i></li> <li>➤ <i>Assess program readiness to be evaluated</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify the purpose of the evaluation</i></li> <li>➤ <i>Identify knowledge from driver education evaluations</i></li> <li>➤ <i>Identify potential users and uses of the evaluation</i></li> <li>➤ <i>Identify key evaluation questions and targets</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Identify the Purpose of the Evaluation*

It is important to clearly articulate and document the evaluation's purpose and key goals. What is the overall intent of the evaluation? In general terms, what do you want to see accomplished as a result of the evaluation? Some examples are to:

- Assess whether program content is logically linked to the program's objectives.
- Determine customer satisfaction levels.
- Identify areas for program improvement.
- Gain new knowledge about the outcomes of program activities on students' driving.
- Determine the program effects on crash rates of graduates.

These broad goals will be translated into specific program evaluation targets as the evaluation definition and planning activities proceed.

### ➤ *Identify Knowledge from Previous Driver Education Evaluations*

If the program or parts of it have been evaluated in the past, assess the information from these evaluations. Guidance can also be found, of course, in previous evaluations of other driver education programs and reviews. Studying previous evaluation approaches, their findings, and how they were used can provide ideas of what to do and not to do as a new evaluation is planned.

The overview of current and past driver education evaluations found in the literature review in Appendix A also provides an understanding of how to improve the evaluation of driver education programs. The following recommendations were based on lessons learned from the reviewed evaluations:

- Carefully describe the program, and develop a program logic model if an underlying program theory is not evident.
- Seek input from users and stakeholders.
- Expand the evaluation to consider the full range of formative and summative activities, as appropriate.
- Determine evaluation methods based on an assessment of evaluation capability and document rationale.
- Ensure an ongoing role for evaluation in program improvement activities.

All of these recommendations are addressed at various points in the *Guidelines*.

➤ *Identify Potential Users and Uses of the Evaluation*

Think about who will use the evaluation findings and how this information will be used. Some of the users may be individuals external to the program. In this case, obtain their views on the focus of the evaluation and questions to be answered, and input this information into the planning process.

What will be done with what is learned from the evaluation is also important to consider. Will the findings be used by the program manager to make changes to instructional materials, such as handouts, pamphlets, videos, or tests? Will they be used to change the emphasis on certain areas of the curriculum or drop certain audiovisual materials? Will an external community agency use them to decide whether to partner with the program in support of impaired driving education or responsible citizenship on the road? Information about evaluation users and uses provides important input into the next tasks.

➤ *Identify Key Evaluation Questions and Targets*

Establishing the purpose of the evaluation will lead directly to looking back at the information describing the program in Step 1A. This helps determine which parts of the program need to be evaluated to meet the evaluation's goals. If, for example, the evaluation's main purpose is to learn more about the effectiveness of the classroom materials in increasing student knowledge, then these materials will be the focus for the evaluation. As will be seen in Step 2 on evaluation design and methods, formative evaluation might or might not be the sole focus. The broader the evaluation, the more will be learned about how to improve the program's effects.

Having identified the program areas to be evaluated, questions can now be created to help specify the aspects of the program to be evaluated. Examples of evaluation questions for a driver education program are:

- Does the in-class instruction component of the program result in satisfactory or expected changes in students' knowledge?
- Do the students who successfully complete this program have fewer crashes in their first year of driving after becoming licensed than comparable new drivers who did not take driver education?

A checklist can help identify evaluation questions and assign priority to the information from the answers. Table 7 presents a hypothetical example to illustrate how an evaluation team might assign priorities to a general set of questions.

**Table 7. Sample Evaluation Questions Checklist**

Activities and Processes	How important are the answers to these questions for the evaluation?
Are program activities being implemented as intended?	High
Do staff think they are well prepared to teach the course?	High
What factors limit the implementation of the program?	Medium
How are partnerships working?	Low
How well do the program activities work?	High
How are the program resources being expended?	High
Are the current resources sufficient?	High
Is the program at the point where it could be expanded?	Low
Target Groups	
How many students attend the course each year?	High
What is the potential market for the program?	Medium
Are potential participants aware of the program?	High
What is known about customer satisfaction?	High
What is known about the program's reputation?	Medium
How do participants find out about the program?	Medium
Outcomes	
Have the program's outcomes been achieved?	
• Increased knowledge?	High
• Improved car-handling skills?	High
• Increased pass rates?	Medium
Have the program's impacts been achieved?	
• Increased licensure rates?	High
• Decreased re-test rates?	Medium
• Decreased crash rates?	Low
How can the program be improved?	High

Adapted from *A Program Evaluation Tool Kit*, Porteous, Sheldrick, and Stewart 1997.

Once the list of the evaluation questions is created, the team decides what priority each question should have in the current evaluation cycle—high, medium, or low. Priorities may change in later evaluation cycles, as might be the case in this example for crash rates. The team then tallies the number of questions in each level, and assesses the number that can be addressed immediately. The team starts with the high-priority questions and then moves to the medium- and low-priority ones if time and resources are available to allocate to them. A worksheet to use when developing your evaluation questions is found in Appendix E.

Several factors should be included in determining the priority of questions, such as what the need for the information is, why it is important, and how it will be used. Be clear on the rationale for these decisions. Some evaluators use the “SMART” principle to check the feasibility, adequacy, and priority of evaluation questions.

THE SMART PRINCIPLE

<b>S</b>	<b>pecific</b>
<b>M</b>	<b>easurable</b>
<b>A</b>	<b>ctionable</b>
<b>R</b>	<b>elephant</b>
<b>T</b>	<b>imely</b>

Ensure that the questions are:

- Specific** Evaluation questions must be specific, clear, and easily understood by all evaluation team members.
- Measurable** There must be specific measures that will provide the answers to all the questions.
- Actionable** The answers must provide the information needed to make decisions about the program.
- Relevant** High-priority questions must provide needed, rather than “nice-to-have” information. Knowing who needs the information, why it is needed, and what will be done with it will help decide if each question is a high priority.
- Timely** Make sure asking each question “right now” is important.

Adapted from *A Program Evaluation Tool Kit*, Porteous, Sheldrick, and Stewart 1997.

If a question fails to meet any of these five criteria, revise it or eliminate it as a high priority for this evaluation cycle. This is also an opportune time for evaluation team members to think about the expectations they have about each of the evaluation questions. Ask what the program is expected to accomplish and how the outcomes can be quantified. These questions then lead to identifying the evaluation targets—the specific aspects of the program that are going to be evaluated. Table 8 presents a comprehensive list of general targets for driver education evaluation activities.

**Table 8. Driver Education Evaluation Targets**

Evaluation Type	Program Area	General Evaluation Targets
Formative Evaluation	Program Logic	User needs
		Program logic model or theory
		Evaluability
	Program Context	Stakeholder expectations
		Regulatory environment
		Contractual environment
	Business Processes	Operations management
		Quality control
		Marketing
		Customer service
		Sustainability
	Program Standards	Benchmarking and certification
		Transportability of program
	Instructional Products	Curriculum materials
		Tests and measurement
Instructional Processes	Instructor preparation	
	Curriculum delivery and in-car practice	
	Instructional facilities	
Summative Evaluation	Student Outcomes	Knowledge outcomes
		Skill outcomes
		Motivation outcomes
		Mobility outcomes
		Behavioral outcomes
	Social Impacts	Crash reduction impacts
		Injury reduction impacts
	Socioeconomic impacts	
Metaevaluation	Evaluation Quality	Evaluation effectiveness

Say, for example, one of the questions the evaluation team asks is “How satisfied are our customers with our program?” Customer service (under the program area “Business Processes”) then becomes an evaluation target. Both the question and the evaluation target, however, are very general and need to be made more specific. What aspects of customer satisfaction do you really need to know about? These will need to be specified at this stage of evaluation planning.

As the team works through this process, it may become clear that two specific customer service issues are important. Are students satisfied with in-car lesson scheduling, and are parents satisfied with the feedback that is intended to keep them informed of their teenager’s progress? As a result, the team establishes specific evaluation targets: 1) percentage of students highly satisfied with in-car lesson scheduling; 2) assessment of their recommendations for improvement; 3) percentage of parents highly satisfied with feedback processes; and 4) assessment of their recommendations for improvement.

The Evaluation Targets Worksheet provided in Appendix E can be used to help specify what is going to be evaluated.

## STEP **1C** APPLY EVALUATION STANDARDS

### Overview

Considering the relevant program evaluation standards (pages 37-39) is important as evaluation planning takes place. Standards in all four groups—Utility, Feasibility, Propriety and Accuracy, apply to Step 1.

1A. DESCRIBE THE PROGRAM	1B. PLAN THE EVALUATION	<b>1C</b> APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify stakeholders, and user and program needs</i></li> <li>➤ <i>Identify the program’s vision, goals, and objectives</i></li> <li>➤ <i>Identify and document program activities, resources, and context</i></li> <li>➤ <i>Develop a program logic model</i></li> <li>➤ <i>Assess program readiness to be evaluated</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify the purpose of the evaluation</i></li> <li>➤ <i>Identify knowledge from driver education evaluations</i></li> <li>➤ <i>Identify potential users and uses of the evaluation</i></li> <li>➤ <i>Identify key evaluation questions and targets</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

As you work through this step, refer to the checklist of questions in the following chart. They will help ensure the standards are an integral part of the evaluation process. Similar checklists appear at the end of each evaluation step. Definitions of these standards can be found in Appendix C.



### Evaluation Standards Checklist

Step 1 Standards	Step 1 Questions
<b>UTILITY</b> Stakeholder identification	<input type="checkbox"/> Have individuals and organizations that will be affected by the evaluation been identified, and if appropriate, included on the evaluation team?
Evaluator credibility	<input type="checkbox"/> Are the participants on the evaluation team trustworthy and competent?
<b>FEASIBILITY</b> Political viability	<input type="checkbox"/> Have political interests and the needs of relevant groups been considered?
<b>PROPRIETY</b> Service orientation	<input type="checkbox"/> Has the evaluation been planned to assist with addressing the needs of the program's target groups and stakeholders?
Complete and fair assessment	<input type="checkbox"/> Has the evaluation been planned to ensure its completeness and fairness in order to build on program strengths and address problem areas?
Conflict of interest	<input type="checkbox"/> Have conflict of interest issues been identified and dealt with openly and honestly to ensure the evaluation processes and results are not compromised?
Fiscal responsibility	<input type="checkbox"/> Are the evaluation's resource allocation and expenditures prudent and ethically responsible, and do they reflect sound accountability?
<b>ACCURACY</b> Program documentation	<input type="checkbox"/> Has the program been described and documented clearly and accurately?
Context analysis	<input type="checkbox"/> Has the context within which the program exists been examined in enough detail to identify the influences on the program?
Described purposes and procedures	<input type="checkbox"/> Have the purposes and procedures of the evaluation been described in enough detail that they can be identified and later assessed?

Once the targets are set, the appropriate evaluation methods and the sources from which to gather data can be determined. Evaluation methods are the subject of Step 2, and data collection and analysis are found in Steps 3 and 4. The evaluation team should also begin thinking about how the evaluation findings will be used and how to ensure they are used to positively affect the program. This will be discussed in detail in Step 5.

# STEP 2

## SELECT EVALUATION METHODS

The next step is to determine the type of evaluation to undertake and select the appropriate research methods. Step 2 is a critical point in the evaluation. It is here that careful consideration and assessment of the evaluation process and the evaluation team’s capabilities takes place.

Step 2 involves working out the details of the evaluation approach and design, ensuring that the appropriate methods are selected. Important methodological and ethical issues are addressed, and the relevant program evaluation standards applied. The three major tasks and their related activities are summarized in the chart below.

### Summary of Activities

2A DETERMINE EVALUATION APPROACH	2B DETERMINE EVALUATION DESIGN	2C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify evaluation approach options</i></li> <li>➤ <i>Determine evaluation level</i></li> <li>➤ <i>Select research methods</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Develop research design</i></li> <li>➤ <i>Determine sample sizes</i></li> <li>➤ <i>Develop ethics and rights of human subjects procedures</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## STEP 2A DETERMINE THE EVALUATION APPROACH

### Overview

When this step is completed, some of the most important decisions about the evaluation will have been made. The overall approach to the evaluation will be agreed upon, and the research methods finalized. The following chart shows the activities for this step.

**2A**

<b>2A DETERMINE EVALUATION APPROACH</b>	<b>2B DETERMINE EVALUATION DESIGN</b>	<b>2C APPLY EVALUATION STANDARDS</b>
<ul style="list-style-type: none"> <li>➤ <i>Identify evaluation approach options</i></li> <li>➤ <i>Determine evaluation level</i></li> <li>➤ <i>Select research methods</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Develop research design</i></li> <li>➤ <i>Determine sample sizes</i></li> <li>➤ <i>Develop ethics and rights of human subjects procedures</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Identify Evaluation Approach Options*

There are three important types of evaluations, defined by their intended purposes: 1) formative evaluation, which provides information to improve some aspects of the program; 2) summative evaluation, which provides information to help demonstrate the effects of the program; and 3) metaevaluation, which improves the quality and impact of evaluations.

### Formative Evaluation

*Formative evaluation* is used to help develop and improve programs, and is appropriate to use when the evaluation addresses:

- Program theory including the program’s theoretical and logical bases.
- Program context including political, economic, and social environments.
- Instructional products such as the program’s materials and content.
- Instructional processes or educational and training delivery methods.
- Business processes including the program’s implementation, operation, and management.
- Program standards including principles, regulation, and governance.

Formative evaluation is the basic level of evaluation that should be an integral part of the development and ongoing management of all programs. It helps to better understand and improve the quality of all aspects of a driver education program. Virtually any

organization capable of delivering a program can and should carry out at least a modest scale formative evaluation. As will be seen later, a variety of qualitative and quantitative methods can be used in formative evaluations, and some can certainly be employed by non-specialists. Furthermore, help can be found at modest cost for methods requiring special knowledge and skills.

### Summative Evaluation

*Summative evaluation* is used to understand and demonstrate the effects of a program. Summative evaluations in driver education will most likely address the effects in two target areas:

- Student outcomes—the direct effects of the program on students’ knowledge, skills, motivations, mobility, exposure to risk, and driving behavior; and
- Social impacts—crash reduction, injury reduction, and other less-direct socioeconomic effects of the program.

Evaluation of student outcomes is a basic part of quality management of a driver education program, as well as a suitable target for special evaluation research projects. At a basic level, it can be as simple as tracking test scores over time. At the most complex levels, it involves exhaustive testing, surveys, and sophisticated data analysis.

Most documented driver education evaluations have been summative evaluations focused on safety impacts. Typically, safety impact has been defined as differences in total crashes, of all severity levels, as measured either by official collision records or self-report on surveys. This remains an important approach, and more and better summative evaluations of safety impact should be encouraged. Past research suggests, however, that a number of driver education programs have failed to show a measurable effect on crash rates.

With only summative information on safety impacts, little direction is available for program improvement, especially when the results are disappointing. Unless an evaluator has intermediate measures of student learning and other outcomes that show promise, or has other reasons to believe that a program has a chance of showing substantially improved safety impacts, summative evaluation of crash effects is not warranted. Past experience has shown that summative evaluation of safety impacts, by itself or as a first step, is not

sufficient; however, as part of a wider evaluation program, including student outcomes, summative evaluation of safety impacts will continue to be useful and important. It is especially important if summative evaluation can be:

- a) Expanded to provide information on a wider range of outcomes and impacts, including those occurring outside driver record crash rates; or
- b) Refined to provide a more detailed look at the impacts associated with sub-groups of students.

A comprehensive summative evaluation includes a range of effects. Even if no reductions in new driver crashes are found, summative evaluation can show where the program effects stopped in the chain of effects before reaching the desired safety impacts. It can also show whether factors such as changes in exposure to risk might offset potential safety benefits.

These two evaluation types, formative and summative, seem at first to be more important than the third, metaevaluation. This type of evaluation might seem interesting mainly to evaluation specialists; however, because the ultimate effectiveness of both formative and summative evaluations is dependent on evaluation quality, the assessment of evaluation is important. It is valuable for evaluators to know that others will look at their evaluations and metaevaluate them. Good evaluations will contribute to progress in driver education programs and improvement of driver education evaluations. Weak evaluations will bring discredit or suspicion upon the evaluator and the program.



*. . . evaluation should not be a one-shot effort but an ongoing part of a cycle of continuous improvement.*

Driver education has suffered from a number of weak evaluations. Thus, metaevaluation takes on significant importance. Evaluating and improving driver education evaluations are critical to program improvement. At the end of each evaluation project or cycle, the quality of the process and credibility of the results must be critically assessed and shared with others. Keep in mind that evaluation should not be a one-shot effort but an ongoing part of a cycle of continuous improvement. This requirement is reiterated in several places as the evaluation process continues through the five steps.

## ➤ *Determine Evaluation Level*

With the general evaluation plan completed, and the range of possible approaches identified, the evaluation team has reached a critical point in the process. It is time to carefully assess the resources and time available for the evaluation, and determine the specific details of the evaluation. We call this process determining the evaluation “level.” This is the point at which the team asks the question “Given the size of our program, the resources available, and what we already know about the program, what is the best level of evaluation to undertake right now?”

There are different evaluation levels, all of which can be useful. Not every organization will be capable of performing evaluations at the higher levels. While precise sets of activities will vary, four broad evaluation levels are identified, and examples of evaluation activities suitable to each level are suggested. These general levels can help determine the scope of an evaluation, taking into consideration program goals and objectives, size, number of students, needs, and resources.

### **Level 1**

The first level is entirely formative and mainly qualitative. Virtually any well-managed local program provider or school authority can aspire to this level of evaluation, and it is the least demanding of resources. Level 1 starts with good recordkeeping and using records to monitor program operations. It includes the program and evaluation planning activities that provide the foundation for moving on to a more extensive evaluation when fewer constraints exist. Even if this is not likely in the foreseeable future, the activities in Level 1 will provide important information that can be used to guide program improvements, including:

- Taking steps to build program evaluation and development capability;
- Describing the program structure and environment;
- Building a logic model for the program;
- Benchmarking the program curriculum structure and materials to industry standards\*;
- Evaluating customer satisfaction levels; and
- Evaluating student reactions to materials and instruction methods.

---

\*Refer to ADTSEA, DSAA, and NIDB standards in Appendix I. Also see Appendix F for an example of a self-assessment tool.

## Level 2

Once the first steps of program and evaluation planning have been completed, resources can be freed up to build on these achievements. Work can begin on the next activities leading toward a more comprehensive and informative evaluation. Level 2 can be considered by evaluation teams that have more resources available than those required by Level 1. It can also be considered by teams who have completed most of the Level 1 activities and are ready to undertake a more active evaluation process. Level 2 activities add two important evaluation targets to those of Level 1:

- Evaluating student knowledge outcomes and skills through testing; and
- Committing to continuous improvement through evaluation and development cycles.

## Level 3

As capability and resources increase, more substantial evaluation levels can be developed. Ongoing formative evaluation to support program development can be supplemented by more demanding organizational quality benchmarking. Summative evaluations of student outcomes using quantitative methods can be developed. Major program providers, large materials suppliers, industry associations, and smaller state or provincial governments will have resources for evaluation at this level. Level 3 activities include those in Levels 1 and 2, and add:

- Auditing compliance with standards and regulations;
- Certifying quality management (e.g., ISO 9000);
- Evaluating student skill and knowledge outcomes through testing and observation;
- Evaluating safety and mobility impacts using quasi-experiments; and
- Assessing evaluation and development activities.

## Level 4

The most advanced evaluation level requires substantial resources and expertise, typically available only to national governments and larger state or provincial governments. Level 4 activities include those identified in the other three levels plus:

- Evaluating student skill and knowledge outcomes using instrumented vehicles and simulators;
- Evaluating driver education context and policy approaches;
- Evaluating safety impacts through ecological studies, quasi-experiments, and experiments; and
- Evaluating socioeconomic impacts through cost/benefit analyses.

With all available options identified, evaluation teams can assess their present capability to evaluate their driver education program, and undertake the most appropriate level of evaluation. Table 9 provides a summary of the activities that can be included in each level.

**Table 9. Suggested Activities for Evaluation Levels**

ACTIVITIES	Evaluation Level			
	1	2	3	4
• Take steps to build program evaluation and development capability	X	X	X	X
• Describe the program structure and environment	X	X	X	X
• Build a logic model for the program	X	X	X	X
• Benchmark the curriculum structure and materials to industry standards	X	X	X	X
• Evaluate customer satisfaction levels	X	X	X	X
• Evaluate student reactions to materials and instruction methods	X	X	X	X
• Evaluate student knowledge outcomes and skills through testing		X	X	X
• Commit to continuous improvement through evaluation and development cycle		X	X	X
• Audit compliance with standards and regulations			X	X
• Certify quality management (e.g., ISO 9000)			X	X
• Evaluate student skill and knowledge outcomes through testing and observation			X	X
• Evaluate safety and mobility impacts using quasi-experiments			X	X
• Assess evaluation and development activities			X	X
• Evaluate student skill and knowledge outcomes using instrumented vehicles and simulators				X
• Evaluate driver education context and policy approaches				X
• Evaluate safety impacts through ecological studies and experiments				X
• Evaluate socioeconomic impacts through cost/benefit analyses				X

It is important to determine in advance that the resources required for the planned evaluation are available and committed. Then proceed to the next task in the development of the evaluation—to look at research approaches and finalize the evaluation design.

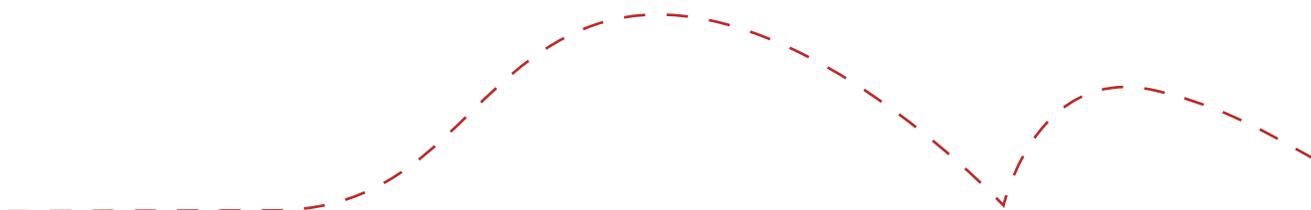
➤ *Select Research Methods*

A wide range of qualitative and quantitative research methods can be used to produce data for the evaluation. Qualitative methods involve accessing information from relatively small, non-representative groups of people. These methods ask in-depth questions of “what?” and they provide a rich understanding of people’s behaviors, opinions, feelings, and beliefs. Quantitative methods, by comparison, typically ask questions of “how much?” They may provide a less rich understanding of behaviors, opinions, feelings, and beliefs, but they provide an objective understanding of their strength and distribution, that is, how many people share those behaviors, opinions, feelings, and beliefs. Table 10 summarizes the differences between these two approaches to research.

**Table 10. Comparing Qualitative and Quantitative Research**

Qualitative	Quantitative
Richer, deeper understanding	Broader understanding
Verbal information from a small number of people	Numerical information from a relatively large number of people
Identifies what people think (i.e., range)	Identifies how many people think what (i.e., distribution)
Not readily quantifiable	Quantifiable, permits statistical analysis
Not representative of wider populations	May be generalized to whole populations

A wide variety of research methods are available in both qualitative and quantitative approaches. Table 11 provides examples of the types of methods evaluation teams can consider as the evaluation is designed.



**Table 11. Examples of Qualitative and Quantitative Methods**

Qualitative Methods	Quantitative Methods
Focus groups	Sample surveys
In-depth interviews	Statistical records analysis
Participant observation	Systematic observation
Diaries, Logs	Log book exposure surveys
Policy analysis	Testing
Critical review	Meta-analysis
Checklists	Cost/benefit analysis
Case study	Mathematical modeling

No single evaluation will use all of these methods, but consideration should be given to using a manageable number of different approaches appropriate to your evaluation. Attempting to answer evaluation questions from a number of different perspectives can help provide more confidence in finding the best answer.

Both qualitative and quantitative methods have appropriate applications to formative evaluations in most program areas. Expanding and improving formative evaluation research are important, but unfortunately, neglected aspects of driver education evaluation. Most program operators have a subjective, qualitative “feel” for the strengths and weaknesses of their products and operations; however, more objective and systematic qualitative research approaches can provide a broader, deeper, and more reliable understanding.

The checklist in Table 12 identifies methods that can be considered for the different evaluation types and program areas. Review these and select the ones that best meet the purpose of your evaluation and evaluation targets. Make sure they are manageable within the available resources and evaluation capability, and that they reflect the priorities identified in Step 1.

*No single evaluation will use all of these methods, but consideration should be given to using a manageable number of different approaches appropriate to your evaluation.*

**Table 12. Research Methods Checklist**

Evaluation Type	Program Area	Qualitative Methods	Quantitative Methods
Formative Evaluation	Program Logic	<input type="checkbox"/> Needs assessment <input type="checkbox"/> Literature review <input type="checkbox"/> Document review <input type="checkbox"/> Logic model development <input type="checkbox"/> Program content analysis	<input type="checkbox"/> Problem definition research <input type="checkbox"/> Meta-analysis
	Program Context	<input type="checkbox"/> Stakeholder analysis <input type="checkbox"/> Compliance checklist <input type="checkbox"/> Policy review <input type="checkbox"/> Legal review	<input type="checkbox"/> Stakeholder survey <input type="checkbox"/> Compliance audit
	Business Processes	<input type="checkbox"/> Customer satisfaction interviews <input type="checkbox"/> Customer satisfaction focus groups <input type="checkbox"/> Staff interviews <input type="checkbox"/> Case study	<input type="checkbox"/> Accounting audit <input type="checkbox"/> Quality control analysis <input type="checkbox"/> Marketing analysis <input type="checkbox"/> Customer satisfaction survey
	Program Standards	<input type="checkbox"/> Benchmarking <input type="checkbox"/> Certification	<input type="checkbox"/> Industry survey <input type="checkbox"/> Jurisdiction survey
	Instructional Products	<input type="checkbox"/> Content analysis <input type="checkbox"/> Focus groups <input type="checkbox"/> Interviews <input type="checkbox"/> Instructor log books	<input type="checkbox"/> Pilot testing <input type="checkbox"/> Student survey <input type="checkbox"/> Instructor survey
	Instructional Processes	<input type="checkbox"/> Student and parent interviews <input type="checkbox"/> Student and parent focus groups <input type="checkbox"/> Observation <input type="checkbox"/> Benchmarking <input type="checkbox"/> Instructor log books <input type="checkbox"/> Case study	<input type="checkbox"/> Quality control analysis <input type="checkbox"/> Pilot testing <input type="checkbox"/> Records analysis <input type="checkbox"/> Student survey <input type="checkbox"/> Parent survey <input type="checkbox"/> Instructor survey
	Summative Evaluation	Student Outcomes	<input type="checkbox"/> Focus groups <input type="checkbox"/> Interviews <input type="checkbox"/> Student diaries <input type="checkbox"/> Case study
Social Impacts		<input type="checkbox"/> Policy analysis <input type="checkbox"/> Case study	<input type="checkbox"/> Surveys- student, parent, stakeholder, general public <input type="checkbox"/> Records study <input type="checkbox"/> Cost/benefit analysis
Metaevaluation	Evaluation Quality	<input type="checkbox"/> Benchmarking against Program Evaluation Standards <input type="checkbox"/> Evaluation review for technical quality and usefulness of findings <input type="checkbox"/> Peer review <input type="checkbox"/> Stakeholder interviews	<input type="checkbox"/> Meta-analysis

Summative evaluations of outcomes and impacts require the objectivity and ability to be replicated provided by quantitative methods. Even in these cases, qualitative information can be useful in planning quantitative research. Knowing what participants think about program outcomes for them and their peers can help guide the formulation of quantitative research questions.

The evaluation team will need to consider available skill sets and other resources for a feasibility check on desirable research approaches. Even highly qualified evaluators will not necessarily be knowledgeable in all approaches and methods. It is important to recognize when help from experts is needed in areas beyond the evaluation team’s expertise. As mentioned earlier, Appendix J provides guidance on how and when to hire an external evaluator.

## STEP **2B** DETERMINE THE EVALUATION DESIGN

### Overview

This step will complete the evaluation methods work and finalize the details of the evaluation design. Critical decisions about the design, including data validity, reliability, comparisons, and sample sizes will be made here. The activities included in this step are listed in the following chart. For more complex evaluations, these activities may require assistance from a statistician. Simpler evaluations may not need to consider all these activities in detail.

2A DETERMINE EVALUATION APPROACH	<b>2B</b> DETERMINE EVALUATION DESIGN	2C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify evaluation approach options</i></li> <li>➤ <i>Determine evaluation level</i></li> <li>➤ <i>Select research methods</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Develop research design</i></li> <li>➤ <i>Determine sample sizes</i></li> <li>➤ <i>Develop ethics and rights of human subjects procedures</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Develop Research Design*

After determining the evaluation approach and deciding which types of qualitative and quantitative methods to include, it is necessary to specify how data are to be used. An evaluation is a research project. Its design must meet certain standards if it is to produce valid and reliable results and if it is to be seen as credible by professional peers. Important considerations as the research design is developed are discussed next.

Valid comparisons are the core of evaluation research design. Data alone provide a description of something and do not constitute an evaluation. As seen in Step 1, description is part of getting ready for evaluation, but it is not actual evaluation. A group of students, for example, might be found to produce a certain average score on a test. Or a school's instructors might have a certain number of hours of instruction, or a curriculum analysis of its content might indicate a certain number of topics addressed. So what? If the data are going to be used in an evaluation, something appropriate is needed to compare the data to. One group of students' scores, for example, can be compared to those from another group of students who use a different curriculum. Instructor qualifications can be compared to a regulatory requirement, and the curriculum content to a benchmark curriculum standard.

Comparisons for driver education evaluation data can be made using a few basic methods, as Table 13 indicates. Understanding the differences among these methods helps decide which ones to include in the evaluation.

**Table 13. Data Comparison Methods for Driver Education Evaluations**

Comparison Methods	Definition
Benchmarking	Comparing data to an established standard
Longitudinal studies	Comparing new data to data from repeated measurements on the same subjects taken at different points in time
Quasi-experiments	Comparing new data to similar data gathered from a pre-existing comparison group
Experiments	Comparing new data to similar data from a specially created control group

These basic methods can appear in a multitude of variants to facilitate data comparisons. Many driver education evaluations have included experiments, quasi-experiments, or longitudinal studies. Benchmarking and other more qualitative efforts occur, but they are less apparent, in part at least, because they are not typically published in the research literature.

Longitudinal or time-series comparisons can be used to evaluate a program change. Such a design typically involves collecting data at many points in time before and after the change. The data in such a case could be, for example, student test scores on a final test. A difference seen in the data around the time of the change *might* be a result of the change. The “might” represents the typical difficulty in interpreting such a finding because the possibility that something else may have changed to cause the difference cannot be ruled out. A comparison with data from students not exposed to the change can add confidence to a time-series study. A variant on this design known as “ecological” has been used to evaluate impacts on crash records in a whole state or country following changes in requirements or support for driver education.

*Many driver education evaluations have included experiments, quasi-experiments, or longitudinal studies. Benchmarking and other more qualitative efforts occur, but they are less apparent . . .*

Basic quasi-experimental and experimental designs collect data at one or a few times from two or more groups of people and compare the results between the groups. These can be strong designs for attributing a causal connection between the program and the resulting differences observed in the data from the two groups. They only help, however, if we can be reasonably sure no other extraneous differences beside the program exist between the groups that could cause the observed differences in results. Extraneous factors are those that might affect the results but are not the ones that we are trying to study. For example, if groups given different driver education also differ in some other factor that affects crash rates, such as age, sex, or amount of driving, these extraneous factors could create differences in later crash rates between the groups or mask any real effects of the different programs.

## Randomized Controlled Trials (RCTs): The Comparison Problem in Beginner Driver Education Evaluation

A controversial issue in the consultations for these *Guidelines* involved the use of the experimental study design known as randomized controlled trials (RCTs). In this design, people are assigned randomly to either a treatment group, which receives the target treatment, or to a control group, which receives no treatment. This allows a reasonable level of confidence, which can be calculated, that different outcomes are actually caused by the different treatments. In other designs, one cannot be sure whether differences resulting from self-selection or other confounding factors may have biased the results.

Because of the “clean” comparison an RCT permits, it is the gold standard for establishing causal relationships in the experimental sciences. In evaluation of beginner driver education, in particular, there are basic practical problems with RCTs. Problems arise around choosing a comparison condition against which to assess the results of the target program. Researchers have suggested that informal driver training by parents is the most suitable comparison. It is not, of course, possible to compare driver education to no training at all, since new drivers have to learn somehow.

The difficulty is in assigning students who apply for driver education to informal training, at least in part because it may mean foregoing an insurance premium discount. Many students may seek other formal training despite their group assignment, and this reduces the resulting differences between the groups. Other forms of driver training, such as driver improvement programs, can be withheld from a randomly assigned no-treatment group, and the no-treatment group is unlikely to seek training. But this is not the case with driver education.

Alternative approaches may get around the comparison problem. A California RCT study simply compared two different forms of driver education, without an informal training control group. This study found one form of training had a substantial safety benefit over the other. It has been ignored by most reviewers, apparently because of the lack of a true control group, which leads to uncertainty whether the “better” treatment had a real positive impact. It is possible that the better treatment had no actual effect and the “worse” treatment yielded a negative effect. It is even possible that both had negative safety impacts, but the “better” treatment was just the less detrimental of the two treatments. These possibilities cannot be dismissed lightly, as some driver education evaluations have found negative impacts.

Quasi-experimental studies can be used to look for differences between self-selected driver education students and those who learn to drive in other ways. This can be a good design, but only when steps are taken to control for the ways in which driver education grads may be different from comparison groups other than training. Self-selected groups often differ in many ways, such as age, sex, income, or other socioeconomic factors, and these extraneous factors may contribute to observed differences in outcome measures. Without proper statistical methods to control extraneous factors, it is not possible to tell what causes any differences in outcomes observed between self-selected groups. In other words, bias between the groups can “confound” the comparison and jeopardize the validity or accuracy of the results.

The bottom line is that it is difficult to use RCTs to evaluate beginner driver education because of the difficulty in assigning people to informal training. RCTs should be part of a broader and more systematic evaluation approach, which includes other study designs, a wide range of output and outcome measures, and a variety of carefully planned comparisons. The evaluation “gold standard” title should be passed along to a comprehensive, systematic evaluation program that includes a variety of approaches.

## Reducing Bias

Bias means that there is some kind of systematic error in establishing equivalent comparison groups. Extraneous factors are present that can affect results and compromise the validity of conclusions about what caused any differences in results between the groups. Self-selection bias occurs when groups are comprised of subjects who choose which group they will belong to, rather than being assigned randomly.

**B***ias means that there is some kind of systematic error in establishing equivalent comparison groups.*

Experimental studies randomly assign subjects to groups to make the groups more or less equivalent on all extraneous factors, at least at the time of assignment. Random assignment is the process of assigning two or more groups by chance, which overcomes self-selection bias and is a critical aspect of experimental design. Because other biases can occur, even between randomly assigned groups, however, it is important to add controls to the design to reduce biases.

One such bias (called attrition bias or differential mortality) occurs when people drop out of one group at different rates than from the comparison group; one group loses people in a different pattern than the other. Subjects are lost from assigned groups for a variety of reasons. They may not complete the course to which they were assigned, or they may get sick, move, experience changes in their schedules, or decide they do not want to participate in follow-up surveys or interviews. Different attrition can make groups that were equivalent at the start no longer equivalent when it is time to measure results.

One way of checking for differences between groups in experiments and quasi-experiments is to add “before” and “after” measurements to the design. Data are gathered from both groups before and after the program or change takes place. The “before” data can give an indication of the pre-existing differences between the groups and show whether the random assignment actually resulted in equivalent groups.

Another way to enhance the validity of the conclusion is to control for potentially confounding extraneous factors. Statistical procedures try to produce a model that specifies that nature and extent of the connection between the program and the results, free from the confounding factors contributing to the pre-existing differences between the groups. They do this by assessing and adjusting for the effect of known confounding factors on the results.

While this approach is highly promising, it requires a lot of information to describe the two groups. This results in added complexity and the cost of identifying information about the possible confounding group differences. There are other advantages of gathering more data from the groups, however. Recall that early evaluations failed to look at intermediate outcomes—knowledge, skills, attitudes, intentions, values, or behaviors that had, or had not, been affected. Thus, ways to improve programs remained unclear. The more that is known about the groups, the greater the likelihood that any special effects on sub-groups will be found.

### ➤ *Determine Sample Sizes*

Selecting the people to participate in the evaluation and whose data will, therefore, be used is important, especially for evaluations that depend on data from individuals or groups of people, such as surveys or record studies. This selection process is known as sampling. While seemingly a simple matter, it occupies much time and causes much dispute among researchers.

## **Sample Selection**

There are various ways to select a sample. The easiest is a convenience sample, which simply means that anybody easily found who can provide relevant information is included. For some purposes, this might be adequate, but typically a convenience sample is not considered representative of any wider population of people, and research findings may result from some peculiarity of the sample rather than from the target of the research. The possible peculiarities of a convenience sample are another potential source of bias. A biased sample would not be representative of students, drivers, or people in general. A commonly used relative of the convenience sample is a volunteer sample. The ways in which volunteers for research might differ from a wider population is called volunteer bias.

It is important to try to make a sample as representative as possible. If we want to be able to say that the evaluation results can be applied beyond the group from which the data have been gathered, such as a sample of driver education students, the sample has to be representative of the wider group, all the driver education students who have taken the program. For this to be the case, the sample should be randomly selected and assigned so that every person in the wider population has an equal chance of contributing data.

Sometimes rather than selecting individual people, whole groups can be sampled. This is called cluster sampling. An example might be listing a number of schools and selecting all the students of a sample of the schools. If possible, schools should be selected randomly from a wider grouping of schools that they are supposed to represent. While this is a good approach, it does complicate matters for the evaluation's statistician in terms of sample size calculations and data analysis.

In evaluation designs that compare data from two or more groups of people, it is important to know whether the groups are equivalent both before and after the program is delivered. In other words, do the groups represent the same wider population? If the people to be sampled are selected at random from the same population and randomly assigned to groups, then, on average, the groups will be equivalent before the program, and this would be verified by analysis of pre-test data. Statistical analyses of the after-treatment results indicate the probability that the observed result could have occurred by chance rather than as a result of the program.

In fact, no sample in practical research is completely representative of people in general, and no comparison or control groups are exactly equivalent. Even if we could start with perfectly equivalent groups, by the time the research is completed, the groups will usually be different to some degree. In large part this is due to attrition. People drop out of all kinds of activities for many different reasons. Who and how many drop out are often different in different groups; for example, if the students are randomly split into treatment and control groups, they will start out equivalent. Once the treatment group is assigned to a course of study and the control group to a no-treatment control condition, the two groups will have different attrition rates and may, by the end of the research, no longer be equivalent. Some of the students in the treatment group will fail to attend or complete the course. The no-treatment control does not require its members to do anything, so there is no dropout or "failure-to-complete" group.

*. . . no sample in practical research is completely representative of people in general, and no comparison or control groups are exactly equivalent. Even if we could start with perfectly equivalent groups, by the time the research is completed, the groups will usually be different to some degree.*

A common way to correct for differential attrition between groups is to analyze the data by assignment, that is, to use the results of all subjects who were assigned to groups, including those who fail to start or complete the requirements of the group to which they were assigned. This is adequate for some research purposes, but in evaluation, it can bias the results toward a finding of “no effect,” since it includes in the training group those who were not actually exposed to the training. As was seen in the DeKalb study, quite different results can occur when analyzing by assignment or by completion. Where feasible, using both approaches and reporting both results add value to the evaluation. When analyzing by completion, some of the bias can be corrected by statistical means, but only if enough information exists about the people to assess how the dropouts differ from those who, for example, complete the course or survey.

## Sample Size

Aside from comparability of study groups, the other major issue in driver education evaluation samples is size—the number of people selected for the sample. In a typical research design, a population characteristic, for example, people’s opinion of driver education, can be measured using a survey questionnaire. A single sample of people is surveyed and the results are used to estimate the real opinion of the wider population. As seen above, even a random sample will usually have some differences from the population it represents. Larger samples, on average, provide estimates closer to the true population value.

Say that 50% of a sample of people surveyed were found to be happy with a school’s facilities. The range of opinions likely in the whole population of customers needs to be estimated from the sample results. If the sample size were 20 people, we could be 95% sure that the “real” favorable percentage is within 22 percentage points on either side of our sample value, 50%. This range (28% to 72%) is known as the 95% confidence interval for the true population value. To put it another way, if samples of 20 were repeatedly drawn, 95% of the estimates would fall between 28% and 72%. To show the importance of sample size, if the sample were 200, the 95% confidence interval would be 43% to 57%. This is still a fairly wide margin, but we could be much more comfortable saying the true value is “about half.”

Sampling variation and uncertainty of sample estimates from small samples affect the ability to compare groups such as those who take a course at one facility and those who attend another. Researchers would apply statistical significance tests to be able to say whether there is a statistically reliable or “significant” difference between them. Even if a fairly large difference between groups is observed, we cannot always be sure the apparent difference is “real” if the sample size is too small and the confidence intervals associated

with the apparent difference are too wide. Unless observed differences can be shown to be statistically significant by appropriate statistical procedures, we cannot reject the idea that the observed differences occurred by chance. A number of evaluations have been carried out with sample sizes that were too small to detect even fairly large program effects. This lack of statistical “power” can lead to errors in inferences about what might be a real program effect. Of course, it is also possible to make errors in the other direction—concluding that a real effect exists where there is none. A proper study design and statistical analysis can balance the risks of these two kinds of errors.

Required sample sizes become very large when trying to measure a relationship between driver education and crash rates because reported crashes are relatively rare. Sample size requirements can be calculated from statistical texts. Sample size is based on a number of factors including:

- How the characteristic being measured is distributed in the population (e.g., the prevailing crash rate for 16-year-old drivers);
- The size of the difference one hopes to be able to detect;
- How certain one wants to be that any apparent effects found are not due to chance (significance level); and
- How much chance one is willing to take of overlooking a real effect (statistical power).

Assume, for example, a 5% level of significance ( $\alpha=0.05$ ) and statistical power of 80% ( $1-\beta=0.80$ ). If prevailing crash rates in your jurisdiction suggest that about 20% of new drivers can be expected to crash during the period examined, then to detect a 10% reduction associated with your program (i.e., a crash rate reduction from 20% to 18%) with 95% confidence, a sample of roughly 6,000 students per group would be required. If larger effects (e.g., a 20% reduction) were expected, however, and the evaluator was prepared to dismiss a smaller reduction as unimportant, a sample of roughly 1,400 students per group would suffice. In a jurisdiction where only about 10% of new drivers were expected to crash, the detection of a true difference of 10% with 95% confidence would require about 13,000 students per group, and even the detection of a 20% difference would require upwards of 3,000.

In contrast to the hypothetical sample sizes suggested here, many past driver education evaluations that have examined crash rates have used samples so small that only truly enormous differences would have been detectable at traditional levels of statistical significance. As illustrated in Appendix A, there have been evaluations that have used only a few hundred drivers. When considering driver record-based evaluations, experts in statistical methods and handling large data sets should be consulted.

## Statistical Significance and Statistical Power

In the social sciences, researchers usually aspire to detect differences at the 5% level of significance. This means that the probability of observing the difference that is observed in the evaluation, by pure chance, assuming that no “real” difference exists, is 5% or smaller. Therefore, a difference whose probability of arising due to chance alone is calculated to be greater than 5% is usually declared “not statistically significant,” whereas a difference whose probability is smaller than 5% is declared “significant.” An evaluation aspiring to detect a 10% difference in crash rates, significant at the 1% level, would demand an enormous sample, but provide the evaluator with great confidence in the result; whereas detecting a difference of the same size at the 10% level of significance would allow for a smaller sample to be used, albeit at the expense of an increased risk of wrongly concluding that a chance difference was “real.”

A related concept, statistical power, indicates the probability that a difference of specified size will be detected at the desired level of significance, given that the underlying difference really does exist. For example, if one aspires to detect a 10% difference in crash rates, at the 5% level of significance, with a statistical power of 80%, this means that if we assume the program in question actually decreases crash rates by 10%, roughly four out of five evaluation attempts should return a statistically significant result; however, about one in five would still conclude that the program has no statistically significant effect on crash rates.

Statistical power comes at a price. More power can be “bought” by decreasing the confidence level at which a given difference will be declared significant, or by increasing the sample size. There are always tradeoffs to be made between the risk of incorrectly concluding that an ineffective program is effective—when some apparent decrease in crash risk is really due to random chance and not an effect of the program—and failing to muster statistically significant evidence that a program is effective when it really is effective. In practice, the number of students enrolled in a program often dictates the size of the available sample. In this case, sample size is essentially non-negotiable, but a decision still must be made, either explicitly or implicitly, regarding the tradeoff between “confidence” and “power.”

The example sample size requirements just discussed can provide some perspective on just how large a sample is needed under various conditions. When designing an evaluation, the evaluator needs to have a thorough knowledge of these concepts, and the program manager should have at least a basic working knowledge of them, so that informed decisions can be made before the evaluation actually begins. For all but the largest programs, it is likely that a single class of driver education graduates would provide far too small of a sample to detect statistically significant differences in performance measures such as crash rates or violation rates, even if the differences were fairly large. If the program manager considers a 10% decrease in crashes to be hugely important, but the evaluator or the statistical consultant determines that nothing smaller than a 50% decrease in crashes could be detected statistically using the available sample and sound methodology, then the evaluation team must make an informed decision regarding whether including crashes as an outcome measure in the evaluation is sensible.

Implementing the parts of a Level 4 evaluation that might directly demonstrate safety effectiveness is relatively costly and difficult, and requiring large samples contributes to much of the difficulty. This is especially a problem for small programs, which can only produce small samples of graduates. In theory, however, indirect ways may exist to avoid the problem of small samples and infer probable safety effectiveness of small programs. In the future, small programs could be modeled on larger ones that have been shown to have safety benefits. If appropriate product, process, and student outcome evaluations showed the model being implemented effectively, the small program can probably be inferred to be effective, too. Inferring safety effectiveness indirectly through use of standardized performance test results that have been validated for a strong statistical relation to safety benefits may also be possible in the future. At present, both of these indirect methods of reducing sample size requirements will have to wait for further program development.

Additional assistance on research design and sample sizes can be found in evaluation and general texts on research methods for the health and social sciences as well as business management. Several sources are provided in the Evaluation Resources section on page 173.

➤ *Develop Ethics and Rights of Human Subjects Procedures*

Establishing ethical procedures for the evaluation and ensuring the rights of human subjects are adequately protected are extremely important tasks. They should be undertaken fairly early in the evaluation planning.

If your organization has its own ethics requirements, they will direct the evaluation's documentation and implementation of ethics procedures. If not, examine the review criteria and related actions outlined in Table 14. They provide guidance on the common elements of an ethics document, although not all criteria will necessarily apply to your evaluation. Some of the required information may already be available from earlier evaluation activities. This document is important to have available to share with stakeholders or sponsors, and to refer to, should questions arise about the evaluation's ethics procedures.

**E***stablishing ethical procedures for the evaluation and ensuring the rights of human subjects are adequately protected are extremely important tasks.*

**Table 14. Ethics and Protection of Human Rights Checklist**

Ethics Review Criteria	Required Action
Evaluation Purpose	<input type="checkbox"/> Prepare a brief summary of the evaluation’s purpose.
Evaluation Methodology	<input type="checkbox"/> Prepare a description of the evaluation approach and design, including participant recruitment procedures, expectations of participants, and data collection and analysis procedures.
Potential Risks to Participants	<input type="checkbox"/> Explain any potential risks (physiological or psychological) to participants in the evaluation. <input type="checkbox"/> If there is apparent risk, identify steps being taken to reduce it.
Rationale for Deception	<input type="checkbox"/> If deception is part of the evaluation, provide an explanation of why an alternative methodology cannot be used. <input type="checkbox"/> Document the evidence that the deception does not put any participant at risk.
Benefits of Evaluation	<input type="checkbox"/> Describe the potential benefits of the evaluation.
Informed Consent Measures	<input type="checkbox"/> Ensure informed consent of participants, if required. <input type="checkbox"/> Prepare and distribute cover letters and consent forms to all participants. <input type="checkbox"/> For telephone surveys, prepare a statement of introduction. <input type="checkbox"/> Prepare informed consent procedures to be used with participants who are legally under the age of consent (18 years of age), including consent forms for parents or legal guardians with provision for permitting or not permitting the participation of their teenagers. <input type="checkbox"/> Ensure that informed consent procedures include the right of participants to withdraw from the evaluation at any time without penalty of any kind.
Anonymity and Confidentiality	<input type="checkbox"/> Outline the procedures to be used to guarantee confidentiality and anonymity for participants. This is particularly important for populations such as students, who may be concerned about the power of the evaluator in a context related to, but not part of, the evaluation.
Storage of Data	<input type="checkbox"/> Ensure that the data will be securely stored for a specified period of time (e.g., 5-7 years) in a secure location.
Evaluation Partners	<input type="checkbox"/> State who will be involved in the evaluation from outside the organization, and provide evidence that such parties have agreed to the ethics procedures.
Dissemination of Results	<input type="checkbox"/> State how the evaluation findings will be disseminated and how the participants will be made aware of these findings.

If any of the required information, such as recruitment procedures or data collection and analysis decisions, is not yet available, be sure to add it as soon as it is. This document should be shared with evaluation partners and others to indicate the intent and importance of conducting and adhering to an ethical evaluation.

Note that the Joint Committee’s Program Evaluation Standards used throughout the *Guidelines* also address ethics in the “Rights of Human Subjects” and “Human Interactions” standards of the Propriety category (refer to Appendix C). Although these requirements overlap with some of those discussed here, they help to focus attention on specific parts of the evaluation where additional effort is needed to ensure that ethics considerations are implemented.

## STEP **2C** APPLY EVALUATION STANDARDS

### Overview

During Step 2, be sure to review the evaluation standards and keep in mind the importance of ensuring that they are met.

2A DETERMINE EVALUATION APPROACH	2B DETERMINE EVALUATION DESIGN	<b>2C</b> APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Identify evaluation approach options</i></li> <li>➤ <i>Determine evaluation level</i></li> <li>➤ <i>Select research methods</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Develop research design</i></li> <li>➤ <i>Determine sample sizes</i></li> <li>➤ <i>Develop ethics and rights of human subjects procedures</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

### Activities

Use the following checklist to consider the relevant standards for Step 2. If gaps are identified, correct them before proceeding.

**STEP  
2**



**Evaluation Standards Checklist**

Step 2 Standards	Step 2 Questions
<b>UTILITY</b> Evaluation impact	<input type="checkbox"/> Is the evaluation being conducted to ensure effective use of its results?
<b>FEASIBILITY</b> Cost effectiveness	<input type="checkbox"/> Has the evaluation been designed to ensure efficiency, the value of the information, and justification of the staff time and money expended?
<b>PROPRIETY</b> Service orientation	<input type="checkbox"/> Has the evaluation been designed to assist in addressing the needs of the target groups and stakeholders?
Rights of human subjects	<input type="checkbox"/> Has the evaluation been designed to respect and protect the rights and welfare of participants, and are the necessary procedures in place? <input type="checkbox"/> Does the evaluation design include sound, ethical, and consistent procedures to ensure that the findings are correct?
Human interactions	<input type="checkbox"/> Do the evaluation procedures respect human dignity and worth, to ensure that participants are not threatened or harmed?
Complete and fair assessment	<input type="checkbox"/> Does the design ensure that the evaluation is complete and fair in its assessment of the program's strengths and weaknesses?
Fiscal responsibility	<input type="checkbox"/> Are the evaluation's resource allocation and expenditures prudent and ethically responsible, and do they reflect sound accountability?
<b>ACCURACY</b> Described purposes and procedures	<input type="checkbox"/> Have the purposes and procedures of the evaluation design and methods been described in enough detail that they can be identified, monitored, and later assessed?

It's now time to move on to the third step—developing the data collection plan and tools.

# STEP 3

## DEVELOP DATA COLLECTION PLAN AND SELECT DATA COLLECTION TOOLS

In Step 2, the details of the evaluation approach and design have been carefully considered and finalized. Step 3 moves the evaluation process along by focusing on the type of information to be gathered, developing a data collection plan, selecting the data collection tools, and ensuring their quality. The Step 3 activities are listed in the following chart.

### Summary of Activities

3A DEVELOP DATA COLLECTION PLAN	3B SELECT AND ASSESS DATA COLLECTION TOOLS	3C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Determine appropriate data types and data gathering methods</i></li> <li>➤ <i>Specify data and sources</i></li> <li>➤ <i>Identify indicators for program success</i></li> <li>➤ <i>Assess feasibility of data collection plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Select, modify or develop tools</i></li> <li>➤ <i>Conduct quality assessment of tools and revise</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## STEP **3A** DEVELOP THE DATA COLLECTION PLAN

### Overview

The data collection plan documents the decisions that the evaluation team makes about what data are to be collected, from whom, and how they will be obtained. These activities are identified in the following chart.

**3A**

<b>DEVELOP DATA COLLECTION PLAN</b>	<b>3B. SELECT AND ASSESS DATA COLLECTION TOOLS</b>	<b>3C. APPLY EVALUATION STANDARDS</b>
<ul style="list-style-type: none"> <li>➤ Determine appropriate data types and data gathering methods</li> <li>➤ Specify data and sources</li> <li>➤ Identify indicators for program success</li> <li>➤ Assess feasibility of data collection plan</li> </ul>	<ul style="list-style-type: none"> <li>➤ Select, modify or develop tools</li> <li>➤ Conduct quality assessment of tools and revise</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

## Activities

### ➤ Determine Appropriate Data Types and Data Gathering Methods

For evaluation purposes, useful data can be many different things. What they cannot be are vague impressions. Useful data are pieces of information that can be captured objectively and preserved. Table 15 provides examples of useful data across the range of program areas and evaluation types. Refer to this list as you think about what types of data are to be collected in the evaluation.

**Table 15. Data Types**

Evaluation Type	Program Area	Qualitative Data Types	Quantitative Data Types
Formative Evaluation	Program Logic	<ul style="list-style-type: none"> <li>• User needs</li> <li>• Review of previous evaluation findings</li> <li>• Document review findings</li> <li>• Logical links among program components</li> <li>• Content analysis results</li> </ul>	<ul style="list-style-type: none"> <li>• Results from problem definition research</li> <li>• Meta-analysis findings</li> </ul>
	Program Context	<ul style="list-style-type: none"> <li>• Stakeholder expectations</li> <li>• Compliance issues</li> <li>• Policy analysis issues</li> <li>• Legal issues</li> </ul>	<ul style="list-style-type: none"> <li>• Stakeholder opinions</li> <li>• Compliance audit data</li> </ul>
	Business Processes	<ul style="list-style-type: none"> <li>• Customer concerns</li> <li>• Staff input</li> <li>• Interview results</li> <li>• Focus group results</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Accounting audit data</li> <li>• Quality control data</li> <li>• Marketing data</li> <li>• Customer satisfaction data</li> </ul>

Evaluation Type	Program Area	Qualitative Data Types	Quantitative Data Types
Formative Evaluation <i>(continued)</i>	Program Standards	<ul style="list-style-type: none"> <li>• Benchmarks</li> <li>• Certification reports</li> </ul>	<ul style="list-style-type: none"> <li>• Industry and jurisdiction standards</li> </ul>
	Instructional Products	<ul style="list-style-type: none"> <li>• Content benchmarks</li> <li>• Focus group and interview results</li> <li>• Instructor feedback</li> </ul>	<ul style="list-style-type: none"> <li>• Pilot test data</li> <li>• Student opinions</li> <li>• Instructor opinions</li> </ul>
	Instructional Processes	<ul style="list-style-type: none"> <li>• Focus group and interview results</li> <li>• Observed actions</li> <li>• Benchmarking</li> <li>• Instructor feedback</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Program delivery consistency</li> <li>• Pilot test data</li> <li>• Records data</li> <li>• Student and parent opinions</li> <li>• Instructor opinions</li> </ul>
Summative Evaluation	Student Outcomes	<ul style="list-style-type: none"> <li>• Focus group and interview results</li> <li>• Student diary entries</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Instrumented vehicle data</li> <li>• Video observation records</li> <li>• Test scores</li> <li>• Student opinions, exposure data</li> </ul>
	Social Impacts	<ul style="list-style-type: none"> <li>• Policy change recommendations</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parent, stakeholder and general public opinions</li> <li>• Records data on crash and casualty rates</li> <li>• Cost/benefit results</li> </ul>
Meta-evaluation	Evaluation Quality	<ul style="list-style-type: none"> <li>• Program Evaluation Standards Benchmarks</li> <li>• Results of technical quality and usefulness of findings review</li> <li>• Peer review results</li> <li>• Stakeholder opinions</li> </ul>	<ul style="list-style-type: none"> <li>• Meta-analysis data</li> </ul>

While many types of data are relevant to evaluating driver education programs, there are relatively few basic methods for generating and gathering data, as shown in Table 16.

**Table 16. Definitions of Data Generation and Data Gathering Methods**

Method	Definition
Record extraction	Documentary or electronic records
Observation	Watching and recording actual behavior
Interviews, focus groups	Individuals or groups answering verbal questions in depth
Survey	Groups answering questions in a standardized, structured format
Testing	Recording performance in simulated behavior

Data gathering can be organized in many different ways; for example, interviews can be conducted over the phone with a random sample of new drivers, or a focus group can be held with a volunteer sample of driving instructors. Who is selected to provide the data and how the data gathering is administered will vary depending on the design of the evaluation. Data collection always has costs, usually for both the researchers and those who provide the data, so careful choices are necessary to avoid collecting less-important data or data that will not actually be used.

As part of the evaluation design, data collection must be systematic. All data collection procedures should be documented so they can be replicated either by the same or another evaluator. Being clear about how data were obtained and handled is important, regardless of whether they are qualitative or quantitative. Methods for extracting qualitative data from interviews or focus groups should be clearly documented. In quantitative methods there are usually missing data, such as survey questions left unanswered. Missing data can be another source of bias that should be documented and dealt with consistently by, for example, deleting cases or estimating missing data. Guidance for dealing with missing data can be found in statistical texts or from experts.

➤ *Specify Data and Sources*

The data collection plan must also clearly identify the specific information to be collected (the data), where and from whom the information will be obtained or collected (data sources), and when to collect it. Table 17 provides a framework for identifying data sources.

**Table 17. Identifying Data Sources for Driver Education Evaluations**

Program Area	Evaluation Targets	Qualitative Data	Quantitative Data	Data Sources
<b>Formative Evaluation</b>				
Program Logic	User needs	<ul style="list-style-type: none"> <li>Needs identified from interview and focus group transcripts</li> <li>Previous evaluation findings</li> </ul>	<ul style="list-style-type: none"> <li>Problem definitions from research</li> <li>Meta-analysis findings</li> </ul>	<ul style="list-style-type: none"> <li>Students, parents, stakeholders, research on new drivers, experts on new drivers</li> </ul>
	Program logic model or theory	<ul style="list-style-type: none"> <li>Logical links among program components</li> <li>Document review findings</li> <li>Content analysis results</li> </ul>		<ul style="list-style-type: none"> <li>Educational research experts, program developers, staff, program records and documents</li> </ul>

Program Area	Evaluation Targets	Qualitative Data	Quantitative Data	Data Sources
Program Logic <i>(continued)</i>	Evaluability	<ul style="list-style-type: none"> <li>• Program review findings</li> <li>• Staff input via log books, interviews, and meetings</li> <li>• Content analysis results</li> </ul>		<ul style="list-style-type: none"> <li>• Evaluation experts, program developers, staff, stakeholders</li> </ul>
Program Context	Stakeholder expectations	<ul style="list-style-type: none"> <li>• Stakeholder opinions</li> </ul>	<ul style="list-style-type: none"> <li>• Stakeholder survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Stakeholders</li> </ul>
	Regulatory environment	<ul style="list-style-type: none"> <li>• Compliance faults</li> <li>• Policy analysis reports</li> </ul>	<ul style="list-style-type: none"> <li>• Compliance audit data</li> </ul>	<ul style="list-style-type: none"> <li>• Management, program records, governments</li> </ul>
	Contractual environment	<ul style="list-style-type: none"> <li>• Legal opinion</li> </ul>		<ul style="list-style-type: none"> <li>• Management, lawyers, sponsors, partners, franchisers</li> </ul>
Business Processes	Operations management	<ul style="list-style-type: none"> <li>• Staff input</li> <li>• Interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Accounting data</li> <li>• Operations data</li> </ul>	<ul style="list-style-type: none"> <li>• Program records, management, staff, instructors, regulators</li> </ul>
	Quality control	<ul style="list-style-type: none"> <li>• Quality control faults</li> <li>• Customer complaints</li> <li>• Staff input</li> </ul>	<ul style="list-style-type: none"> <li>• Accounting audit data</li> <li>• Quality control data</li> <li>• Customer survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, managers, regulators</li> </ul>
	Marketing	<ul style="list-style-type: none"> <li>• Interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Marketing data</li> </ul>	<ul style="list-style-type: none"> <li>• Management, staff, instructors, students, parents</li> </ul>
	Customer service	<ul style="list-style-type: none"> <li>• Focus group and interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Customer survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, staff, managers, regulators</li> </ul>
	Sustainability	<ul style="list-style-type: none"> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Accounting data</li> </ul>	<ul style="list-style-type: none"> <li>• Managers, regulators</li> </ul>
Program Standards	Benchmarking and certification	<ul style="list-style-type: none"> <li>• Benchmark failures</li> <li>• Certification faults</li> </ul>	<ul style="list-style-type: none"> <li>• Industry and jurisdiction survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Managers, regulators, industry members, jurisdiction representatives</li> </ul>
	Transportability of program	<ul style="list-style-type: none"> <li>• Policy analysis reports</li> </ul>	<ul style="list-style-type: none"> <li>• Operations data</li> </ul>	<ul style="list-style-type: none"> <li>• Managers, regulators</li> </ul>

Program Area	Evaluation Targets	Qualitative Data	Quantitative Data	Data Sources
Instructional Products	Curriculum materials	<ul style="list-style-type: none"> <li>• Content benchmarks</li> <li>• Focus group and Interview transcripts</li> <li>• Instructor log book entries</li> </ul>	<ul style="list-style-type: none"> <li>• Student survey responses</li> <li>• Instructor survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, managers, regulators</li> </ul>
	Tests and measurement	<ul style="list-style-type: none"> <li>• Expert opinion</li> </ul>	<ul style="list-style-type: none"> <li>• Pilot test data</li> <li>• Student survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, comparison groups, parents, instructors, managers, regulators</li> </ul>
Instructional Processes	Instructor preparation	<ul style="list-style-type: none"> <li>• Benchmarking</li> </ul>	<ul style="list-style-type: none"> <li>• Compliance audit data</li> <li>• Pilot test data</li> <li>• Records data</li> </ul>	<ul style="list-style-type: none"> <li>• Instructors, managers, regulators</li> </ul>
	Curriculum delivery and in-car practice	<ul style="list-style-type: none"> <li>• Observed actions</li> <li>• Instructor log book entries</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Parent and student survey responses</li> <li>• Instructor survey responses</li> <li>• Inspection report data</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, managers, regulators</li> </ul>
	Instructional facilities	<ul style="list-style-type: none"> <li>• Instructor feedback</li> <li>• Focus group and interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Parent and student survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, managers, regulators</li> </ul>
<b>Summative Evaluation</b>				
Student Outcomes	Knowledge outcomes	<ul style="list-style-type: none"> <li>• Focus group and interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Test scores</li> <li>• Survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, licensing authorities</li> </ul>
	Skill outcomes	<ul style="list-style-type: none"> <li>• Focus group and interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Test scores</li> <li>• Survey responses</li> <li>• Video observation data</li> <li>• Instrumented vehicle data</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents, instructors, licensing authorities</li> </ul>
	Motivation outcomes	<ul style="list-style-type: none"> <li>• Focus group and interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Video observation data</li> <li>• Survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Students, parents</li> </ul>
	Mobility outcomes	<ul style="list-style-type: none"> <li>• Focus group transcripts</li> <li>• Student diary entries</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Test scores</li> <li>• Survey responses</li> </ul>	<ul style="list-style-type: none"> <li>• Licensing authorities, students, parents</li> </ul>
	Behavioral outcomes	<ul style="list-style-type: none"> <li>• Focus group transcripts</li> <li>• Student diary entries</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Licensing test scores</li> <li>• Survey responses</li> <li>• Log book data</li> </ul>	<ul style="list-style-type: none"> <li>• Students, licensing authorities</li> </ul>

Program Area	Evaluation Targets	Qualitative Data	Quantitative Data	Data Sources
Social Impacts	Crash reduction impacts	<ul style="list-style-type: none"> <li>• Policy analysis reports</li> <li>• Case study data</li> </ul>	<ul style="list-style-type: none"> <li>• Survey responses</li> <li>• Driver records</li> <li>• Insurance records</li> </ul>	<ul style="list-style-type: none"> <li>• Licensing authorities, students, parents, insurers</li> </ul>
	Injury reduction impacts	<ul style="list-style-type: none"> <li>• Policy analysis reports</li> </ul>	<ul style="list-style-type: none"> <li>• Survey responses</li> <li>• Driver records</li> <li>• Insurance records</li> </ul>	<ul style="list-style-type: none"> <li>• Licensing authorities, students, parents, insurers</li> </ul>
	Socioeconomic impacts	<ul style="list-style-type: none"> <li>• Policy analysis reports</li> </ul>	<ul style="list-style-type: none"> <li>• Cost/benefit results</li> </ul>	<ul style="list-style-type: none"> <li>• Licensing authorities, insurers, stakeholders</li> </ul>
<b>Metaevaluation</b>				
Evaluation Quality	Evaluation effectiveness	<ul style="list-style-type: none"> <li>• Program Evaluation Standards benchmarks</li> <li>• Documentation of usefulness and relevance of findings</li> <li>• Peer review data</li> <li>• Stakeholder interview transcripts</li> </ul>	<ul style="list-style-type: none"> <li>• Meta-analysis results over different evaluations</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluators, managers, staff, stakeholders</li> </ul>

Using the evaluation targets, the corresponding data sources can now be identified. Be as specific as possible when determining the data needed, where they are located, and how or from whom they can be obtained. Data availability and accessibility must be considered when the type of evaluation is being decided. They should not be the only factors, however, that determine what data to collect. If the evaluation points to the need for a type of data that is not available, then finding a way to access the needed data can be integrated into the next evaluation cycle. It is also important to identify data that are essential to the evaluation in contrast to data that are “nice to know” but not critical.

➤ *Identify Indicators for Program Success*

Next, the success indicators associated with your evaluation targets are identified. These are the criteria that will be used to determine the effectiveness of your program. Examples that can be considered as the data collection plan is developed are shown in Table 18.

**Table 18. Examples of Indicators for Program Success**

Program Areas	Evaluation Targets	Indicators
<b>Formative Evaluation</b>		
Program Logic	User needs	<ul style="list-style-type: none"> <li>Needs linked to objectives</li> </ul>
	Program logic model or theory	<ul style="list-style-type: none"> <li>Objectives linked to content</li> </ul>
	Evaluability	<ul style="list-style-type: none"> <li>Program ready to be evaluated</li> </ul>
Program Context	Stakeholder expectations	<ul style="list-style-type: none"> <li>Expectations for program and evaluation addressed</li> </ul>
	Regulatory environment	<ul style="list-style-type: none"> <li>Constraints and obligations met</li> </ul>
	Contractual environment	<ul style="list-style-type: none"> <li>Obligations to sponsors, partners, franchisers clarified and met</li> </ul>
Business Processes	Operations management	<ul style="list-style-type: none"> <li>Staff retained</li> <li>Staff motivated with few concerns</li> <li>Operations efficient</li> </ul>
	Quality control	<ul style="list-style-type: none"> <li>Quality standards met</li> <li>Program consistent</li> </ul>
	Marketing	<ul style="list-style-type: none"> <li>Program reach optimized</li> <li>Participation rates stable or growing</li> </ul>
	Customer service	<ul style="list-style-type: none"> <li>Standards met</li> <li>Complaints addressed</li> </ul>
	Sustainability	<ul style="list-style-type: none"> <li>Positive cash flow, revenues increasing</li> <li>Growth acceptable</li> <li>Political and customer support</li> </ul>
Program Standards	Benchmarking and certification	<ul style="list-style-type: none"> <li>Applicable benchmarks met</li> <li>Certifiability requirements met</li> </ul>
	Transportability of program	<ul style="list-style-type: none"> <li>Potential for expansion in other jurisdictions established</li> </ul>
Instructional Products	Curriculum materials	<ul style="list-style-type: none"> <li>Efficacious</li> <li>User friendly</li> <li>Up-to-date</li> </ul>
	Tests and measurement	<ul style="list-style-type: none"> <li>Reliable</li> <li>Valid</li> <li>Practical to implement and assess</li> </ul>
Instructional Processes	Instructor preparation	<ul style="list-style-type: none"> <li>Benchmarks met</li> <li>Training quality control standards met</li> <li>Re-training standards met</li> </ul>
	Curriculum delivery and in-car practice	<ul style="list-style-type: none"> <li>Educational standards met</li> <li>Course completion and pass rates met</li> </ul>
	Instructional facilities	<ul style="list-style-type: none"> <li>Customer satisfaction</li> <li>Staff satisfaction</li> <li>Safety standards met</li> </ul>

Program Areas	Evaluation Targets	Indicators
<b>Summative Evaluation</b>		
Student Outcomes	Knowledge outcomes	<ul style="list-style-type: none"> <li>• Knowledge gain targets met</li> </ul>
	Skill outcomes	<ul style="list-style-type: none"> <li>• Skill acquisition targets met</li> </ul>
	Motivation outcomes	<ul style="list-style-type: none"> <li>• Attitude change targets met</li> </ul>
	Mobility outcomes	<ul style="list-style-type: none"> <li>• Targets for course pass and license rates met</li> </ul>
	Behavioral outcomes	<ul style="list-style-type: none"> <li>• Exposure to risk limited</li> </ul>
Social Impacts	Crash reduction impacts	<ul style="list-style-type: none"> <li>• Targets for crash rates met</li> </ul>
	Injury reduction impacts	<ul style="list-style-type: none"> <li>• Targets for casualty rates met</li> </ul>
	Socioeconomic impacts	<ul style="list-style-type: none"> <li>• Cost/benefit effects positive</li> <li>• Other economic effects positive</li> <li>• Side effects minimized</li> </ul>
<b>Metaevaluation</b>		
Evaluation Quality	Evaluation effectiveness	<ul style="list-style-type: none"> <li>• Program Evaluation Standards met</li> <li>• Evaluation results useful and relevant</li> <li>• Stakeholder expectations met</li> </ul>

➤ *Assess Feasibility of Data Collection Plan*

The feasibility of collecting the data identified in the data collection plan must be carefully considered. It would be rare for an evaluator to be able to collect as much or as many different kinds of data as desired. It is better to keep initial evaluations simple and manageable. Smaller amounts of high-quality data are usually more valuable than lots of low-quality data. Well-defined indicators will help ensure that high-quality, reliable, and useful data are collected.

Data quality depends on a number of factors. Most have to do with clarity and confidence that: 1) the sampling and assignment were carried out as planned; 2) the actual data are reasonably close to the data that were intended to be obtained; 3) the data have been handled carefully to maintain integrity; and 4) missing data and sample attrition have been accounted for. Other factors that can affect data quality include data collection tool design, training of data collectors, data source selection, data coding and entering, data management, and routine error checking. Identify actions that are being planned or implemented to manage these factors in the data collection plan. These actions will influence the activities to be undertaken next in Step 3B.

## STEP **3B** SELECT AND ASSESS DATA COLLECTION TOOLS

### Overview

A wide range of data collection or measurement tools can be used to produce and gather data for driver education evaluations. As discussed in Step 2, they can be: 1) qualitative tools such as reviews, checklists, diaries, open-ended interviews, and focus groups; and 2) quantitative tools such as records, observation logs, tests, and surveys.

Different tools are appropriate for different types of evaluations. Formative evaluations use both qualitative and quantitative tools, while summative evaluations require quantitative tools. Qualitative tools, however, can be helpful in developing background data and in generating ideas for research questions for quantitative studies. As the following chart indicates, once the tools are selected, modified, or developed, it is essential to assess their quality before beginning the data collection phase of the evaluation.

3A. DEVELOP DATA COLLECTION PLAN	<b>3B</b> SELECT AND ASSESS DATA COLLECTION TOOLS	3C. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Determine appropriate data types and data gathering methods</li> <li>➤ Specify data and sources</li> <li>➤ Identify indicators for program success</li> <li>➤ Assess feasibility of data collection plan</li> </ul>	<ul style="list-style-type: none"> <li>➤ Select, modify or develop tools</li> <li>➤ Conduct quality assessment of tools and revise</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

### Activities

#### ➤ *Select, Modify or Develop Tools*

There are relatively few data collection tools available specifically for driver education evaluations. Examples of existing tools are found in Appendices D-F. In many cases,

items or measures from different tools can be combined to create a new tool that meets the needs of a specific evaluation.

Ideally, there will eventually be well-established, reliable, and valid standardized tests and other tools, with established norms for comparison, but these still need to be developed. Nevertheless, there are benefits to using existing tools where possible. Because they have been used before, information may be available on their validity and reliability. They also may have been revised and improved over time, which means they will generate higher quality data.

### Help with Developing New Tools for Driver Education Evaluation

To find help with developing new tools or modifying existing ones, talk to colleagues and contact any driver education evaluators that you are aware of. Research organizations that may have information on driver education evaluation tools include:

- AAA Foundation for Traffic Safety (AAAFTS)
- American Driver and Traffic Safety Education Association (ADTSEA)
- National Highway Traffic Safety Administration (NHTSA)
- Texas Transportation Institute (TTI)
- Traffic Injury Research Foundation (TIRF)
- Transportation Research Board (TRB)
- University of Michigan Transportation Research Institute (UMTRI)
- University of North Carolina Highway Safety Research Center (UNC-HSRC)

Refer to the Evaluation Resources section on page 173 for the web addresses of these organizations. The checklist in Table 19 can be used to help determine the type of tools to look for. Review existing tools to make sure they will collect all the data needed for your evaluation. Also review the reliability and validity of the tools to ensure that, if used, they will generate high-quality information.

**Table 19. Data Collection Tools Checklist**

Program Area	Evaluation Targets	Qualitative Tools	Quantitative Tools
<b>Formative Evaluation</b>			
Program Logic	User needs	<input type="checkbox"/> Needs assessment <input type="checkbox"/> Critical review of research literature <input type="checkbox"/> Interview guide <input type="checkbox"/> Focus group guide	<input type="checkbox"/> Problem definition research <input type="checkbox"/> Meta-analysis
	Program logic model	<input type="checkbox"/> Documentation of logical program linkages <input type="checkbox"/> Logic model <input type="checkbox"/> Content analysis checklist	
	Evaluability	<input type="checkbox"/> Document review checklist <input type="checkbox"/> Interview guide	
Program Context	Stakeholder expectations	<input type="checkbox"/> Interview guide <input type="checkbox"/> Workshop format	<input type="checkbox"/> Questionnaires
	Regulatory environment	<input type="checkbox"/> Policy analysis <input type="checkbox"/> Compliance fault checklist	<input type="checkbox"/> Business and audit records <input type="checkbox"/> Cross-jurisdictional studies
	Contract environment	<input type="checkbox"/> Legal review	
Business Processes	Operations management	<input type="checkbox"/> Log books <input type="checkbox"/> Interviews <input type="checkbox"/> Staff meetings	<input type="checkbox"/> Accounting records <input type="checkbox"/> Operations records
	Quality control	<input type="checkbox"/> Log books <input type="checkbox"/> Interview guide <input type="checkbox"/> Certification checklist	<input type="checkbox"/> Operations records <input type="checkbox"/> Questionnaires
	Marketing	<input type="checkbox"/> Interview guide <input type="checkbox"/> Focus group guide	<input type="checkbox"/> Business records <input type="checkbox"/> Marketing reports
	Customer Service	<input type="checkbox"/> Interview guide <input type="checkbox"/> Focus group guide	<input type="checkbox"/> Questionnaires
	Sustainability	<input type="checkbox"/> Case study	<input type="checkbox"/> Accounting and business records
Program Standards	Benchmarking and certification	<input type="checkbox"/> Benchmark checklist <input type="checkbox"/> Certification checklist	<input type="checkbox"/> Questionnaires
	Transportability of program	<input type="checkbox"/> Policy analysis	<input type="checkbox"/> Business records
Instructional Products	Curriculum materials	<input type="checkbox"/> Content analysis checklist <input type="checkbox"/> Benchmark checklist <input type="checkbox"/> Interview guide <input type="checkbox"/> Focus group guide <input type="checkbox"/> Log books	<input type="checkbox"/> Questionnaires
	Tests and measurement	<input type="checkbox"/> Expert review	<input type="checkbox"/> Pilot testing <input type="checkbox"/> Questionnaires

Program Area	Evaluation Targets	Qualitative Tools	Quantitative Tools
Instructional Processes	Instructor preparation	<input type="checkbox"/> Benchmark checklist	<input type="checkbox"/> Records
	Curriculum delivery and in-car practice	<input type="checkbox"/> Observation guide <input type="checkbox"/> Log books <input type="checkbox"/> Case study	<input type="checkbox"/> Questionnaires <input type="checkbox"/> Records <input type="checkbox"/> Observation guide
	Instructional facilities	<input type="checkbox"/> Benchmark checklist <input type="checkbox"/> Interview guide <input type="checkbox"/> Focus group guide	<input type="checkbox"/> Questionnaires
<b>Summative Evaluation</b>			
Student Outcomes	Knowledge outcomes	<input type="checkbox"/> Focus group guide <input type="checkbox"/> Interview guide	<input type="checkbox"/> Tests <input type="checkbox"/> Questionnaires
	Skill outcomes	<input type="checkbox"/> Focus group guide <input type="checkbox"/> Interview guide	<input type="checkbox"/> Tests <input type="checkbox"/> Questionnaires <input type="checkbox"/> Video observation guide <input type="checkbox"/> Instrumented vehicle records
	Motivation outcomes	<input type="checkbox"/> Focus group guide <input type="checkbox"/> Interview guide	<input type="checkbox"/> Video observation guide <input type="checkbox"/> Questionnaires
	Mobility outcomes	<input type="checkbox"/> Focus group guide <input type="checkbox"/> Student diaries <input type="checkbox"/> Case study	<input type="checkbox"/> Tests <input type="checkbox"/> Questionnaires <input type="checkbox"/> Log books
	Behavioral outcomes	<input type="checkbox"/> Focus group guide <input type="checkbox"/> Student diaries <input type="checkbox"/> Case study	<input type="checkbox"/> Tests <input type="checkbox"/> Questionnaires
Social Impacts	Crash reduction impacts	<input type="checkbox"/> Policy analysis <input type="checkbox"/> Case study	<input type="checkbox"/> Questionnaires <input type="checkbox"/> Driver records <input type="checkbox"/> Insurance records
	Injury reduction impacts	<input type="checkbox"/> Policy analysis	<input type="checkbox"/> Questionnaires <input type="checkbox"/> Driver records <input type="checkbox"/> Insurance records
	Socioeconomic impacts	<input type="checkbox"/> Policy analysis	<input type="checkbox"/> Cost/benefit analysis tools
<b>Metaevaluation</b>			
Evaluation Quality	Evaluation effectiveness	<input type="checkbox"/> Program Evaluation Standards checklists <input type="checkbox"/> Critical review of evaluation for usefulness and relevance <input type="checkbox"/> Peer review guide <input type="checkbox"/> Stakeholder interview guide	<input type="checkbox"/> Meta-analysis over different evaluations

The evaluation team may decide to develop some tools to supplement existing ones. Table 19 can also help determine which tools to develop. Some important technical matters, however, should be kept in mind if you are going to develop new tools. The process starts with identifying the main categories of information that the tool is going to address. For example, if increased students' knowledge of the "rules of the road" is one of the evaluation targets, then a tool will obviously include questions to determine whether the students have learned and remembered this information from the course. A written questionnaire is an appropriate tool to collect this type of information. Similarly, a student satisfaction questionnaire will ask students to rate the course and ask for more in-depth information about their views on specific aspects of the program, such as quality of instructional materials and teaching.

Developing good questions to be used in an evaluation tool is important and not always easy. How questions are worded influences the answers, and thus, the quality and meaningfulness of the evaluation data. The following suggestions can help develop questions that will provide high-quality data.

### Developing Effective Questions

- Pay attention to the language in each question, and the literacy levels of the people from whom the data are being collected. Use simple, unambiguous, and familiar words.
- Use standard wording used by other evaluations for common information such as demographics.
- Make sure the questions are short, straightforward, and direct.
- Consider the need for translation of the data collection tools into other languages.
- Don't use judgmental language.
- Watch for common "pitfalls" of asking questions, such as leading, loaded, and double-barreled questions; asking more than one question at a time; using technical terms or jargon; and using slang or acronyms.

Adapted from *A Program Evaluation Tool Kit*, Porteous, Sheldrick, and Stewart 1997.

Once the questions for each data collection tool are drafted and refined, the tool can be organized. Put questions with common themes or intent together to help the flow of the questions, and consider the tool's appearance. Examples of other evaluation tools from driver education evaluations and other fields can be used as references. Also consider details such as the content of the introduction, an easy-to-read font for questionnaires, not splitting questions over pages, and estimating completion time.

➤ *Conduct Quality Assessment of Tools and Revise*

The best way to know how well the data collection tools will work is to assess them before finalizing and actually using them "in the field." Three of the most common ways to do this are:

1. Ask experts, such as driver education and road safety experts, to review the tools' content. If possible, identify 2 or 3 experts and ask them to review the data collection tools using the questions in Table 20 as a guide. Keep in mind that you may get as many different opinions as you have experts.
2. Pilot test the tools by locating a small number (~10) of people who are representative of the sample for each tool, and ask them to complete it and then answer questions about it. Again Table 20 provides sample questions.
3. A more sophisticated test, referred to as a "stability reliability" test, can also be considered if the resources and time are available (Porteous, Sheldrick, and Stewart 1997). This test examines the consistency of a measure, meaning the likelihood that the measure results in consistent data over time. For example, a sample of the target group who are asked to complete a data collection tool, say a questionnaire, will do so twice. The second time will be weeks, or even months, after the first. If the two sets of data are similar, then the tool is reliable. Significantly dissimilar results mean that there are problems with the tool's reliability. In such a case, the tool should be revised and re-tested before being used. Determining sample sizes and analysis procedures for the test will require assistance from an evaluation specialist or researcher.

**Table 20. Questions to Help Assess the Quality of Data Collection Tools**

Assessment	Sample Questions
<b>Expert Review</b>	<ul style="list-style-type: none"> <li>• Are there questions missing?</li> <li>• Are any unnecessary items included in the tool?</li> <li>• Are the questions clear, and is the language straightforward?</li> <li>• Would you recommend any format or design changes?</li> <li>• Are the response categories appropriate?</li> <li>• Do you think the tool measures what it is supposed to measure?</li> <li>• Is it a reasonable length?</li> <li>• Is the tool culturally appropriate?</li> </ul>
<b>Pilot Test</b>	<ul style="list-style-type: none"> <li>• Could you understand all the questions?</li> <li>• Were you able to answer all the questions? Which did you have trouble with and why?</li> <li>• Are the questions clear, and is the language straightforward?</li> <li>• Were you able to follow the instructions?</li> <li>• What is the quality of the translation? (if appropriate)</li> <li>• Is the tool easy to read and follow (e.g., font size, order of questions and skips, format)?</li> <li>• Did you get bored as you were completing the tool? Where?</li> </ul>

Use the information from these checks to revise the data collection tools, and begin the preparations for data gathering. Tracking quality during data collection ensures that problems do not develop.

## STEP **3C** APPLY EVALUATION STANDARDS

### Overview

Again, be sure to review the evaluation standards using the following checklist, and keep in mind the importance of ensuring that they are met.

3A. DEVELOP DATA COLLECTION PLAN	3B. SELECT AND ASSESS DATA COLLECTION TOOLS	<b>3C</b> APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Determine appropriate data types and data gathering methods</li> <li>➤ Specify data and sources</li> <li>➤ Identify indicators for program success</li> <li>➤ Assess feasibility of data collection plan</li> </ul>	<ul style="list-style-type: none"> <li>➤ Select, modify or develop tools</li> <li>➤ Conduct quality assessment of tools and revise</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

## Activities

Use the checklist below to ensure that the Step 3 standards are applied, and then move on to Step 4.



### Evaluation Standards Checklist

Step 3 Standards	Step 3 Questions
<b>UTILITY</b> Information scope and selection	<input type="checkbox"/> Are the data collection tools clearly linked to the data collection methods used to ensure they are related to the evaluation questions as well as user and stakeholder needs?
<b>FEASIBILITY</b> Practical procedures	<input type="checkbox"/> Are the tools practical so that disruption to daily activities of participants is minimized?
Cost effectiveness	<input type="checkbox"/> Have the tools been designed to ensure efficiency, the value of the information, and justification of the expended resources (staff and money)?
<b>PROPRIETY</b> Service orientation	<input type="checkbox"/> Have the tools been designed to assist with addressing the needs of the target groups and stakeholders?
<b>ACCURACY</b> Defensible information sources	<input type="checkbox"/> Have the purposes and procedures related to the data collection tools been described in enough detail that they can be identified, monitored, and later assessed? <input type="checkbox"/> Have the data sources been described in enough detail to assess the adequacy of the information?
Valid information	<input type="checkbox"/> Have the tools been chosen or developed to ensure a valid interpretation for the evaluation's intended use?
Reliable information	<input type="checkbox"/> Have the tools been chosen or developed to ensure sufficiently reliable information for the evaluation's intended use?



# STEP 4

## GATHER, ANALYZE, AND SUMMARIZE DATA

With the completion of Step 3, the data collection tools are now finalized, and it is time to begin preparing to collect the data and proceed with the data gathering and analysis. Step 4 includes the following activities.

### Summary of Activities

4A DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	4B GATHER AND ENTER DATA	4C ANALYZE AND SUMMARIZE DATA	4D APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Develop data collection logistics plan</i></li> <li>➤ <i>Develop procedures to train data collection personnel and conduct training</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Ensure timely and consistent data collection</i></li> <li>➤ <i>Enter data and ensure accuracy</i></li> <li>➤ <i>Ensure confidentiality and security of data</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify data analysis procedures and conduct data analysis</i></li> <li>➤ <i>Assess, synthesize, and summarize data analysis results</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## STEP 4A DEVELOP DATA COLLECTION LOGISTICS PLAN AND TRAINING PROCEDURES

### Overview

The details of how the data are going to be gathered must now be organized. The evaluation team needs to determine when, where, and who will collect the data. The following chart identifies the activities involved in this step of the evaluation.

**4A**

DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	4B. GATHER AND ENTER DATA	4C. ANALYZE AND SUMMARIZE DATA	4D. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Develop data collection logistics plan</i></li> <li>➤ <i>Develop procedures to train data collection personnel and conduct training</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Ensure timely and consistent data collection</i></li> <li>➤ <i>Enter data and ensure accuracy</i></li> <li>➤ <i>Ensure confidentiality and security of data</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify data analysis procedures and conduct data analysis</i></li> <li>➤ <i>Assess, synthesize, and summarize data analysis results</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Develop Data Collection Logistics Plan*

A logistics plan should be developed to provide a guide for the evaluation team as the data collection proceeds. Table 21 provides an example of the types of activities and information to include in the plan.

This plan helps track progress and keep data collection activities on schedule. If data are going to be collected from people outside the program, such as parents or community organizations, decisions about how to reach them should be made as part of this activity.

A logistics worksheet is provided in Appendix E. Checking and updating the plan regularly can help the evaluation team accommodate changes as data collection proceeds. A separate plan for each type of evaluation tool can be developed if warranted by the scale of the evaluation. This plan assumes that all Step 3 tasks have been completed, and that the content and format of the data collection tools have been determined and agreed upon.

The data collection procedures that ensure appropriate ethics and rights of human subjects should also be verified as part of Step 4. As explained in Step 2, ensuring privacy, confidentiality, and ethical procedures as the data are collected is essential to every evaluation. Often the law or organizational or program standards require these procedures. No evaluation should proceed without the appropriate procedures in place. As a safeguard, the evaluation standards for this step include rights of human subjects and human interactions standards (see Appendix C).

**Table 21. Data Collection Logistics Plan**

Activity	Details	Suggested Timing	Who is Responsible
Prepare for and schedule activities required to collect data	<ul style="list-style-type: none"> <li>• Set up meetings with evaluation target groups to explain, schedule evaluation, and recruit participants</li> <li>• Organize participation in focus groups</li> <li>• Obtain addresses/phone numbers for surveys</li> <li>• Determine telephone interview/mail-out questionnaire schedules</li> <li>• Consult with records data personnel to determine access and timing</li> <li>• Prepare overall schedule for data collection activities, and obtain agreement</li> </ul>	8 weeks prior to initiation of data collection	Evaluation team member responsible for data collection
Verify ethics and human rights procedures	<ul style="list-style-type: none"> <li>• Verify procedures appropriate to data collection tools, such as informed consent forms and provisions for confidentiality and privacy</li> </ul>	8 weeks prior to initiation of data collection	Evaluation team member responsible for data collection
Organize tools and plan for data entry	<ul style="list-style-type: none"> <li>• Identify tools (e.g., assign ID numbers on questionnaires and interview forms)</li> </ul>	4 weeks prior to initiation of data collection	Assigned evaluation team or staff person
Order supplies	<ul style="list-style-type: none"> <li>• Obtain pens, pencils, envelopes, labels, paper, reminder postcards, postage</li> </ul>	4 weeks prior to initiation of data collection	Assigned evaluation team or staff person
Produce addressed materials	<ul style="list-style-type: none"> <li>• If conducting a mail survey, produce labels, envelopes, covering and reminder letters with appropriate salutations and addresses</li> </ul>	4 weeks prior to initiation of data collection	Assigned evaluation team or staff person
Produce data collection tools	<ul style="list-style-type: none"> <li>• Make copies of data collection tools, such as questionnaires, interview guides, focus group formats, data recording forms, covering letters, informed consent forms, reminder letters</li> </ul>	4 weeks prior to initiation of data collection	Assigned evaluation team or staff person

➤ *Develop Procedures to Train Data Collection Personnel and Conduct Training*

With data collection logistics in place, the individuals who are going to administer the data collection activities involving direct contact with target groups must be trained to follow predetermined data collection formats. These are the data collectors who will

moderate focus groups, conduct telephone interviews, administer questionnaires, or take direct observations. For some of these tasks, a trained evaluator is required. For example, in-car observations require an experienced researcher or driver evaluator. But for many tasks, program staff can be trained as the data collectors.

Training of data collectors is very important. The reason for this is that each tool should be administered exactly the same way each time it is used. In sound evaluation, every effort is made to ensure that all participants receive the same information about the evaluation and the same instructions for completing the tool. Participants should also have as similar an understanding as possible of each question in, for example, an interview or a questionnaire. Data collectors must be clear that improvisation and interpretation of the information they provide to participants is not acceptable. Developing a training guide to assist data collectors and procedures to train personnel will ensure that this does not happen.

Next, training sessions on how to administer the data collection sessions and collect the data should be conducted. Role playing for focus group moderators and interviewers, as well as practice sessions for those administering surveys could be included. These activities can also be used to generate sensitivity and positive attitudes amongst the people who will be taking on this important task.

## STEP **4B** GATHER AND ENTER DATA

### Overview

As the following chart indicates, data gathering is the next step and involves the tasks of obtaining the data, entering the data, and ensuring its confidentiality and security.

4A. DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	<b>4B</b> GATHER AND ENTER DATA	4C. ANALYZE AND SUMMARIZE DATA	4D. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Develop data collection logistics plan</i></li> <li>➤ <i>Develop procedures to train data collection personnel and conduct training</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Ensure timely and consistent data collection</i></li> <li>➤ <i>Enter data and ensure accuracy</i></li> <li>➤ <i>Ensure confidentiality and security of data</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify data analysis procedures and conduct data analysis</i></li> <li>➤ <i>Assess, synthesize, and summarize data analysis results</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Ensure Timely and Consistent Data Collection*

The data collection logistics plan will help ensure that data collection activities are well organized, carefully scheduled, and administered in a timely manner. The evaluation team member responsible for data collection should keep track of the schedule for data collection activities, such as interview appointments, focus group meetings, and questionnaire completions. Setting realistic completion dates for the data collection is important as well as appointing someone to be responsible for tracking progress and deadlines.

### ➤ *Enter Data and Ensure Accuracy*

Converting the data that have been collected into a format appropriate to the type of analysis selected for the evaluation is very important. First the data have to be transferred from the completed data collection forms into a format that is useable for analysis purposes. Answers from a written questionnaire, for example, must be assigned codes before they can be transferred into a format that allows them to be analyzed. Recorded responses from individual interviews and focus groups have to be transcribed, and the data collected in a records study have to be checked for errors and cleaned to ensure their accuracy.

Quantitative data can be entered into spreadsheet or analysis software; for example, multiple choice answers on a questionnaire can be entered into software that will count the number of responses and calculate the percentages in each category. Make sure the information is converted carefully and accurately. The person entering the data must check to ensure that no mistakes have been made.

### ➤ *Ensure Confidentiality and Security of Data*

Actions to ensure the confidentiality of the data must be in place as data collection tools are being developed and carefully implemented while data are being gathered. The data collectors need to obtain signed informed consent forms, if required, to communicate privacy, confidentiality, and security procedures, and manage the data to ensure confidentiality and security. Any documents or data files with personal identifiers or other confidential information must be stored and eventually disposed of securely.

# STEP **4C** ANALYZE AND SUMMARIZE DATA

## Overview

Data analysis is the process of compiling or aggregating your data and figuring out what they mean. This analysis involves systematically applying logical and statistical techniques to describe, summarize, and compare the data. Different types of data require different analysis approaches and techniques. The evaluation methods selected in Step 2 will direct decisions about which types of analysis to use and which processes need to be in place to conduct appropriate and sound analyses. The tasks in this step are identified in the following chart.

4A. DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	4B. GATHER AND ENTER DATA	<b>4C</b> ANALYZE AND SUMMARIZE DATA	4D. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Develop data collection logistics plan</i></li> <li>➤ <i>Develop procedures to train data collection personnel and conduct training</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Ensure timely and consistent data collection</i></li> <li>➤ <i>Enter data and ensure accuracy</i></li> <li>➤ <i>Ensure confidentiality and security of data</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify data analysis procedures and conduct data analysis</i></li> <li>➤ <i>Assess, synthesize, and summarize data analysis results</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Identify Data Analysis Procedures and Conduct Data Analysis*

Several factors determine the types of data analysis to be undertaken. These factors include the types of evaluation being implemented (formative, summative, metaevaluation); methods being used (qualitative and quantitative); and the type of data collected (interview, survey, record data). Resources available for data analysis and the analytical capability of the evaluation team are also important considerations. How the analysis is carried out is determined by several factors, the most important being: 1) which types of analysis have been agreed upon; 2) who is going to do the data analysis; and 3) what in-house resources are available. Table 22 provides a summary of the range of possible analysis procedures.

**Table 22. Examples of Data Analysis Procedures for Qualitative and Quantitative Data**

Data Category	Data Type	Analysis Options
<p><b>Qualitative</b></p>	<p>Individual interviews</p>	<ul style="list-style-type: none"> <li>• Identification of response categories, trends, and themes</li> <li>• Documentation of detailed descriptive information that helps address specific program issues</li> <li>• Qualitative software analysis</li> </ul>
	<p>Focus groups</p>	<ul style="list-style-type: none"> <li>• Identification of response categories, trends, and themes</li> <li>• Documentation of detailed descriptive information that helps address specific program issues</li> <li>• Qualitative software analysis</li> </ul>
<p><b>Quantitative</b></p>	<p>Mail-out or telephone survey</p>	<ul style="list-style-type: none"> <li>• Descriptive statistics—counts, differences, cross-tabulations, averages</li> <li>• Chi square analysis</li> <li>• Correlational analysis</li> <li>• t-Tests</li> <li>• Analysis of variance</li> <li>• Factor analysis</li> <li>• Regression analysis</li> </ul>
	<p>In-class survey</p>	<ul style="list-style-type: none"> <li>• Descriptive statistics—counts, differences, cross-tabulations, averages</li> <li>• Correlational analysis</li> <li>• t-Tests</li> <li>• Analysis of variance</li> <li>• Factor analysis</li> <li>• Regression analysis</li> </ul>
	<p>Observations</p>	<ul style="list-style-type: none"> <li>• Descriptive statistics</li> <li>• Chi square analysis</li> <li>• t-Tests</li> <li>• Analysis of variance</li> <li>• Correlational analysis</li> </ul>
	<p>Records</p>	<ul style="list-style-type: none"> <li>• Descriptive statistics</li> <li>• Chi square analysis</li> <li>• Correlational analysis</li> <li>• Regression analysis</li> <li>• Multi-variate models</li> <li>• Survival analysis</li> </ul>

In many quantitative and some qualitative data sets, there will be too much data to handle the analyses manually. Spreadsheet software, such as Microsoft Excel, is relatively user-friendly, does not require a lot of training, and will be adequate for many evaluation plans. With larger evaluations, however, more specialized data analysis software will be required to conduct the analysis. If no one on the evaluation team is familiar with computer-assisted analysis, seek assistance from an evaluation specialist, a statistician, or a graduate student who has expertise in health or social science research methods.

➤ *Assess, Synthesize, and Summarize Data Analysis Results*

Findings resulting from the data analysis depend on the types of data analyses undertaken. Once the data analysis is completed, decisions about how to organize, clas-

**F**ull details of the analysis are important to the evaluation team, but they should be summarized and displayed in straightforward, clear, and easily understandable formats for other audiences.

sify, compare, and display the information need to be made. If more than one evaluation method has been used, different sets of information will need to be assessed, combined, or synthesized. Common themes to look for include what the different sets of information show, whether they support or contradict each other, and whether any of the findings are surprising or unexpected.

Full details of the analysis are important to the evaluation team, but they should be summarized and displayed in straightforward, clear, and easily understandable formats for other audiences. Charts, bar graphs, and histograms help make the findings easily understandable by the widest possible audience.



**STEP 4D APPLY EVALUATION STANDARDS**

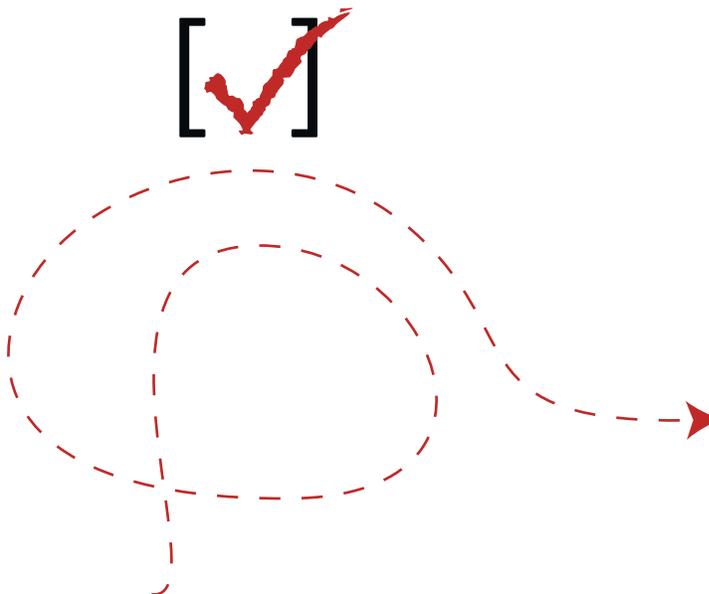
**Overview**

Again, be sure to review the evaluation standards, and keep in mind the importance of making sure they are met.

4A. DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	4B. GATHER AND ENTER DATA	4C. ANALYZE AND SUMMARIZE DATA	4D APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Develop data collection logistics plan</i></li> <li>➤ <i>Develop procedures to train data collection personnel and conduct training</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Ensure timely and consistent data collection</i></li> <li>➤ <i>Enter data and ensure accuracy</i></li> <li>➤ <i>Ensure confidentiality and security of data</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Identify data analysis procedures and conduct data analysis</i></li> <li>➤ <i>Assess, synthesize, and summarize data analysis results</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

As the work in this step proceeds, keep the questions in the following checklist in mind. They will assist in understanding and achieving the standards and ensure they are an integral part of the evaluation process.



**STEP  
4**



**Evaluation Standards Checklist**

Step 4 Standards	Step 4 Questions
<b>UTILITY</b> Information scope and selection	<input type="checkbox"/> Will the data analysis address the key evaluation questions and be responsive to the needs and interests of program users and stakeholders?
<b>FEASIBILITY</b> Practical procedures	<input type="checkbox"/> Are the data collection procedures practical so that disruption of daily activities of participants is minimized?
<b>PROPRIETY</b> Rights of human subjects	<input type="checkbox"/> Have the data collection procedures been designed to respect and protect the rights and welfare of participants, and are the necessary procedures in place?
Human interactions	<input type="checkbox"/> Do the data collection procedures respect human dignity and worth, to ensure that participants are not threatened or harmed?
<b>ACCURACY</b> Valid and reliable information	<input type="checkbox"/> Do the data collection procedures address internal validity and reliability issues?
Systematic information	<input type="checkbox"/> Is there a system in place to identify and check for errors in data entry?
Analysis of quantitative and qualitative information	<input type="checkbox"/> Have the quantitative and qualitative data been appropriately and systematically analyzed so that the evaluation questions are effectively answered?

# STEP 5

## INTERPRET AND ACT UPON THE EVALUATION FINDINGS

Step 5 initiates the evaluation activities that link the findings of the data analysis to decision making and future program changes. In this step, all the effort that has been put into conducting a good evaluation comes together and makes a difference to the program being evaluated. Each evaluation can also make a contribution to driver education evaluation in general. Other program managers and evaluators will want to know about your evaluation and what happened next. Step 5 includes the following activities.

### Summary of Activities

5A INTERPRET AND DOCUMENT FINDINGS	5B MAKE RECOMMENDATIONS AND TAKE ACTION	5C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Interpret findings</i></li> <li>➤ <i>Prepare conclusions and make judgments</i></li> <li>➤ <i>Document evaluation process and findings in evaluation report</i></li> <li>➤ <i>Undertake peer review</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Prepare recommendations</i></li> <li>➤ <i>Ensure feedback, follow-up, and dissemination of evaluation results</i></li> <li>➤ <i>Undertake actions to ensure use of evaluation and share lessons learned</i></li> <li>➤ <i>Determine changes to implement in next evaluation cycle and prepare action plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## STEP 5A INTERPRET AND DOCUMENT THE EVALUATION FINDINGS

### Overview

With the results of the data analysis complete, it is time to assess and synthesize these results and decide what they mean. Then the conclusions of the evaluation will be determined, and judgments about the value of the findings related to program standards can

be made. The evaluation process and findings should be documented in a report from which recommendations can be generated. You may also want to ask an outside expert to provide a preliminary review of the findings, or undertake a peer review by qualified outside experts to assess the evaluation and identify initial interpretations or limitations. The following chart identifies the activities involved in this evaluation step.

<b>5A</b> <b>INTERPRET AND DOCUMENT FINDINGS</b>	<b>5B. MAKE RECOMMENDATIONS AND TAKE ACTION</b>	<b>5C. APPLY EVALUATION STANDARDS</b>
<ul style="list-style-type: none"> <li>➤ <i>Interpret findings</i></li> <li>➤ <i>Prepare conclusions and make judgments</i></li> <li>➤ <i>Document evaluation process and findings in evaluation report</i></li> <li>➤ <i>Undertake peer review</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Prepare recommendations</i></li> <li>➤ <i>Ensure feedback, follow-up, and dissemination of evaluation results</i></li> <li>➤ <i>Undertake actions to ensure use of evaluation and share lessons learned</i></li> <li>➤ <i>Determine changes to implement in next evaluation cycle and prepare action plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Interpret Findings*

Depending on the type of evaluation conducted, the evaluation data and the approach used to interpret the findings can be quite different. Quantitative data from a record study, for example, may have been analyzed using sophisticated statistical procedures. To understand what these data mean, assistance from an experienced evaluator or researcher will be needed. An in-house evaluator or program staff person will most likely be more comfortable interpreting results obtained from smaller-scale evaluations, such as a student survey or a series of parent interviews. Questions to ask as the findings are interpreted include:

- Does the new information help answer our evaluation questions, and what do we know about our evaluation targets that we didn't know before?
- What new information is available about the program's products and processes, outcomes, and possibly impacts?

### ➤ *Prepare Conclusions and Make Judgments*

With answers to these questions in mind, document the conclusions about the outcomes of the evaluation and the implications for the program. The conclusions can also be reviewed with stakeholders. It is important to judge the usefulness of the evidence that has been gathered through the evaluation. Program and evaluation staff should be clear about justifying their conclusions. Consider alternative explanations for evaluation findings, and determine whether there is evidence to support them. Reaching different, but equally supported, conclusions is not unusual, and you should assess each conclusion by examining its strengths and weaknesses.

### ➤ *Document Evaluation Process and Findings in an Evaluation Report*

Although writing an evaluation report may seem unnecessary to those directly involved in the evaluation, it is an essential part of the process. It provides a complete overview of how the evaluation was conceived, implemented, and concluded, and should document:

- Who was consulted;
- What decisions were made about how the evaluation was to be conducted;
- Who was involved;
- How the evaluation was carried out;
- What the data analysis consisted of; and
- What the findings and conclusions were.

The report is the principal source of all the data analysis, assessment, and synthesis information, as well as evaluation conclusions and subsequent decision-making. It will also help with future evaluations, which can build upon and further improve the quality and usefulness of evaluations for improving programs.

### ➤ *Undertake Peer Review*

Depending on the type of evaluation conducted, submitting the draft evaluation report for peer review may be appropriate. It might be organized to address the specific project, or a report on the project could be submitted to a peer-reviewed research journal in transportation, evaluation, education, social sciences, or health.

# STEP **5B** MAKE RECOMMENDATIONS AND TAKE ACTION

## Overview

This step, outlined in the following chart, focuses on what happens after the evaluation process and results are documented. This is one of the most important aspects of the entire evaluation. It is here that critical decisions about the implications of the evaluation for the program are made, and actions about what should happen as a result of the findings are determined. Disseminating information about the evaluation and sharing lessons learned increase awareness and communication about improving driver education evaluations and programs. Creating an action plan ensures that the appropriate changes are incorporated into the next evaluation cycle and that next steps are identified and implemented.

5A. INTERPRET AND DOCUMENT FINDINGS	<b>5B</b> MAKE RECOMMENDATIONS AND TAKE ACTION	5C. APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Interpret findings</i></li> <li>➤ <i>Prepare conclusions and make judgments</i></li> <li>➤ <i>Document evaluation process and findings in evaluation report</i></li> <li>➤ <i>Undertake peer review</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Prepare recommendations</i></li> <li>➤ <i>Ensure feedback, follow-up, and dissemination of evaluation results</i></li> <li>➤ <i>Undertake actions to ensure use of evaluation and share lessons learned</i></li> <li>➤ <i>Determine changes to implement in next evaluation cycle and prepare action plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

### ➤ *Prepare Recommendations*

With the evaluation results documented, recommendations can now be developed. These represent decisions to be made and actions to consider as a result of the evaluation. Draft recommendations should be available to key stakeholders and program personnel to ensure that implications for all affected are carefully considered. The pros and cons (benefits and disbenefits) of all recommendations must be considered before they are finalized to ensure the relevance, credibility, and uptake of evaluation results.

➤ *Ensure Feedback, Follow-Up, and Dissemination of Evaluation Results*

The main audiences for the evaluation report are program staff and management, clients, stakeholders, peers and colleagues, and others who may have assisted with the evaluation. The evaluation report may be too detailed to communicate actions and results to the general community, industry colleagues, students, and parents. Shorter, simplified communications pieces, such as a brief summary in a newsletter or on a web site, should be considered for these audiences.

➤ *Undertake Actions to Ensure Use of Evaluation and Share Lessons Learned*

Evaluation team members must ensure that using the findings and avoiding their misuse are high priorities at this point in the evaluation. Several factors can influence this—evaluator credibility, report clarity, timeliness and impartiality, disclosure of findings, and being committed to undertaking the identified program changes (adapted from *An Evaluation Framework for Community Health Programs*, The Center for Community Based Public Health 2000). As mentioned in Step 1, it is important to think about how the evaluation results will be used during the early planning activities as well as toward the end of the evaluation.

### **Positive Use of Evaluation Results**

Types of activities that ensure positive use of evaluation results include:

- Ensuring the evaluation team is fully informed.
- Making program users aware of the evaluation results and actions being taken in response to the findings.
- Making stakeholders aware of the evaluation results and seeking feedback.
- Using the findings as input into program change decisions.
- Demonstrating how results will be used to improve the program.
- Uncovering new opportunities to bring the benefits of evaluation to users, stakeholders, and interested community organizations.

Sharing what has been learned from the evaluation validates the time and resources spent, and reinforces the need to incorporate evaluation activities into ongoing program planning and improvement cycles. This process also provides support to and acknowledges the evaluation team's efforts.

➤ *Determine Changes to Implement in the Next Evaluation Cycle and Prepare an Action Plan*

The final activity of your evaluation is to determine the changes you will make in the next round of evaluation. What was learned in this evaluation, and which questions are still unanswered can guide your planning of a new set of evaluation objectives, which then become the basis for ongoing program evaluation and improvement. Making evaluation and program improvement a routine part of program management is an important activity for program managers, staff, and possibly users and other stakeholders to be involved in. An action plan that specifies what is going to be done, by whom, and in what timeframe, becomes a standard for ongoing program evaluation and improvement.

## STEP **5C** APPLY EVALUATION STANDARDS

### Overview

During Step 5, be sure to review the evaluation standards, and keep in mind the importance of ensuring that they are met.

5A. INTERPRET AND DOCUMENT FINDINGS	5B. MAKE RECOMMENDATIONS AND TAKE ACTION	<b>5C</b> APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ <i>Interpret findings</i></li> <li>➤ <i>Prepare conclusions and make judgments</i></li> <li>➤ <i>Document evaluation process and findings in evaluation report</i></li> <li>➤ <i>Undertake peer review</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Prepare recommendations</i></li> <li>➤ <i>Ensure feedback, follow-up, and dissemination of evaluation results</i></li> <li>➤ <i>Undertake actions to ensure use of evaluation and share lessons learned</i></li> <li>➤ <i>Determine changes to implement in next evaluation cycle and prepare action plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

## Activities

As the final activities are completed, use the following checklist to ensure the standards continue to be an integral part of the evaluation process.



### Evaluation Standards Checklist

Step 5 Standards	Step 5 Questions
<b>UTILITY</b> Values identification	<input type="checkbox"/> Have the perspectives, procedures, and rationale used to interpret the evaluation findings been carefully described, so that the basis for value judgments is clear?
Report clarity	<input type="checkbox"/> Does the evaluation report clearly describe the program so that the essential information is provided and easily understood?
Report timeliness and dissemination	<input type="checkbox"/> Have interim and final findings and reports been prepared and disseminated so that they can be used in a timely manner?
Evaluation impact	<input type="checkbox"/> Has the evaluation been reported in ways that encourage follow-through by stakeholders so that the evaluation is likely to be used?
<b>PROPRIETY</b> Disclosure of findings	<input type="checkbox"/> Has the evaluation team ensured that the full set of evaluation findings along with limitations are made accessible to those affected by the program and others who have a right to receive the findings?
<b>ACCURACY</b> Justified conclusions	<input type="checkbox"/> Have the conclusions of the evaluation been explicitly justified so that stakeholders can assess them?
Impartial reporting	<input type="checkbox"/> Does the evaluation report impartially and fairly reflect the findings?
Metaevaluation	<input type="checkbox"/> Has the evaluation been formatively and summatively evaluated against relevant standards to ensure assessment of the quality of the evaluation and determination of its strengths and weaknesses?

The next two chapters present two hypothetical driver education program evaluation scenarios based on the five evaluation steps.





## 4. Basic Driver Education Program Evaluation Scenario

Two hypothetical program evaluation scenarios are presented in this chapter and the next. These will help evaluators and practitioners apply the evaluation guidelines. Key activities that an evaluation team can undertake as part of their program evaluation are described. These scenarios take an evaluation team through three key aspects of an evaluation: 1) developing a logic model; 2) making decisions about what to evaluate; and 3) designing the evaluation.

The first scenario represents a hypothetical basic program evaluation that could be undertaken by a medium-size driving school with several locations or a medium-size high school driver education program, each with approximately 700-900 students per year and about 10-15 instructors. It assumes that no evaluation process has previously been in place. Management, however, has identified the need for initiating an ongoing evaluation activity. As a result, the program manager has begun to think about program planning processes that need to be in place, and a staff team has been created to oversee the program's evaluation.

This scenario also assumes that no technical in-house evaluation expertise is available to the team; however, the manager and senior staff supervisor are knowledgeable about the purposes of program evaluation. They agree that there are important reasons to introduce an ongoing evaluation process into the program's operational plans. With this decision has come the commitment of a modest yearly budget (~\$20,000) within which to create and manage an annual evaluation plan.

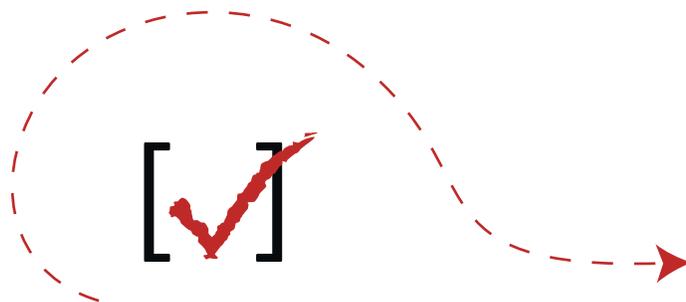
*. . . the program manager has begun to think about program planning processes that need to be in place, and a staff team has been created to oversee the program's evaluation.*

## Program Logic Model

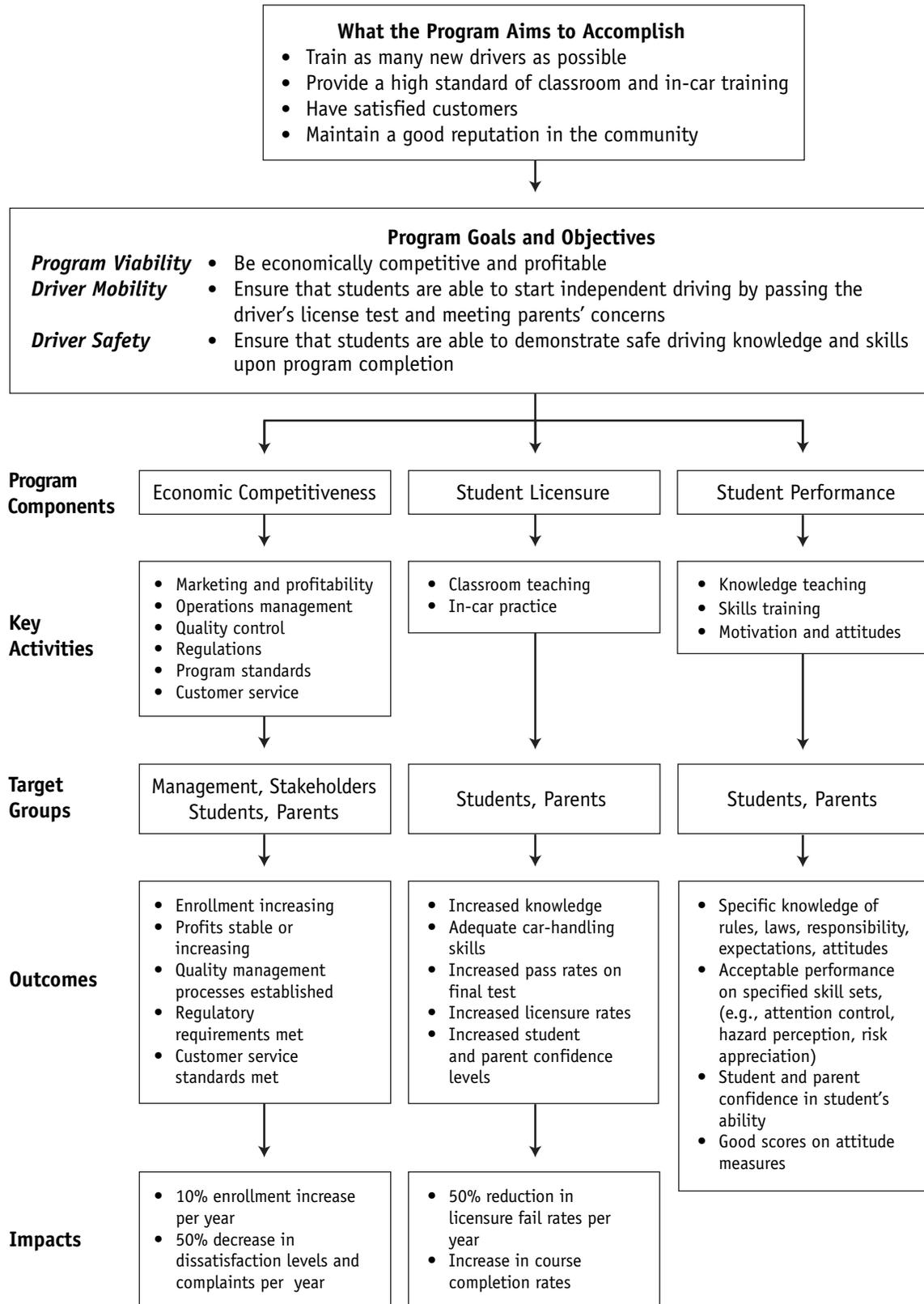
The evaluation team has started working through Step 1 of the *Guidelines*. The members have thought about the program's stakeholders and user and program needs. They have decided to develop a logic model. This process will help them clarify the goals and objectives their program is meant to achieve and the specific activities that are or should be related to these goals and objectives. To do this, the team begins by discussing expected program effects, program activities and resources, and the program's context. Team members examine all the various aspects of the program and organize this information using the Program Organization Worksheet from the *Guidelines* on page 63 and in Appendix E. They are now ready to create the program's logic model.

Using the generic logic model from Chapter 2 (page 42) and the Driver Education Logic Model from Step 1A (page 67) as guides, the team decides to try to capture the most important information in a one-page chart. This is a group effort that takes a series of meetings to complete. Wisely, the program manager uses this opportunity to build commitment and involvement among the staff. The final logic model that team members agree adequately summarizes their program and will serve to focus their evaluation is shown in Figure 7.

With the program logic model completed, the evaluation team members continue to work through Step 1 of the *Guidelines*. They agree the program is ready to be evaluated and begin planning the evaluation. They review current information available about driver education program evaluation, and identify who will use the evaluation and how. They then start to focus on the purpose of the evaluation and what aspects of the program to evaluate.



**Figure 7. Hypothetical Program Logic Model for a Medium-Size Driver Education Program**



## Evaluation Decisions

The team members use their logic model and the evaluation framework from the *Guidelines* to help organize their thoughts on which program areas and key activities to evaluate. They also begin to discuss the scope of the evaluation and how comprehensive an evaluation they are able to carry out. They document their decisions, as described in Table 23.

**Table 23. Evaluation Decisions for a Medium-Size Driver Education Program**

<b>Goal:</b> Operate a viable driver education program (Program Viability)			
<b>Objective:</b> Be economically competitive, profitable, and manage a quality program			
Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Formative Evaluation</b>			
Program Logic	User needs	YES	Since not much is known about student or parent needs related to the program, the team decides that this is important information to begin collecting. Team members believe it can be collected using methods that will serve other evaluation purposes as well.
	Program logic model	YES	The development of a program logic model has been completed as part of initiating an ongoing evaluation process.
	Evaluability	NO	With a program planning process in place, and a logic model now created for a well-established program, the evaluation team decides that the program is ready to be evaluated. Further evaluability assessment is not needed.
Program Context	Stakeholder expectations	NO	While this is an important aspect of the program's operations, the team decides that initially focusing on the expectations of the program's direct users—students and parents—is most important. This information will be incorporated into evaluation activities related to user needs, customer satisfaction, and student input on course effectiveness.
	Regulatory environment	NO	Government relations and regulatory compliance are not issues of concern, as all requirements are being met.
	Contractual environment	NO	There are no immediate concerns about contracts or obligations with partners, and the program does not have franchisers.

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
Business Processes	Operations management	NO	Operations management is OK—the workforce is stable, there is little turnover of staff, and the condition of the facilities is up to standards.
	Quality control	YES	Little is known about quality control issues related to the program’s organization and delivery. Team members agree these should be a target area for the evaluation.
	Marketing	NO	Marketing initiatives and response are OK, as program growth is at acceptable levels.
	Customer service	YES	Although some information is collected from students via a form they are asked to complete at the end of the course, team members feel that better information from students and input from parents should be obtained to help identify areas for program improvement. This area is identified as a priority for the evaluation.
	Sustainability	NO	Revenues, consumer support, and growth are all at acceptable levels, and the evaluation team feels that as the evaluation proceeds, important and useful information will be obtained to assist in future deliberations on expansion and revenue growth.
Program Standards	Benchmarking and certification	YES	The evaluation team decides to check the curriculum against industry standards, such as those of ADTSEA and NIDB, and to investigate certification by the DSAA. This benchmarking will make an important contribution to program evaluation efforts and subsequent program improvement.
	Transportability of program	NO	No current plans exist to expand into other jurisdictions, and the organization is not a franchise operation, so evaluation in this area is not relevant at this time.

**Goal:** Operate a program that successfully prepares students for independent driving (Driver Mobility)

**Objective:** Ensure that students are able to start independent driving by passing the driver's license test and meeting parents' concerns

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Formative Evaluation</b>			
Instructional Products and Processes	Classroom curriculum and materials  Tests and measurements	YES	Some evaluation activities can be used to obtain information to assist both driver mobility and driver safety goals. The team decides that feedback on the course content and materials is important and needed. This is an area to be included in the evaluation. Students are given a written knowledge test at the end of the classroom course to ensure knowledge mastery, but nothing else is done with the results. It is decided that this information will be used to assess the knowledge test and also be given to the students at the beginning of the course to collect information on what they know at the outset. More detailed assessment of the knowledge test will take place in a later evaluation cycle.
	Instructor preparation  Curriculum delivery and in-car practice  Instructional facilities	YES	Overlap between the two goals is also apparent for the in-car training component of the program. Not knowing how many students pass their license test is a serious gap in program information. It is decided that efforts to begin obtaining this information from students is a priority for the evaluation. In addition, finding out how much practice students get with their parents, and finding out about students' and parents' confidence levels are to be included in the evaluation.

**Goal:** Operate a program that successfully prepares students to be safe drivers (Driver Safety)

**Objective:** Ensure that students are able to demonstrate safe driving knowledge and skills upon program completion

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Summative Evaluation</b>			
Student Outcomes	Knowledge outcomes	YES	See comments under Driver Mobility—classroom curriculum and materials. Using the same evaluation activities to provide information relevant to both driver mobility and safety goals will help keep the evaluation within budget as well as make progress in gathering information that can be used to improve both aspects of the program.
	Skills outcomes	YES	See comments under Driver Mobility—curriculum delivery and in-car practice. There is presently no final in-car test conducted with students, and a longer-term objective is to begin working on a test that will be safety-oriented and meet higher performance standards than the existing license test.
	Motivation teaching and outcomes	YES	The team decides that finding out about students’ motivations and attitudes is important for the program and should be included in the evaluation. The team does not think this will require a separate evaluation tool and thus is feasible to consider within the present plan.
	Mobility outcomes	NO	Current resources, skills, and number of students are not adequate to undertake the more in-depth and sophisticated evaluation required in this area.
	Behavioral outcomes	NO	The team members understand that gathering information to assess on-road performance after licensure, driving choices, amount of practice, and habit formation is a very important part of driver education evaluation. These are all important aspects of students’ driving behaviors after leaving the course. They agree, however, that undertaking this type of evaluation before the foundations of the evaluation process are in place is premature.

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
Social Impacts	Crash and injury reduction impacts	NO	Similarly, even more comprehensive studies to investigate students' crash and injury rates after becoming licensed are premature at this time.
Evaluation Quality	Evaluation effectiveness and usefulness	YES	The team has reviewed the program evaluation standards for each step in the <i>Guidelines</i> and has agreed that it is important to ensure that their evaluation is checked against them. They want their evaluation to be as effective as possible, and as their experience with program evaluation increases, they plan to improve each evaluation cycle.

This provides the team with an overall guide for proceeding to identify evaluation questions and targets.

### Evaluation Targets

Following careful consideration of evaluation questions identified from the decision-making process, the team decides that the program's evaluation targets are going to be:

- User needs
- Quality control processes of curriculum materials and delivery methods
- Customer service and satisfaction
- Program benchmarking against established industry standards
- Effectiveness of key curriculum materials
- Student and parent confidence levels in student knowledge and ability
- Safe driving knowledge and ability of students
- Student attitudes
- Licensure rates of students
- Effectiveness of the evaluation processes

Team members are concerned that they may be taking on too much, but also want to ensure that this first cycle of evaluation is as comprehensive and effective as possible. They decide to discuss this with the external evaluator whom they have decided to hire, and agree to consider scaling back the evaluation, if necessary.

## Evaluation Design

With the foundation for the evaluation in place, the team now begins to think about the research approach that their evaluation plan requires. Team members decide to hire an experienced evaluator on an as-needed basis to provide guidance as they move into the design and implementation phases of the evaluation. The team feels comfortable carrying out some of the evaluation tasks, but because no one has been involved in a program evaluation before, team members recognize that outside evaluation expertise is needed.

Because the initial annual budget is modest, the program manager suggests that the team seek out a graduate student from the community who is experienced in evaluation. Contacts with stakeholders and community organizations lead to identifying a graduate student who is knowledgeable and has some experience with both formative and summative types of evaluation. This person is interviewed and subsequently hired on a part-time basis.

The first task for the evaluator is to provide an assessment of the evaluation target list. Although the list seems lengthy, the evaluator agrees with the team's assessment that several of the targets can be evaluated using the same tools. Information on student needs, satisfaction levels, and confidence levels, for example, can be gathered through student focus groups and questions on a student survey at the end of the course. Similarly, a survey developed for parents can ask about their needs as users of the program, their views on customer service, and their satisfaction with the program.

With input from the evaluator, the team decides to undertake a staged evaluation in which the initial activities will take place over two years. During the first year, team members will conduct a formative evaluation of the program's instructional products and processes, and gather information on user needs, customer satisfaction, and quality control. They will also undertake the data collection required to initiate some summative evaluation activities related to student outcomes. Then in the second year, improvements based on the formative evaluation will be initiated and evaluated, and attention will be given to developing a systematic summative evaluation to track student outcomes. From there, an ongoing evaluation process will be established, monitored, and possibly expanded.

The formative evaluation will consist of both qualitative and quantitative methods, and the summative evaluation will use quantitative methods. Again, with input from the evaluator, the team develops an evaluation design to identify the specific types of evaluation and data collection methods to implement. The resulting evaluation design includes the methods described in Table 24.

**Table 24. Sample Evaluation Design for a Medium-Size Driver Education Program**

Evaluation Type	Evaluation Targets	Qualitative Methods	Quantitative Methods
Formative Evaluation	User needs	<ul style="list-style-type: none"> <li>• Student focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Parent needs and satisfaction survey</li> </ul>
	Quality control Customer service	<ul style="list-style-type: none"> <li>• Quality control interviews with instructors to identify improvements and new processes</li> </ul>	<ul style="list-style-type: none"> <li>• Parent needs and satisfaction survey</li> <li>• Student exit survey</li> </ul>
	Curriculum and materials	<ul style="list-style-type: none"> <li>• Benchmarking of program*</li> <li>• Student focus groups</li> <li>• Instructor log books</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge test</li> <li>• Student exit survey</li> </ul>
	Curriculum delivery and in-car practice	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Instructor log books</li> </ul>	<ul style="list-style-type: none"> <li>• Student exit survey</li> </ul>
Summative Evaluation	Tests and measurements	<ul style="list-style-type: none"> <li>• Parent feedback sessions</li> <li>• Student focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge test</li> <li>• Student exit survey</li> </ul>
	Motivation and attitudes		
Metaevaluation	Evaluation effectiveness	<ul style="list-style-type: none"> <li>• Benchmark evaluation against Program Evaluation Standards</li> </ul>	

\*Procedures for benchmarking the program against industry standards will be created by program staff based on the specific standards used.

The evaluation types and methods that are going to be implemented in this hypothetical evaluation will not require sophisticated sampling methods. The sample of each population, be it students, parents, or instructors will be obtained by requesting each group's participation. There is little the evaluation team can do to insist on participation of parents, for example, other than provide several opportunities and reminders to complete a survey or attend a feedback session. Incentives could be offered to increase participation.

Students, on the other hand, can be considered a “captive” target group because they will be asked to complete the exit survey during the final class of the course (which is mandatory). Students who participate in the focus groups, however, are volunteers, but the numbers required are small (~5-7 per group). It is recognized that those who choose to participate in the evaluation processes are probably different from those who decline. If refusal rates become too high (more than 40%), steps should be taken to see what the key differences might be.

Involvement of staff and clients can also enrich qualitative data collection. Instructors are expected to participate in any evaluation activities requested of them, but it is important to recognize that they may initially feel threatened. People may mistake evaluation for judging. A participatory approach can reduce potential antagonism and anxiety toward the evaluation process, thereby creating buy-in early on for its findings. By framing evaluation as a continuous process of program improvement, staff involvement can lead to developing new skills and a broader understanding of their work.

**B**y framing evaluation as a continuous process of program improvement, staff involvement can lead to developing new skills and a broader understanding of their work.

The team also requests assistance from the evaluator to establish the evaluation’s ethical procedures. A document is prepared to identify the actions that will be included in the evaluation to meet ethics criteria. These actions include documenting the purpose and methodology of the evaluation, informed consent measures, anonymity and confidentiality procedures, and methods of disseminating the evaluation findings.

A data collection plan is then developed, and the data sources identified. In this evaluation, the data sources are the program’s instructors, students, and their parents. In addition, the benchmarking of program standards will include assessing curriculum materials and processes. Indicators of program success for each evaluation target are identified, and existing and relevant data collection tools are modified where possible, and where not, developed.

Based on the evaluation design, the instruments or tools the team needs to identify or develop include:

#### **Qualitative Tools:**

1. An interview guide for conducting interviews with instructors to discuss existing, new, and improved quality control processes in the classroom.
2. A focus group guide for conducting facilitated discussions with students about the curriculum materials and instructional processes used in the classroom and the in-car sessions, and motivations and attitudes about safe and responsible driving.
3. Log books for instructors to gather information about instructional materials and processes during their classroom sessions and in-car lessons.
4. A guide for parent feedback sessions at the end of each course.

#### **Quantitative Tools:**

1. Customer satisfaction survey for parents.
2. Baseline knowledge test given to students at the beginning and end of each course.
3. Student exit survey to assess the effectiveness of curriculum materials and instructional processes, determine satisfaction levels, and investigate motivation and attitudes about safe and responsible driving.

The team recognizes that including these seven tools in the evaluation is time consuming, but with the assistance of the evaluator, the team feels it is feasible to proceed. Once developed, these tools will be used on an ongoing basis, and thus, the initial investment of time and resources is believed to be cost-effective in the longer term. Samples of each of these tools and examples of relevant existing instruments are found in Appendices F and H.

Once the development of the data collection tools is complete, the evaluator carries out a small pilot test of each one to identify problems areas. The tools are revised and the

final tools produced. The evaluation proceeds with training the data collectors, after which the data collection commences.

Another task the evaluator has taken on is identifying data analyses that should be undertaken for the different sets of data. Upon completion of the data collection, these analyses are carried out by the evaluator with assistance from team members, and the evaluator helps the team understand and interpret the findings. The evaluation team then prepares evaluation conclusions, documents them in an evaluation report, and presents this to the program management and staff. Recommendations for responses to the evaluation and actions to be taken related to program improvements and the next evaluation cycle are drafted. The team and management agree that an action plan will be prepared to identify required changes and a time frame for implementation.

## **Evaluation Quality**

Throughout the evaluation process, the team has followed the evaluation standards from the *Guidelines* to ensure their evaluation is effective. As team members work through each step, they check their evaluation against the evaluation standards. In this way, the team develops a sound process upon which to base an ongoing evaluation of their driver education program.





## 5. Advanced Driver Education Program Evaluation Scenario

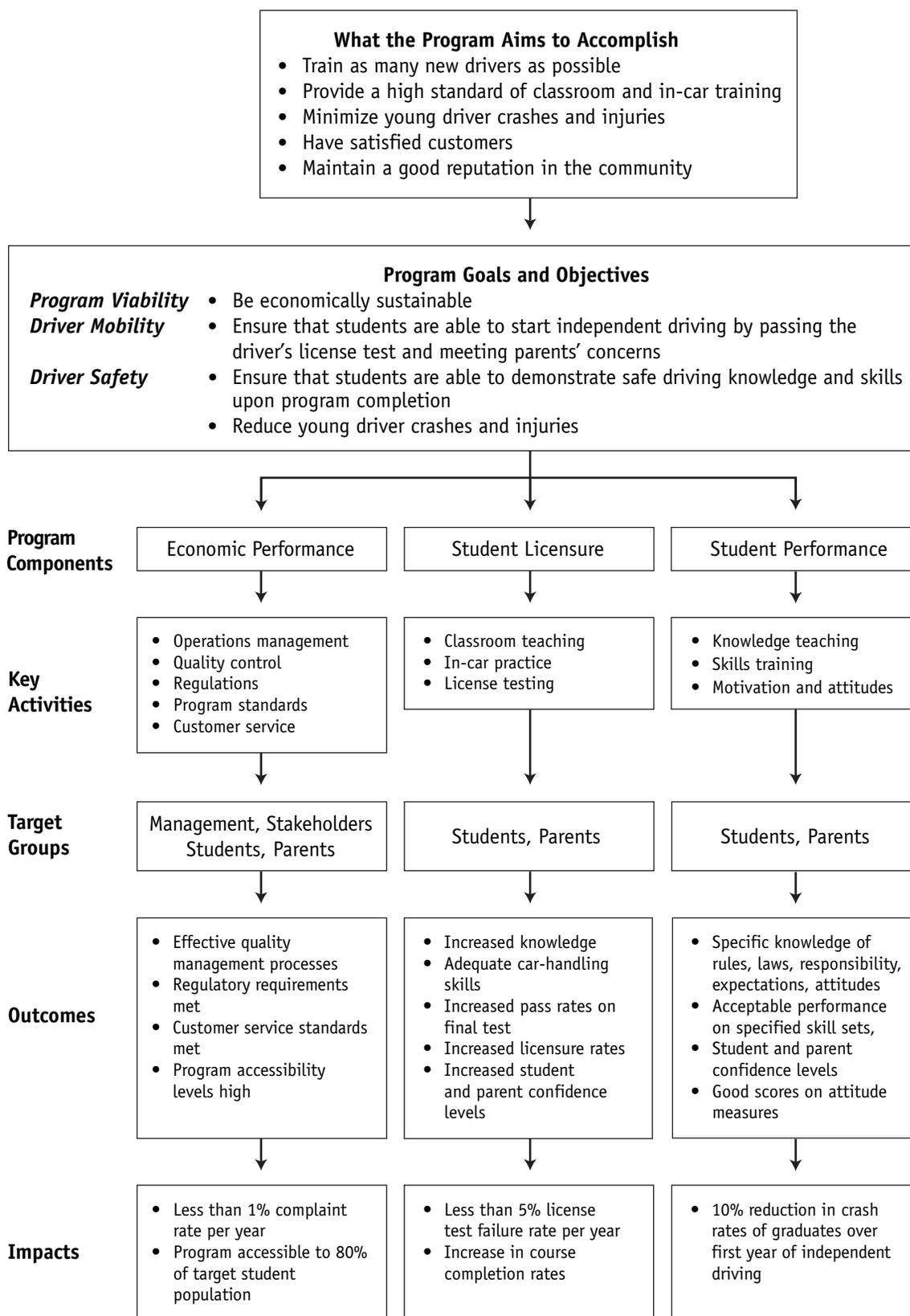
This chapter presents a more advanced evaluation scenario, which builds on the basic evaluation in Chapter 4. Again, three key aspects of an evaluation are described—developing a logic model; making decisions about what to evaluate; and designing the evaluation. This scenario represents a hypothetical evaluation that could be undertaken by a large driver education program being operated by a state or province. It is assumed that the program services at least 10,000 students per year with a few hundred instructors, and it has a head office and staff for administration and research and development. The jurisdiction’s Driver Education Office manages the program, which is delivered through high schools and commercial driving schools.

This scenario also assumes the driver education program has conducted evaluation and program development activities in the past. The program’s strategic planning, operational goals, objectives, and program standards are also well established. The program content and operations have previously been assessed and improved through formative evaluations, which allows the present evaluation to focus on program outcomes and impacts.

*People responsible for the program recognize that the comprehensive evaluation requires a wide range of technical expertise . . .*

This scenario further assumes that technical in-house evaluation expertise is available. An evaluation team has been created that includes an evaluator; head office program staff and field staff members; as well as representatives from the driver licensing department, the education department, and the central road safety coordinating office. People responsible for the program recognize that the comprehensive evaluation requires a wide range of technical expertise in education evaluation, transportation safety program evaluation, and in research design and statistical analysis of large surveys and driver record databases. Additional expertise will be sought from outside, and management has approved the initiation of a more comprehensive, ongoing evaluation process for the program. An annual budget of \$150,000 has been allocated to cover internal costs and outside help.

**Figure 8. Hypothetical Program Logic Model for a Large Driver Education Program**



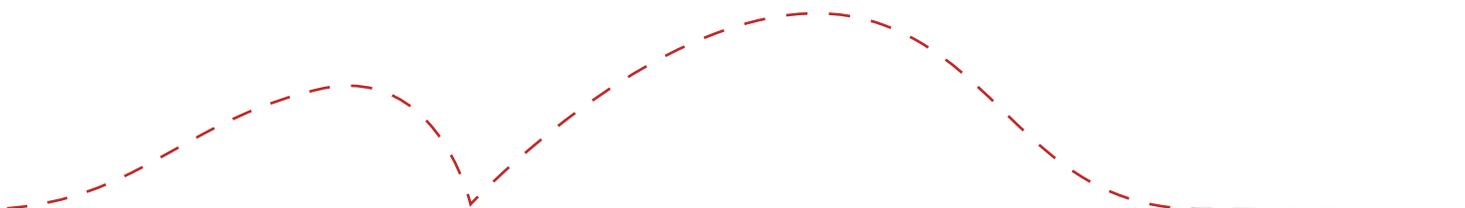
## Program Logic Model

The evaluation team has studied Step 1 of the *Guidelines* and discussed user and program needs as well as stakeholder expectations. Because a program logic model has not previously been developed, team members have decided to develop a model to document the linkages between program goals, objectives, activities, outcomes, and impacts. This will ensure that the understanding of these relationships is logical, appropriate, and complete. Team members begin by examining expected program effects, program activities and resources, and program context. They look at all relevant program information and determine that no major gaps exist. A team member takes on the responsibility of coordinating efforts to develop a logic model. A series of meetings take place to prepare a draft and present it to the program manager and senior management. The approved logic model shown in Figure 8 focuses the evaluation and identifies priority areas.

With the program logic model completed, the evaluation team works through the remaining tasks of Step 1. The detailed planning of the evaluation targets is the next priority. The team agrees that the planning is to be thorough and with careful attention to detail. This will establish the foundations for an effective and ongoing evaluation process. Current information available about driver education program evaluation is reviewed. The team also documents how the evaluation findings are expected to be used. Team members then start to focus on the purpose of the evaluation and the specific aspects of the program to evaluate in the present cycle.

## Evaluation Decisions

Team members use the logic model and the evaluation framework from the *Guidelines* to identify the program areas and key activities to include in their evaluation. They discuss evaluation scope, focusing on the requirements of a comprehensive evaluation. They document their decisions, as shown in Table 25. This provides an overall guide to identifying evaluation questions and targets.



**Table 25. Evaluation Decisions for a Large Driver Education Program**

<b>Goal:</b> Operate a viable driver education program (Program Viability)
<b>Objective:</b> Be economically sustainable and manage a quality program

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Formative Evaluation</b>			
Program Logic	User needs	NO	The program monitors the young driver research literature and collects information from students and parents about their needs. These data are used in the program’s annual content review; therefore, this area does not need to be included in the evaluation.
	Program logic model	YES	A program logic model has been completed as part of initiating this ongoing evaluation process.
	Evaluability	NO	With strategic and program planning processes in place, a logic model created, a well-established program, and some evaluation activities already implemented, it is clear that the program is ready to be evaluated for outcomes and impacts.
Program Context	Stakeholder expectations	YES	Stakeholders have never been actively involved in program evaluation, and the team believes their input is important. It is timely to initiate a process of information gathering with stakeholders as part of this new evaluation process. This is crucial to identifying evaluation questions, anticipating criticisms, and ensuring the utility of the evaluation results. The team recommends that stakeholder input be included in the evaluation.
	Regulatory environment	NO	All requirements of regulatory compliance are being met, and there is no need to include this area in the evaluation.
	Contractual environment	NO	There are no contractual requirements that need attention.
Business Processes	Operations management	NO	Operations management is OK; ISO 9000 certification is in place.
	Quality control	NO	The program has quality management and control processes in place, and they are operating satisfactorily. There is no need to include them in the evaluation.

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
Business Processes <i>(continued)</i>	Marketing	NO	The program's marketing activities are approved, and ongoing assessment is already in place.
	Customer service	NO	Customer service monitoring is in place and working well, e.g., student and parent course evaluations, satisfaction levels, and identification of areas for operational improvements.
	Sustainability	NO	Revenues, consumer support, and growth are all at acceptable levels.
Program Standards	Benchmarking and certification	NO	Program benchmarking is undertaken annually against the jurisdiction's regulated standards and industry benchmark standards.
	Transportability of program	NO	Evaluation in this area is not currently relevant to this driver education program.

<b>Goal:</b> Operate a program that successfully prepares students for independent driving (Driver Mobility)
<b>Objective:</b> Ensure that students are able to start independent driving by passing the driver's license test and meeting parents' concerns

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Formative Evaluation</b>			
Instructional Products and Processes	Classroom curriculum and materials	NO	The effectiveness of the classroom materials has been evaluated in prior evaluation work, and the team is comfortable with the program. Team members recognize the utility of continual formative research and agree to incorporate it into the ongoing evaluation process that is to be part of the follow-up of the present process.
	Curriculum delivery and in-car practice	NO	The team feels similarly about the in-car component of their program. It is routinely assessed and its evaluation will be incorporated into the ongoing evaluation process.

**Goal:** Operate a program that successfully prepares students to be safe drivers (Driver Safety)

**Objective:** Ensure that students are able to demonstrate safe driving knowledge and skills upon program completion

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
<b>Summative Evaluation</b>			
Student Outcomes	Knowledge outcomes	YES	The effectiveness of a revised curriculum in ensuring more complete knowledge retention is key to evaluation concerns, so it is included as a priority. Textbook-based tests, and data from license knowledge tests are available for earlier graduates, and these are identified as suitable knowledge criterion measures for comparing current knowledge outcomes with those of the earlier program.
	Skills outcomes	YES	The effectiveness of the new curriculum in achieving a higher and more consistent level of skill in graduates is also a key concern. The team decides to find or develop a test that will be safety-oriented and meet higher performance standards than the existing license test.
	Motivation outcomes	YES	Students' motivations and attitudes toward safe and responsible driving are an important component of the curriculum, and little is known about how effective the program is in influencing these factors. Team members decide this area should be included in the evaluation, and suitable measurement instruments found or developed.
	Mobility outcomes	YES	The team agrees that this important outcome area of the program needs to be evaluated. Team members want to be able to track licensure rates, license retention or loss rates, and identify driving exposure measures that could also be tracked.
	Behavioral outcomes	YES	The team feels that information about on-road performance after licensure, driving choices, amount of driving, and habit formation is a very important part of driver education evaluation. This is a priority for the evaluation.

Program Area	Program Activities and Evaluation Targets	Inclusion in Evaluation	Reasons
Social Impacts	Crash and injury reduction impacts	YES	A comprehensive investigation of students' crash and injury rates after becoming licensed is also recommended as a high priority. This will be a significant advance in the program's evaluation process and in gaining a better understanding of program impacts.
Evaluation Quality	Evaluation effectiveness and usefulness	YES	The evaluation team has reviewed the Program Evaluation Standards and recommends the evaluation be checked against them. The evaluation needs to be as effective as possible given its importance and the resources being allocated to it.

Lists of evaluation questions and specific evaluation targets are generated from this decision-making matrix, and the following evaluation targets are recommended to and approved by senior management:

- Stakeholder expectations
- Safe driving knowledge and ability of students
- Student attitudes toward safe and responsible driving
- Mobility of students after licensing
- Post-license performance measures including on-road performance, driving choices, amount of driving, and habit formation
- Crash and injury rates during the first year of licensed driving
- Effectiveness of the evaluation processes

## Evaluation Design

With the evaluation targets approved, the team now develops the evaluation research approach and then the more detailed evaluation design. The approach includes a relatively comprehensive summative evaluation of outcomes and impacts. Approval has been received to hire outside people to assist with the more complex aspects of a comprehensive evaluation. The team requires research design and statistics expertise, and contracts with an experienced evaluator and a graduate statistics student from a transportation research center located in a nearby university.

The evaluator and statistician draft an overall approach to undertake a staged evaluation that will take place over 2½ to 3 years to allow time for accumulation of driver records and follow-up surveys of drivers during their first two years of independent driving. Student outcome and safety impact measures will be compared:

1. On a before/after basis between cohorts of program graduates; and
2. On a partially controlled quasi-experiment basis with new drivers who have no known formal driver education.

The evaluation will include both qualitative and quantitative methods. The evaluator helps the team identify specifically which types of evaluation and data collection methods to implement. At this point, the statistician also participates in the design development to ensure that the specific data, their location, and accessibility are considered. The evaluation methods are shown in Table 26.

**Table 26. Sample Evaluation Design for a Large Driver Education Program**

Evaluation Type	Evaluation Targets	Qualitative Methods	Quantitative Methods
Formative Evaluation	Stakeholder expectations	<ul style="list-style-type: none"> <li>• Interviews</li> <li>• Stakeholder workshop</li> </ul>	
Summative Evaluation	Knowledge and skill outcomes	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Comparison group focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge test</li> <li>• Skill test</li> <li>• Advanced skill test</li> <li>• Student exit survey</li> </ul>
	Motivation and Attitudes	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Comparison group focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Student survey</li> <li>• Comparison group survey</li> </ul>
	Mobility outcomes	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Comparison group focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Student survey</li> <li>• Comparison group survey</li> </ul>
	Behavioral outcomes	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Comparison group focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Student survey</li> <li>• Comparison group survey</li> </ul>
	Crash and injury reductions	<ul style="list-style-type: none"> <li>• Student focus groups</li> <li>• Comparison group focus groups</li> <li>• General public focus groups</li> </ul>	<ul style="list-style-type: none"> <li>• Student survey</li> <li>• Comparison group survey</li> <li>• Parent survey</li> <li>• General public survey</li> <li>• Longitudinal record study</li> </ul>
Meta-evaluation	Evaluation effectiveness	<ul style="list-style-type: none"> <li>• Benchmark evaluation against Program Evaluation Standards</li> </ul>	

Preparatory tasks will include the following activities:

### **Plan and organize stakeholder interviews and workshop**

Gathering input from stakeholders about their views and program expectations will involve individual interviews with several key stakeholders, including representatives from the automotive and insurance industries, licensing officials, enforcement agencies, and others. Each person will be contacted by phone and requested to participate by providing an hour of his or her time for the interview. A guided interview format will be used to ensure collection of consistent information, and the interviews will be recorded with agreement from the participants. Email contact will be used as a backup when stakeholder representatives are not easily contacted by phone or prefer to communicate via email.

A stakeholder workshop will also be organized with representatives from key organizations. This will provide a forum for additional input about stakeholder expectations and also an opportunity for consensus building on important issues for the evaluation process and beyond.

### **Plan and organize the focus groups**

In-depth information will be gathered through focus groups with students and comparison groups of youth who did not take the target driver education course. Participants will be asked about:

- Knowledge and skills they learned in the course
- Motivation and attitudes about safe and responsible driving
- Post-licensing driving including trip patterns, reasons, times, amount of driving, vehicle occupants, skill levels, and risk taking
- Crash frequency, results, and lessons learned

Students will be asked to participate in a one-hour group discussion at the end of the course. Comparison group participants will be identified through driver and licensing records and contacted by phone. Large numbers of participants are not required, but the evaluator recommends at least 4 groups each be held. A discussion guide will be developed that session moderators will follow, and the sessions will be recorded. Since transcribing these discussions can be time consuming, the evaluators will train a staff person to undertake this task.

## **Select or develop tests and surveys**

A review of existing knowledge and in-car tests will be undertaken. The evaluators will help team members decide whether they need to develop new tests or are comfortable with the suitability of the tests that have been identified. Existing surveys will be reviewed and items used as appropriate. Again, decisions will be made with guidance from the evaluator about how much modification is required to create the appropriate tools for this evaluation.

## **Pilot test the instruments**

Once the data collection instruments are finalized, the evaluators will help pilot test them with small volunteer samples of the target populations. This ensures the instruments are readable, clear, straightforward, and understandable. Based on this feedback, revisions will be made, and preparations for distribution undertaken.

## **Plan and organize the surveys**

The evaluators recommend mail-out surveys to all target groups due to the time and cost associated with telephone interviews. The logistical details of creating, reproducing, and labeling questionnaires, and their preparation for coding must be organized and scheduled. Procedures for selecting respondents as well as creating mailing lists, labels, cover letters, and reminder cards need to be scheduled.

## **Build databases to house, test, record, and survey results**

The procedures for housing all evaluation data will be established. Comparable databases will be created to locate each unique set of data and allow for data matching, where desirable. Confidentiality requires secure storage.

## **Identify driver record data system**

Several aspects of obtaining and handling the driver record data will be important to organize early in the evaluation. Access to the data must be organized and the extraction procedures designed. Depending on where the data are located, this can be a fairly time-consuming activity. Driver record systems are typically very large, complex files designed primarily for administrative purposes rather than research. Getting permission to use, and actually obtaining, government driver record data can be difficult, time consuming, and may involve a cost. Handling large data sets and extracting reliable evaluation data from them are complex tasks requiring special expertise.

## **Design the survey and record study sampling plans**

The sampling plans for the surveys and record study are critical aspects of the evaluation design, which will also require attention early in the evaluation. The team will rely heavily on the evaluators to identify the appropriate sample sizes and procedures for optimizing survey response rates.

## **Design statistical analyses of the survey and record data**

The evaluators will also design the statistical analyses for the surveys and record data, and ensure the data collection tools are formatted to facilitate data coding and entry.

## **Determine ethical procedures**

The team establishes ethical procedures for the evaluation. A document is prepared that identifies the actions that will be included in the evaluation to meet ethics criteria. These actions include documenting the purpose and methodology of the evaluation, informed consent measures, anonymity and confidentiality procedures, data storage arrangements, and methods of disseminating the evaluation findings.

A data collection plan is then developed, and data sources identified. In this evaluation, the data sources include the program's instructors, students and their parents, comparison (control) group participants, and members of the general public. Indicators of program success for each evaluation target are identified and, with assistance from the evaluators, existing and relevant data collection tools are located, and new tools developed where needed.

**H***andling large data sets and extracting reliable evaluation data from them are complex tasks requiring special expertise.*

Based on the evaluation design, the tools or instruments the team needs to identify or develop include:

### **Qualitative Tools:**

1. An interview guide for conducting telephone or email interviews with officials and representatives from key stakeholder groups to identify key program issues and expectations.
2. A workshop framework for organizing and conducting a workshop with stakeholder groups to seek consensus on key issues, expectations, and appropriate actions with particular focus on program evaluation.
3. A focus group guide for conducting facilitated discussions about driving knowledge and skills, motivations and attitudes toward driving, driving behaviors, and crash experiences, with graduated students who have obtained their driving license.
4. A focus group guide for conducting facilitated discussions with comparison groups of students who did not take the driver education program but are licensed. The same discussion topics will be used.

### **Quantitative Tools:**

1. The existing knowledge test, which has been given to students at the end of the course, is retained and used again in this evaluation, and also given to a comparison group.
2. The routine skills test used at the end of the course is retained for “before” and “after” comparisons.
3. A special, more advanced skills test is found and given to a sample of new graduates and a comparison group of new drivers who learned to drive without formal driver education.
4. A student exit survey is given to all students at the last class of the course. Questions are added to capture limited personal information to help calibrate later follow-up survey returns.

5. A student follow-up survey using a mail-out questionnaire will be administered at intervals after graduation and licensing. It asks about:
  - Personal, geographic, demographic, and socioeconomic information
  - Opinions of the course
  - Knowledge and skills learned in the course
  - Amount of practice driving during course
  - Skill assessment
  - Motivation and attitudes about safe and responsible driving
  - Post-licensing driving, including trip patterns, reasons, times, amount of driving, vehicle occupants, skill levels, and risk taking
  - Crash frequency and results, violations, license actions, and lessons learned
6. A comparison group survey using a mail-out questionnaire asking about:
  - What, if any, driver education instruction they took, and the knowledge and skills they learned
  - Opinion of course taken
  - Amount of practice driving during course
  - Skill assessment
  - Motivation and attitudes
  - Post-licensing driving
  - Crash frequency and results, violations, license actions, and lessons learned
7. A parent survey using a mail-out questionnaire asking about:
  - Assessment of driver education course taken by teenager
  - Effect of driver education on teenager's safe and responsible driving
  - Amount of practice driving during driver education course
8. General public survey using a mail-out questionnaire asking about:
  - Assessment of driver education course provided by jurisdiction
  - Perceived effect of driver education on teenagers' safe and responsible driving
  - Willingness to pay to prevent teen crash casualties
9. Driver records database to capture officially recorded:
  - Convictions
  - Licensing sanctions (suspensions, revocations)
  - Crashes of varying severity

Samples of some of these tools and examples of relevant existing instruments are found in Appendices G and H.

The systematic implementation of the evaluation design is to be carefully planned and monitored. The team recognizes that the plan may have to be changed as obstacles arise. Once the data collection tools have been developed, the team will carry out pilot tests of each tool to identify problem areas. The tools are then revised and finalized. The evaluation proceeds with training the data collectors, and data collection commences.

Another task the evaluators have taken on is identifying data analyses that should be undertaken for the different sets of data. Upon completion of the data collection, these analyses are carried out by the evaluators with assistance from team members. The evaluators help the team understand and interpret the findings.

The evaluation team prepares the conclusions of the evaluation and documents the research and findings in a series of evaluation reports. The team presents these reports to the program management and staff in a series of briefing sessions.

Recommendations for responses to the evaluation and actions to be taken related to program improvements and the next evaluation cycle are drafted. The team and management agree that an action plan will be prepared to identify changes and a timeframe for implementation.

## **Evaluation Quality**

Throughout the evaluation process, the team has followed the evaluation standards from the *Guidelines* to ensure their evaluation is defensible and effective. As team members work through each step, they check the evaluation against the evaluation standards questions. In this way, a sound process is developed upon which to base an ongoing driver education program evaluation.

At the end of the present cycle, an ongoing evaluation process will be established, monitored, and possibly expanded, as further improvements are made to the program and to its coordination with licensing and other safety measures.



## 6. Conclusion

Objective, systematic evaluation is needed to help driver education programs improve and to maximize the likelihood of having a measurable safety impact. Systematic evaluation consists of an ongoing series of stepped evaluation actions used to improve driver education programs and raise the bar of program performance and outcomes. It is based upon a foundation built from evaluation models, program logic, a comprehensive evaluation framework, and program evaluation standards.

The *Guidelines* will help driver education researchers and managers determine which type and scale of evaluation fits their specific circumstances. They can then conduct a program evaluation based on sound and astute decisions about what the evaluation intends to achieve and how it will aid program improvement and impact. These stepped evaluation actions can be used to improve driver education program evaluations to a higher standard than in the past.

The *Guidelines* are expected to promote more consistent and competent application of evaluation principles and methods. Once adopted and implemented on a regular basis, they will establish new standards for driver education evaluation, resulting in improvement in both driver education evaluation and driver education programs. While evaluation is important to improving the effectiveness and efficiency of driver education, recognizing its limitations is also important. This recognition has been lacking in the past and has led to unfortunate policy decisions. Evaluation of driver education, like driver education itself, is evolving and still far from its ultimate conclusion. These *Guidelines* provide the understanding and step-by-step guidance to increase the potential of evaluation to help driver education programs improve their products and processes and become more effective in helping youth become safer drivers.

Remember the companion documents *Evaluating Driver Education Programs: Management Overview*, and *Evaluating Driver Education Programs: How-To Guide* are available for others in your organization who might find the *Guidelines* too detailed. The *Management Overview* provides an introduction to evaluation, and a summary of driver education evaluation. The *How-To Guide* is a manual which explains the steps involved in conducting a formative evaluation.

Use these *Guidelines* to help initiate and carry out effective, ongoing evaluation activities within your driver education program.

## REFERENCES

- Boyce, T. E., and E. S. Geller. 2002. An instrumented vehicle assessment of problem behavior and driving style: Do younger males really take more risks? *Accident Analysis and Prevention* 34 (1):51-64.
- Carstensen, G. 1994. *Evaluation of a new driver education in Denmark*. Linköping, Sweden: Swedish Road Transport Research Institute.
- Carstensen, G. 2002. The effect on accident risk of a change in driver education in Denmark. *Accident Analysis and Prevention* 34 (1):111-21.
- Centers for Disease Control and Prevention. 1999. Framework for program evaluation in public health. *MMWR*. 48 (No.RR-11):1-42.
- Chipman, M. L. 1982. The role of exposure, experience and demerit point levels in the risk of collision. *Accident Analysis and Prevention* 14 (6):475-83.
- Christie, R. 2001. *The effectiveness of driver training as a road safety measure: A review of the literature*. Melbourne, Australia: Royal Automobile Club of Victoria (RACV) Ltd., Public Policy Group.
- Christie, R. 2002. Road safety education and training from a public health perspective. *Road Safety Research, Policing and Education Conference Proceedings*.
- Davis, C. S. 1990. *The DeKalb County, Georgia, Driver Education Demonstration Project: Analysis of its long-term effect*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Donelson, A. C., and D. R. Mayhew. 1987. Driver improvement as post-licensing control: The state of knowledge. In *Nonpunitive approaches to driver improvement: Rewards, incentives, and reinforcers*. Toronto, ON: Ontario Ministry of Transportation and Communications.
- Dreyer, D., and M. Janke. 1979. The effects of range versus nonrange driver training on the accident and conviction frequencies of young drivers. *Accident Analysis and Prevention* 11:179-98.
- Elvik, R., and T. Vaa, eds. 2004. *The handbook of road safety measures*. Amsterdam: Elsevier.
- Engström, I., N. P. Gregersen, K. Hernetkoski, E. Keskinen, and A. Nyberg. 2003. *Young novice drivers, driver education and training: Literature review*. VTI Rapport 491A. Linköping, Sweden: Swedish National Road and Transport Research Institute.
- Evans, L. 1987. Young driver involvement in severe car crashes. *Alcohol, Drugs and Driving* 3 (3-4):63-78.
- Evans, L. 1991. *Traffic safety and the driver*. New York: Van Nostrand Reinhold.

Forsyth, E., G. Maycock, and B. Sexton. 1995. *Cohort study of learner and novice drivers. Part 3: Accidents, offences and driving experience in the first three years of driving*. Crowthorne, Berkshire: Transport Research Laboratory.

Fuller, R., and D. Bonney. 2003. Driver education and training in post-primary schools in Ireland. In *Behavioural research in road safety: Thirteenth seminar*. Crowthorne, Berkshire: Transport Research Laboratory.

Fuller, R., and D. Bonney. 2004. *Driver education and training in post-primary schools. Monograph*. Dublin: National Council for Curriculum and Assessment.

Glad, A. 1988. *Phase 2 Driver Education: Effect on the risk of accident*. Oslo, Norway: Norwegian Center for Research, Department of Transportation.

Goldman, K. D., and K. J. Schmalz. 2001. Theoretically speaking: Overview and summary of key health education theories. *Health Promotion Practice* 2 (4):277-81.

Gonzales, M. M., L. M. Dickinson, C. DiGuseppi, and S. R. Lowenstein. 2005. Student drivers: A study of fatal motor vehicle crashes involving 16-year-old drivers. *Annals of Emergency Medicine* 45 (2):140-46.

Green, L. W., and M. W. Kreuter. 1999. *Health promotion planning: An educational and environmental approach*. 3rd ed. Mountain View, CA: Mayfield Publishing (McGraw-Hill).

Gregersen, N. P. 1994. Systematic cooperation between driving schools and parents in driver education, An experiment. *Accident Analysis and Prevention* 26 (4):453-61.

Gregersen, N. P. 1996. Young drivers' overestimation of their own skill—An experiment on the relation between training strategy and skill. *Accident Analysis and Prevention* 28 (2):243-50.

Haworth, N., N. Kowadlo, and C. Tingvall. 2000. *Evaluation of pre-driver education program*. Victoria, Australia: Monash University Accident Research Centre.

Highway Safety Center. 2002. *Traffic safety education life long learning process*. Indiana, PA: Indiana University of Pennsylvania.

Joint Committee on Standards for Educational Evaluation. 1994. *The Program Evaluation Standards: How to assess evaluations of educational programs*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Katila, A., E. Keskinen, S. Laapotti, and M. Hatakka. 1995. *Changes in slippery road accidents as an effect of renewed driver training in Finland*. Turku, Finland: University of Turku.

Keskinen, E., M. Hatakka, and A. Katila. 1998. *The Finnish way of driver improvement: Second phase of training*. Paper presented at International Driver Improvement Workshop, at Bergisch Gladbach, Germany: Bundesanstalt fur Strassenwesen.

Levy, D. T. 1988. The effects of driving age, driver education, and curfew laws on traffic fatalities of 15-17 year olds. *Risk Analysis* 8 (4):569-74.

- Levy, D. T. 1990. Youth and traffic safety: The effects of driving age, experience, and education. *Accident Analysis and Prevention* 22 (4):327-34.
- Lonero, L., and K. Clinton. 1998. *Changing road user behavior: What works and what doesn't*. PDE Publications. Available from www.drivers.com.
- Lonero, L., and K. Clinton. 2006. *Situational analysis of driver education in Canadian provinces and territories*. Ottawa, ON: Natural Resources Canada.
- Lonero, L., K. Clinton, J. Brock, G. Wilde, I. Laurie, and D. Black. 1995. *Novice driver education model curriculum outline*. Washington, DC: AAA Foundation for Traffic Safety.
- Lonero, L. P., K. M. Clinton, R. C. Peck, D. Mayhew, A. Smiley, and D. Black. Unpublished. *Driver improvement programs: State of knowledge and trends*. Ottawa, ON: Transport Canada.
- Lonero, L. P., K. M. Clinton, B. N. Persaud, M. L. Chipman, and A. M. Smiley. 2005. *A longitudinal analysis of Manitoba Public Insurance Driver Education Program*. Winnipeg, MB: Manitoba Public Insurance.
- Lonero, L. P., K. M. Clinton, G. J. S. Wilde, K. Roach, A. J. McKnight, H. Maclean, S. J. Guastello, and R. Lamble. 1994. *The roles of legislation, education, and reinforcement in changing road user behaviour*. Toronto, ON: Ontario Ministry of Transportation.
- Lund, A. K., A. F. Williams, and P. Zador. 1986. High school driver education: Further evaluation of the DeKalb County study. *Accident Analysis and Prevention* 18 (4):349-57.
- Masten, S. V., and E. A. Chapman. 2003. *The effectiveness of home-study driver education compared to classroom instruction: The impact on student knowledge, skills, and attitudes. Report to the legislature of the state of California*. Report no. CAL-DMV-RSS-03-203. Sacramento, CA: California Department of Motor Vehicles.
- Masten, S. V., and E. A. Chapman. 2004. The effectiveness of home-study driver education compared to classroom instruction: The impact on student knowledge and attitudes. *Traffic Injury Prevention* 5 (2):117-21.
- Maycock, G. 1995. Accidents in the first three years of driving. *TRL Annual Review*: 71-7.
- Maycock, G., and E. Forsyth. 1997. *Cohort study of learner and novice drivers. Part 4: Novice driver accidents in relation to methods of learning to drive, performance in the driving test and self assessed driving ability and behaviour*. Crowthorne, Berkshire: Transport Research Laboratory.
- Mayhew, D. R., and H. M. Simpson. 1995. *The role of driving experience: Implications for the training and licensing of new drivers*. Toronto, ON: Insurance Bureau of Canada.
- Mayhew, D. R., and H. M. Simpson. 1997. *Effectiveness and role of driver education and training in a graduated licensing system*. Ottawa, ON: Traffic Injury Research Foundation.
- Mayhew, D. R., and H. M. Simpson. 2002. The safety value of driver education and training. *Injury Prevention* 8 (Supplement II):3-8.

- Mayhew, D. R., H. M. Simpson, and A. Pak. 2003. Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis and Prevention* 35 (5):683-91.
- McKenna, C. K., B. Yost, R. F. Munzenrider, and M. L. Young. 2000. *An evaluation of driver education in Pennsylvania*. Harrisburg, PA: Pennsylvania Department of Transportation.
- McKnight, A. J., and A. S. McKnight. 2003. Young novice drivers: Careless or clueless? *Accident Analysis and Prevention* 35 (6):921-25.
- National Transportation Safety Board. 2005. *National Transportation Safety Board public forum on driver education and training, October 29-30, 2003. Report of proceedings*. Washington, DC: NTSB.
- Nichols, J. L. 2003. *Driver education review*. Paper presented at the National Transportation Safety Board Forum on Driver Education and Training, October, 2003. Washington, DC: NTSB.
- OECD. 1986. *Effectiveness of road safety education programmes*. Paris: Road Transport Research, Organisation for Economic Co-operation and Development.
- OECD Scientific Expert Group. 1990. *Behavioural adaptations to changes in the road transport system*. Paris: Road Transport Research, Organisation for Economic Co-operation and Development.
- Ottoson, J. 2006. *Using evaluation theory to map evaluation practice*. CDC/AEA Summer Evaluation Institute, Atlanta, GA, June 12-14, 2006.
- Patton, M. Q. 1997. *Utilization-focused evaluation: The new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications.
- Plato, K. C., and A. Rasp, Jr. 1983. *Parental views on issues related to traffic safety education and the licensing of teenage drivers*. Olympia, Washington: Washington Office of the State Superintendent of Public Instruction.
- Porteous, N. L., B. J. Sheldrick, and P. J. Stewart. 1997. *A program evaluation tool kit—A blueprint for public health management*. Ottawa, ON: Public Health Research, Education and Development Program, Ottawa-Carleton Health Department.
- Potvin, L., F. Champagne, and C. Laberge-Nadeau. 1988. Mandatory driver training and road safety: The Quebec experience. *American Journal of Public Health* 78 (9):1206-9.
- Preusser, D. F. 1996. Licensing practices and crash risk in the United States. In *New to the road: Reducing risks for young motorists. Proceedings of the First Annual Symposium of the Youth Enhancement Service, June 8-11, 1995*. H. Simpson (ed.). Los Angeles, CA: Youth Enhancement Services, University of California.
- Preusser, D. F., S. A. Ferguson, and A. F. Williams. 1997. *Effect of teenage passengers on the fatal crash risk of teenage drivers*. Arlington, VA: Insurance Institute for Highway Safety.
- Ray, H. W., M. Sadof, J. Weaver, J. R. Brink, and J. R. Stock. 1980. *Safe Performance Secondary School Driver Education Curriculum Demonstration Project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

- Road Safety Educators' Association. 2005. *Process for curriculum review*. Ontario, Canada: DSAA.
- Roberts, I., I. Kwan, and the Cochrane Injuries Group Driver Education Reviewers. 2004. School based driver education for the prevention of traffic crashes (Cochrane Review). In *The Cochrane Library* (1), Chichester, UK: John Wiley & Sons.
- Robertson, L. S. 1980. Crash involvement of teenaged drivers when driver education is eliminated from high school. *American Journal of Public Health* 70 (6):599-603.
- Robertson, L. S., and P. L. Zador. 1978. Driver education and fatal crash involvement of teenage drivers. *American Journal of Public Health* 68 (10):959-65.
- Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2004. *Evaluation. A systematic approach*. 7th ed. Thousand Oaks, CA: Sage Publications.
- Scriven, M., quoted in Davidson, E. J. 2004. *Evaluation methodology basics*. Thousand Oaks, CA: Sage Publications.
- Shope, J. T., and L. J. Molnar. 2003. Graduated driver licensing in the United States: Evaluation results from the early programs. *Journal of Safety Research* 34 (1S):63-9.
- Siegrist, S. 2003. *Driver training and licensing—The European perspective*. Paper presented at the National Transportation Safety Board Forum on Driver Education and Training, October, 2003. Washington, DC: NTSB.
- Simpson, H., ed. 1996. *New to the road: Reducing the risks for young motorists*. First Annual International Symposium of the Youth Enhancement Service, June 8-11, 1995: Youth Enhancement Service.
- Smiley, A., L. Lonero, and M. Chipman. 2004. *Final report. A review of the effectiveness of driver training and licensing programs in reducing road crashes*. Paris, France: MAIF Foundation.
- Smith, D. L. 1983. The DeKalb Driver Education Project—The same mistakes: Improper criteria. *Journal of Traffic Safety Education* 30 (January):14 & 26.
- Smith, M. F., and J. Blatt. 1987. *Summary of preliminary results: Follow-up evaluation. Safe Performance Curriculum Driver Education Project*. Spokane, WA: American Driver & Traffic Safety Education Association.
- Stock, J. R., J. K. Weaver, H. W. Ray, J. R. Brink, and M. G. Sadof. 1983. *Evaluation of Safe Performance Secondary School Driver Education Curriculum Demonstration Project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Strang, P. M., K. B. Deutsch, R. S. James, and S. M. Manders. 1982. *A comparison of on-road and off-road driver training*. Canberra City, Australia: Department of Transport and Construction Office of Road Safety.

- Stufflebeam, D. L. 2001. Evaluation models. *New Directions for Evaluation* Spring 2001 (89):1-106.
- Stufflebeam, D. L. 2003. *The CIPP model for evaluation*. Paper presented at Annual Conference of the Oregon Program Evaluators' Network, October 3, at Portland, Oregon.
- Taylor-Powell, E., L. Jones, and E. Henert. 2002. *Enhancing program performance with logic models*. Madison, Wisconsin: University of Wisconsin Extension. Available from <http://www1.uwex.edu/ces/lmcourse>.
- The Center for the Advancement of Community Based Public Health. 2000. *An evaluation framework for community health programs*. Durham, NC: CPBH.
- The Health Communication Unit. 2006. *Evaluating health promotion programs*. Toronto, ON: Centre for Health Promotion, University of Toronto.
- Trochim, W. 2001. *The research methods knowledge base*. Cincinnati, OH: Atomic Dog Publishing.
- Vernick, J. S., G. Li, S. Ogaitis, E. J. MacKenzie, S. P. Baker, and A. C. Gielen. 1999. Effects of high school driver education on motor vehicle crashes, violations, and licensure. *American Journal of Preventive Medicine* 16 (1S):40-6.
- Weaver, J. K. 1987. *Summary of preliminary results. Follow-up evaluation. Safe Performance Curriculum Driver Education Project*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Wiggins, S. 2005. *Graduated Licensing Program interim evaluation report - Year 3*. Vancouver, BC: Insurance Corporation of British Columbia.
- Wilde, G. J. S. 1994. Risk homeostasis theory and its promise for improved safety. In *Challenges to accident prevention: The issue of risk compensation behaviour*. R. M. Trimpop and G. J. S. Wilde (eds.). Groningen: Styx Publications.
- Williams, A. F. 2003. Teenage drivers: Patterns of risk. *Journal of Safety Research* 34 (1S): 5-15.
- W.K. Kellogg Foundation. 2004. *W.K. Kellogg Foundation evaluation handbook*. 1998. Available from <http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf>.
- Woolley, J. 2000. *In-car driver training at high schools: A literature review*. Adelaide, Australia: Transport SA.
- Wynne-Jones, J. D., and P. M. Hurst. 1984. *The AA driver training evaluation*. Traffic Research Report No. 33. Wellington, New Zealand: Ministry of Transport.
- Zhao, J., R. E. Mann, M. Chipman, E. Adlaf, G. Stoduto, and R. G. Smart. 2006. The impact of driver education on self-reported collisions among young drivers with a graduated license. *Accident Analysis and Prevention* 38: (1)35-42.

# EVALUATION RESOURCES\*

## 1. Examples of Program Evaluation Guidelines, Tools, and Resource Materials

*A Program Evaluation Tool Kit—A Blueprint for Public Health Management*

Nancy L. Porteous, Barbara J. Sheldrick, and Paula J. Stewart.

Public Health Research, Education and Development Program, Ottawa-Carleton Health Department, Ottawa, Ontario. 1997

[http://ottawa.ca/city\\_services/grants/toolkit/index\\_en.shtml](http://ottawa.ca/city_services/grants/toolkit/index_en.shtml)

*An Evaluation Framework for Community Health Programs*

The Center for the Advancement of Community Based Public Health. 2000

<http://www.cdc.gov/eval/evalcbph.pdf>

*Basic Guide to Program Evaluation*

Carter McNamara. 2000

[http://www.mapnp.org/library/evaluatn/fnl\\_eval.htm](http://www.mapnp.org/library/evaluatn/fnl_eval.htm)

*Evaluating Health Promotion Programs Workbook*

Centre for Health Promotion, University of Toronto, 2006

<http://www.thcu.ca/infoandresources/publications/EVALMasterWorkbookv3.6.03.06.06.pdf>

*Evaluation in Health Promotion: Principles and Perspectives*

WHO, CDC and Health Canada

[http://www.euro.who.int/eprise/main/WHO/InformationSources/Publications/Catalogue/20040130\\_1](http://www.euro.who.int/eprise/main/WHO/InformationSources/Publications/Catalogue/20040130_1)

*Key Evaluation Checklist*

Michael Scriven

<http://www.wmich.edu/evalctr/checklists/kec.htm>

*Knowledge Required to Perform the Duties of an Evaluator*

D. J. Caron

The Canadian Journal of Program Evaluation, 1993, V. 8, No. 1, 59-78

*Professional Development Modules on Key Topics in Evaluation*

Online Evaluation Resource Library

<http://oerl.sri.com/module/modules.html>

---

\*The websites listed in this section were correct at time of printing and are for informational purposes only. AAA Foundation does not endorse any particular organization or website.

*Program Evaluation Kit*

First 5 LA (Los Angeles County Children and Families First Proposition 10 Commission)  
Research and Evaluation Department. 2003  
[http://www.first5.org/docs/Community/CommRsrc\\_EvalKit\\_0603.pdf](http://www.first5.org/docs/Community/CommRsrc_EvalKit_0603.pdf)

*Programme Manager's Planning, Monitoring, and Evaluation Toolkit*

The United Nations Population Fund (UNFPA)  
<http://www.unfpa.org/monitoring/toolkit.htm>

*Project STAR—Support and Training for Assessing Results*

*A comprehensive guide to decisions involved in the evaluation process.*

[http://nationalservicerresources.org/resources/online\\_pubs/perf\\_meas/ac\\_pm\\_worksheets.php](http://nationalservicerresources.org/resources/online_pubs/perf_meas/ac_pm_worksheets.php),  
and [http://nationalservicerresources.org/resources/online\\_pubs/perf\\_meas/usersguide.php](http://nationalservicerresources.org/resources/online_pubs/perf_meas/usersguide.php)

*Resources*

Information about evaluation or assistance in conducting an evaluation project  
CDC Evaluation Working Group  
<http://www.cdc.gov/eval/resources.htm>

*Taking Stock—A Practical Guide to Evaluating Your Own Programs*

Sally L. Bond, Sally E. Boyd, and Kathleen A. Rapp  
Horizon Research, Inc. 1997  
<http://www.horizon-research.com/publications/stock.pdf>

*The Community Toolbox, Part J, Evaluating Community Programs and Initiatives*

Work Group on Health Promotion and Community Development University of Kansas  
[http://ctb.ku.edu/tools/en/part\\_J.htm](http://ctb.ku.edu/tools/en/part_J.htm)

*The Framework for Program Evaluation in Public Health*

CDC Evaluation Working Group  
<ftp://ftp.cdc.gov/pub/Publications/mmwr/rr/rr4811.pdf>

*Utilization-Focused Evaluation Checklist*

Michael Quinn Patton  
<http://www.wmich.edu/evalctr/checklists/ufechecklist.htm>

*W.K. Kellogg Foundation Evaluation Handbook*

W.K. Kellogg Foundation  
<http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf>

*W.K. Kellogg Foundation Logic Model Development Guide*

W.K. Kellogg Foundation  
<http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf>

## 2. Research and Evaluation Methodology and Statistical Resources

Babbie, E. 2005. *The basics of social research*. Belmont, CA: Thomson Wadsworth.

Bourque, L. B., and E. P. Fielder. 1995. How to conduct self-administered and mail surveys. *Survey Kit Series, Volume 3*. A. Fink (ed.). Thousand Oaks, CA: Sage Publications.

Davidson, E. J. 2004. *Evaluation methodology basics*. Thousand Oaks, CA: Sage Publications.

Fink, A. 1995. How to ask survey questions. In *Survey Kit Series, Volume 3*. A. Fink (ed.). Thousand Oaks, CA: Sage Publications.

Fink, A. 1995. How to measure survey reliability and validity. In *Survey Kit Series, Volume 7*. A. Fink (ed.). Thousand Oaks, CA: Sage Publications.

Frey, J. H., and S. Mertens Oishi. 1995. How to conduct interviews by telephone and in person. In *Survey Kit Series, Volume 4*. A. Fink (ed.). Thousand Oaks, CA: Sage Publications.

Herman, J., M. L. Lyons, and C. Fitz-Gibbon. 1987. *Evaluator's handbook*. Thousand Oaks, CA: Sage Publications.

Krueger, R. A. 1994. *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage Publications.

Miller, D. C. 1991. *Handbook of research design and social measurement*. Thousand Oaks, CA: Sage Publications.

Munro, B. H. 2001. *Statistical methods for health care research*. 4th ed. New York: Lippincott.

Neutens, J. J., and L. Rubinson. 2001. *Research techniques for the health sciences*. 3rd ed. San Francisco, CA: Benjamin Cummings.

Patton, M. Q. 1997. *Utilization-focused evaluation: The new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Patton, M. Q. 2001. *Qualitative research and evaluation methods*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2004. *Evaluation. A systematic approach*. 7th ed. Thousand Oaks, CA: Sage Publications.

Rush, B., and A. Ogborne. 1991. Program logic models: Expanding their role and structure for program planning and evaluation. *Canadian Journal of Program Evaluation* 6 (2):96-106.

Scriven, M. 1991. *Evaluation thesaurus*. 4th ed. Thousand Oaks, CA: Sage Publications.

Shadish, W. R., T. D. Cook, and T. D. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shadish, W. R., T. D. Cook, and L. C. Leviton. 1991. *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage Publications.

Trochim, W. M. K. 2001. *The research methods knowledge base*. Cincinnati, OH: Atomic Dog Publishing. Available from <http://www.socialresearchmethods.net>

*Using Excel for Evaluation Data*

<http://www.metrokc.gov/health/APU/healthed/emanual.htm>

*What is a Survey? Series*. American Statistical Association

<http://www.amstat.org/sections/srms/whatsurvey.html>

Wholey, J. 1987. Evaluability assessment: Developing program theory. *New Directions for Program Evaluation*, 33:77-92.

### 3. Organizations

AAA Foundation for Traffic Safety

<http://www.aaafoundation.org>

American Driver and Traffic Safety Education Association (ADTSEA)

<http://adtsea.iup.edu/adtsea>

National Highway Traffic Safety Administration (NHTSA)

<http://www.nhtsa.dot.gov>

National Institute for Driver Behavior (NIDB)

<http://www.nidb.org>

The Driving School Association of the Americas (DSAA)

<http://www.thedsaa.org>

Transportation Research Board

<http://www.trb.org>

University of Michigan Transportation Research Institute (UMTRI)

<http://www.umtri.umich.edu>

University of North Carolina's Highway Safety Research Center (UNC-HSRC)

<http://www.hsrb.unc.edu>

## 4. Evaluation Consultants

### *Getting and Working with Consultants*

Carter McNamara

Management Assistance Program for Nonprofits

Links to information to help organizations find, hire, and work with evaluation consultants

<http://www.mapnp.org/library/staffing/outsrcng/consult/consult.htm>

### *Resume Bank*

American Evaluation Association

Links to resumes of AEA members who are available as evaluation consultants

[http://www.eval.org/find\\_an\\_evaluator/evaluator\\_search.asp](http://www.eval.org/find_an_evaluator/evaluator_search.asp)

## 5. Evaluation Training

### *Building Evaluation Capacity. 72 Activities for Teaching and Training*

H. Preskill and D. Russ-Eft. 2005

Thousand Oaks, CA: Sage Publications

### *Events Directory*

American Evaluation Association

<http://www.eval.org/Training/eventsdir.asp>

The Evaluators' Institute

<http://www.evaluatorsinstitute.com>



# **APPENDIX A: Review of the Literature: Driver Education Evaluation—Past and Present**

## **INTRODUCTION**

This Appendix provides:

- An overview of the background and context for driver education and evaluation;
- A discussion of the conclusions of recent reviews of evaluations; and
- A detailed examination of the methods and findings of recent and some important older individual evaluations.

The aims of the review are to provide a richer understanding of driver education evaluation and perspectives on how driver education evaluation can best be understood and improved in the context of driver education policy, program planning, and program management. It concludes with a discussion of the limitations and implications of the evaluation literature for driver education program practice and development.

### **Young Novice Drivers**

When discussing driver education, it is important to keep in mind the difficulties of young novice drivers that driver education hopes to mitigate. In all motorized jurisdictions around the world, young, inexperienced drivers have much higher crash rates than older, more experienced drivers. The crash rate per mile of U.S. 16-year-olds is 10 times the rate of experienced adult drivers (Williams 2003). Crash risk also declines rapidly over the first few months and the first few hundred miles of driving (Mayhew, Simpson, and Pak 2003; McKnight and McKnight 2003). Sixteen-year-olds have almost three times the crash rate of 18-year-olds (Evans 1987). While risk drops rapidly, it takes a very long time to level off, requiring as much as 10 years of driving before reaching mature rates (Evans 1991).

Limited skills and abilities contribute to the elevated risk of novice drivers. They are less able than experienced drivers to control attention, scan the environment effectively, detect potential hazards early, and make tough decisions quickly. They perceive less risk in some specific violations and high-risk situations, but they may perceive more risk in certain lower-risk situations. Skill deficits, however, are not necessarily the only problem.

Novice drivers also tend to raise their risk through seemingly deliberate choices. They tend to drive too fast, too close to others, to accept small gaps in traffic, have unrealistic confidence in their own abilities, and leave inadequate safety margins (Boyce and Geller 2002; Lonero et al. 1995). Of course, some risky choices may result from the inability to anticipate and perceive risks.

Whether inexperienced driving skills or immature decision making is the key problem has been long argued. Wilde (1994) pointed out that: 1) the excess risk of new drivers is an international phenomenon; 2) this risk holds true both per distance driven and per person; and 3) it is due to both immaturity and inexperience. On the other hand, recent research suggests that inexperience might be most important in the early, very high-risk driving, as the risk declines substantially over the first few months (Mayhew, Simpson, and Pak 2003). Presumably age-related maturity develops over a longer timeframe and cannot be primarily responsible for the rapid change in risk over the first few months of driving, when driving experience builds rapidly.

James and Scott McKnight (2003) studied the records of non-fatal crashes of young novice drivers. They concluded that collision reports typically evidence simple mistakes, seemingly consistent with inexperienced skill failures rather than extravagant risk taking. The most prominent errors included lack of visual search prior to turning left, not watching the car ahead, driving too fast for conditions, and failing to adjust adequately for wet road surfaces. These error patterns did not change across the 16-19 age range.

If non-fatal crashes of young drivers seem to be mainly precipitated by relatively minor and inadvertent errors, one might reasonably expect to see a different pattern of errors in fatal crashes, which differ in many ways from the patterns of less severe crashes. Many young driver fatal crashes involve a single vehicle. A recent study of U.S. Fatality Analysis Reporting System (FARS) data for the State of Colorado (Gonzales et al. 2005) suggests a much higher incidence of violations in young driver fatal crashes than in mature driver fatal crashes (e.g., speeding—1.9 times higher, driving recklessly—4.8 times higher). The researchers also found lower incidence of some risk factors, such as alcohol impairment and adverse weather conditions. This study supports the expectation that young driver fatal crashes are different from older drivers' fatal crashes and from young drivers' non-fatal crashes. The differences that were found, however, although seemingly substantial, do not appear to be statistically significant. Although the study's report does not address significance of the odds ratios found, it does provide 95% confidence intervals, which seem to bracket all the odds ratios presented. Waiting until another study with larger numbers of cases is conducted will be necessary to validate these differences through fatality data.

Nevertheless, inadvertent errors and unsafe choices may both contribute to young novice drivers' excess risk, albeit perhaps not in the same proportions for differing severities of crashes and at different times in the early driving career. This implies that both error avoidance and healthier choices should be effectively addressed in driver education and serve as evaluation targets for longer-term driver education evaluation.

## **PROGRAM CONTEXT FOR DRIVER EDUCATION EVALUATION**

For the purposes of this review, driver education means beginner, pre-licensing driver instruction. Other forms of driver instruction are mainly beyond its scope.

Driver education has long had been mandated to address all possible aspects of the tragically high crash risk of young novice drivers. Courses for beginners have been a popular and convenient

means of achieving independent mobility, important for both young people and their parents. Driver education has strong “face validity” as a safety measure. Parents think it makes their children safer drivers (Fuller and Bonney 2003, 2004; Plato and Rasp 1983).

A recent concise and detailed history of driver education can be found in a paper by Dr. James Nichols (2003) prepared for the 2003 U.S. National Transportation Safety Board (NTSB) hearings on driver education and available in the NTSB (2005) report on driver education. The NTSB hearings and recommendations may be a turning point toward more organized, systematic, and evidence-based development, after a long period of relatively little centralized support or direction for driver education R&D.

In recent years, many changes in the ways in which driver education operates have been the result of technology “push” rather than having been pulled by pedagogical, epidemiological, or evaluation research on driver education and its target audiences. Nevertheless, as a result of recent changes, driver education is now highly diverse in its delivery and organization. As well as traditional high school driver education, evaluators must recognize the importance of driving schools, home schooling, and computer- and web-based instruction.

Even though less development appears in content than in delivery, greater diversity in content now exists. This diversity complicates answering general evaluation questions such as, “Does driver education work?” and “How can its effects be improved?” This review examines how these questions have been asked and answered in the past to provide a more in-depth understanding of driver education evaluation, and to identify evaluation implications for improving driver education policies and practices in the more complex and diverse future.

The term “driver education” has usually been applied to programs consisting of both classroom theory and in-car practical training, particularly in North America. Traditional driver education has taken place in a single stage, before the driver becomes licensed. Indeed, one principal purpose of driver education is to prepare beginners for license testing. In many jurisdictions, most beginner drivers receive some instruction from a paid professional instructor (Maycock and Forsyth 1997).

Driver education has been widely available in public secondary schools in North America, although availability has declined in most jurisdictions since the early 1980s. This is especially clear in the United States, where high school driver education was pervasive. In Canada, driver education has always been more diverse. High school driver education has remained strong in some provinces, never existed in others, and has declined in still others in a manner similar to the United States. A project to assess driver education status across Canada has been completed by Natural Resources Canada (Lonerio and Clinton 2006), and a similar federal project is in development in the United States, so a clearer picture of North American driver education should soon be available.

Traditional high school driver education of 30 hours in the classroom and 6 hours in the car still predominates in the United States. In some Canadian jurisdictions, the in-car requirements

have been expanded to 8 or 10 hours. Classroom methods typically consist of teacher-centered lectures, with some discussion supported by film and video. Many commercial driver education providers use similar program structures.

Driver education also traditionally meant instruction only before the new driver was licensed to drive independently, but that distinction has become somewhat blurred. A less-common but potentially important form of driver instruction is addressed to drivers after they are licensed to drive independently. In a few jurisdictions, such as Michigan, Finland, and Luxembourg, new drivers are required to take a second stage of training after they have been driving as licensed drivers for a short period of time (e.g., Glad 1988; Keskinen, Hatakka, and Katila 1998; Shope and Molnar 2003). Some safety benefits of these second-stage programs have been observed, although in a limited range of evaluations.

In addition to current movement toward multi-stage education, trends toward changes in instructional method and program delivery also exist. Traditionally, all driver education activities involved face-to-face interaction between instructor and learner, although classroom instruction was often supported with film and video media, and sometimes simulators. More recently self-instruction, computer-based instruction, simulation, and even web-based instruction have become prevalent, particularly in parts of the United States.

While good instructional methods can facilitate learning of cognitive and psychomotor skills, better knowledge and skills do not automatically lead to fewer crashes. Improvements in safety probably require safer driving habits. That is, to be safer, better-trained drivers must actually drive differently than they would otherwise. People do not always do what they know how to do to minimize their risk, or even what their attitudes say they should do.

Behavior change toward habitual lower risk behaviors is much harder to accomplish than is generally understood (Lonerio and Clinton 1998; Lonerio et al. 1994). As an effective behavior change intervention, driver education's potential seems excellent on the surface. The typical face-to-face instructional setting, with rehearsal and practice opportunities, and two-way communication and feedback, ought to provide strong support to forming appropriate habits and skills. Indeed, evidence of effective gain in knowledge and skills existed in the DeKalb driver education experimental study (Stock et al. 1983).

While more education is always a popular prescription for improving safety, demonstrated effectiveness in improving safety performance solely through educational measures of any form is relatively rare. In most jurisdictions, a small portion of licensed drivers receive some further instruction, commonly referred to as "defensive driving," often delivered in an employment setting. While evaluations of these programs lie beyond the scope of this review, these programs have not usually been found to have a measurable safety impact, although they occasionally have shown some effect in reducing convictions. Violator schools, widely used as diversions from court and licensing sanction procedures for drivers receiving traffic tickets, have also been shown to be ineffective in reducing crashes (Lonerio et al. unpublished). Road safety is not alone in having difficulties demonstrating beneficial bottom-line effects of education—other health and safety fields have also been unable to do so (Lonerio and Clinton 1998; Lonerio et al. 1994).

## DRIVER EDUCATION EVALUATION OVERVIEW

Most courses are considered successful if learning objectives are met at the end. Driver education's assignment is tougher, because it is expected to produce improved subsequent driving behavior and measurable effects on crashes. Crashes have complex causation, some aspects of which are outside the driver's control, and are rare events even among high-risk drivers.

Compared to the public health and education fields, beginner driver education has seen relatively few evaluations, mostly assessing short-term, direct safety impacts. These impacts have typically been defined as total subsequent reported crashes of graduates compared to new drivers who learned to drive in other ways, and who passed the same licensing tests. Numerous evaluations have compared different forms of formal driver education, with some including a control group of new drivers who learned to drive through their family or some other form of instruction.

The research design structure of driver education evaluation has taken various forms, including:

- Experimental studies that have involved randomly assigning drivers to various training conditions and comparing subsequent crash rates and other measures;
- Quasi-experimental studies that have observed differences between self-selected driver education students and those who learn to drive in other ways; and
- Ecological studies that have considered impacts on crashes following jurisdictional changes in requirements or support for formal driver education.

The largest experimental and most influential driver education evaluation ("DeKalb") involved randomly assigning 16,000 U.S. high school student volunteers to three groups—intensive training, minimal training, or no formal driver education. The results failed to show a dramatic, long-term benefit of a special course (Stock et al. 1983), and reactions to the results had profound effects on driver education. In the United States, driver education market penetration peaked in the early 1980s with about 80% of new drivers being formally trained. Afterward, however, many high school driver education programs were dropped. New Jersey schools offering driver education, for example, dropped from 96% to 40% between 1976 and 1986 (Simpson 1996). It has not been clearly demonstrated whether the DeKalb results in effect caused the decline in U.S. high school driver education, or whether it served as support for budget cutting.

Other studies than DeKalb have also found that driver education failed to produce a direct measurable change in crash rates of graduates compared to others. Although positive findings have also been found, many members of the safety research community have come to believe that "driver education does not work." This conclusion raises questions as to how such a counter-intuitive situation might be possible. Given the limited scope of beginner training, however, as well as its position at the very start of a long learning curve, the driver education experience may be overshadowed by other experiences, overconfidence, increased exposure to risk, and relaxed parental supervision. Since so much of drivers' learning takes place after licensing, potentially beneficial effects of traditional driver education may be offset by other influences.

And as researchers have also suggested, driver education in the past may not have provided the best possible content in the best ways (Mayhew and Simpson 1997).

Unfortunately, as will be discussed in more detail later, evaluations have also been rather unsystematic and limited in quantity. Even the randomized controlled trial (RCT) experiments have suffered methodological problems that make their results less than definitive. Some experimental studies used small samples and lacked statistical power to detect modest effects (Engström et al. 2003). A recent Australian quasi-experimental evaluation observed substantial crash differences between training conditions, but, because the numbers of drivers were so small, it could not conclude that the differences were the result of anything other than chance (Haworth, Kowadlo, and Tingvall 2000). Very large numbers of cases are needed, even to assess effects on total crashes, let alone injury or fatal crashes.

Other study design problems also have reduced evaluation benefits for driver education. Most evaluations have failed to look at intermediate student outcomes—knowledge, skills, attitudes, intentions, or values that have (or have not) been affected, instead of simply looking at crash rates. Ways to improve driver education programs have, therefore, been unclear (Lonero et al. 1994; 1995). Intermediate outcome measures and survey tracking of behavior during the follow-up period can provide something akin to an “audit trail” of program outcomes, such as knowledge, skill, attitudes, and exposure to risk (rather than just looking at final impacts on crash rates) which in turn, lead to safety impacts.

Overall, scientific evaluation of driver education has been quite limited in quantity, quality, and scope. Beginner driver education is particularly hard to evaluate, at least in terms of safety impacts, because suitable comparison groups are hard to establish. Many earlier evaluations compared groups of young drivers who not only received different forms of driver education, but also differed in other ways that might affect their driving record and other results. These “extraneous” or “confounding” factors include location of residence, income, or other important socioeconomic factors. Even when they can be established, comparable treatment and control groups are hard to maintain over time, since assigned or selected groups can have different dropout rates.

Key areas where driver education evaluation has been found lacking include:

- **Program theory:** Theory in the sense used here means the logic that justifies thinking a program should meet its goals—that is, why we think it should work. There has been little evaluation of the theory underlying various driver education programs.
- **Formative evaluation:** Formative evaluation involves applying evaluation tools to improve the content and delivery of a program. There has been little evaluation of intermediate effects, so it is not clear how well driver education students achieve, retain, and use desired skills and knowledge. Driver education courses vary greatly in quality, and limited evaluation of program differences has existed.
- **Methodological soundness:** In the relatively small pool of existing evaluations, problems

of scope, design, and sampling limit unequivocal conclusions about driver education's present ultimate value and how its impact might be improved in the future.

The following review briefly discusses conclusions of the most significant recent reviews of earlier driver education evaluations. The more recent individual quantitative evaluations, along with selected older evaluations, are discussed in the subsequent section.

## **REVIEWS OF DRIVER EDUCATION EVALUATIONS**

Reviews of driver education evaluations predominantly attempt to determine the extent that evaluations indicate whether driver education works and how it can be improved. This review has a somewhat broader focus. It provides a perspective on driver education evaluation in the context of program planning and management, and on how evaluation might be improved to play a stronger and more positive role in future driver education development. Evaluation is a potentially valuable and crucial tool for improving driver education, but it has not yet achieved its full potential.

In the last decade or so, a number of broad reviews of evaluation studies of beginner driver education have appeared (Christie 2001; Engström et al. 2003; Lonero et al. 1994, 1995; Mayhew and Simpson 1997, 2002; Siegrist 2003; Smiley, Lonero, and Chipman 2004; Woolley 2000). Recent reviews have usually addressed driver education in conjunction with other forms of driver instruction or graduated driver licensing (GDL).

GDL introduces stages into the licensing procedure. GDL systems generally require that new drivers be accompanied by a licensed driver in the first stage, either for a fixed time period or until a specified age is reached. This is followed by a period in which some aspects of driving are restricted. The aim of the restrictions is to avoid circumstances known to be associated with high crash risk for young drivers. Restrictions vary widely among jurisdictions and may include driving at night, driving with teenage passengers, and driving while under the influence of alcohol. The requirements to spend an extended period of supervised driving with a licensed driver as the front seat passenger, and curfews of various kinds tend to reduce the amount of driving. Most evaluations have shown some reduction in crashes after GDL systems were introduced.

Early driver education evaluations used totally uncontrolled comparisons between driver education graduates and others. That is, studies failed to control for the effects of extraneous, confounding factors. Nichols summarized the findings of the early evaluations as follows:

Although there were a few early studies which reported negative results, the majority of studies conducted at the time suggested that: (1) while the effects may be short-lived, driver education students had fewer accidents and violations than non-students; (2) complete courses involving both classroom and behind-the-wheel training were more effective than classroom-only courses; and (3) [High School Driver Education] was more effective in reducing accidents and violations than either parent training or commercial driver training. (Nichols, 2003, 20)

Early studies made no effort to control for the ways in which driver education grads were different from comparison groups other than the type of training each group had received. As a result, these uncontrolled quasi-experiments were not considered credible. Some later evaluations were designed as full experiments, that is, randomized controlled trials (RCTs). They controlled for extraneous differences between driver education and comparison groups by random assignment to a training group or a control group (comprised of informally trained novice drivers or those who took a different course).

## **Systematic Reviews**

Two recent systematic reviews of a small number of evaluations have appeared. Vernick et al. (1999) reviewed nine evaluations that met their methodological criteria (of 27 evaluations found). The review's stated intent was broader than most, aimed at finding: 1) whether driver education graduates were less likely to crash or more likely to become licensed to drive; and 2) whether driver education had broader public health effects in lowering community rates of crashes. All but one of the nine studies addressed U.S. high school programs. Five of the studies reviewed were structured as large-scale ecological record modeling studies, including two each from Levy (1988; 1990) and Robertson (1980; Robertson and Zador 1978). The other four studies were experimental RCTs. Three of the RCTs consisted of one of the original DeKalb experiment reports and two re-analyses of this data. The fourth was an RCT with fewer than 800 subjects assigned to four different training conditions, which, not surprisingly given the small sample, found no significant differences in crashes (Strang et al. 1982). The reviewers concluded that no study that met their design criteria showed a "significant individual or community-level beneficial effect of driver education for high school-aged students" (44). No explanation is offered for disregarding the findings of a significant beneficial effect on fatal crashes by Levy (1990).

Using an even narrower selection basis than the Vernick et al. review, Roberts and Kwan (2004) reviewed three RCT experimental evaluations, all from the early 1980s. They also concluded that no evidence showed safety impacts of driver education. The utility of this review of old and well-known studies is difficult to discern. Neither of the systematic reviews included the RCT study by Dreyer and Janke (1979), which found a positive effect on drivers' crash records. The narrow orientation of this approach to systematic review seems to limit its applicability to driver education evaluation, since most evaluations would not meet the criteria for inclusion, and RCTs present a special difficulty in evaluating beginner driver education, as will be discussed more fully at the end of this review.

## **Review of Driver Education in Graduated Licensing**

Graduated licensing that delays independent driving has been the principal initiative to address young driver crashes in recent years and has been shown to be effective in reducing crashes. At the Traffic Injury Research Foundation, Mayhew and Simpson (1997) performed a detailed review of the DeKalb experiment and eight later evaluations of beginner driver education, in the wider context of graduated licensing and other forms of driving instruction. These researchers indicated some positive findings for driver education effectiveness for novice car drivers, including:

- Per-licensed driver analyses of the short-term DeKalb data (Stock et al. 1983);
- Minimal training condition in a long-term follow up in the DeKalb experiment (Smith and Blatt 1987);
- Quebec mandatory training system (Potvin, Champagne, and Laberge-Nadeau 1988);
- Norwegian second-stage training in night driving (Glad 1988);
- Denmark mandatory driver education (Carstensen 1994); and
- Finland mandatory slippery road training for older novices (Katila et al. 1995).

They also pointed out that two U.S. studies in the 1980s showed better safety effects from school-based driver education than commercial schools, but these were uncontrolled quasi-experiments, similar to the early evaluations and have low credibility. The econometric modeling study of driver education over 47 U.S. states (Levy 1990), which also showed a positive effect for driver education, was not included in this review. Also not included was the 1979 California RCT study by Dreyer and Janke.

Over a wide range of driving instruction, the authors concluded that 7 studies showed a positive effect, 16 showed no effect, and about 7 others showed a negative safety effect. They summarized their findings as follows:

On balance, the weight of the available evidence does not favor the hypothesis that formal instruction has safety benefits. Indeed there is precious little in the way of reliable evidence to show that formal instruction provides safety benefits. As counterintuitive as this may seem, it is difficult to reach a different conclusion in the face of the total body of evidence. (Mayhew and Simpson 1997, 45)

Simpson and Mayhew then recommended three key directions for future improvement of driver education:

1. Driver education should be multi-phased;
2. GDL systems should not give a “time discount” for driver education; and
3. Specific changes should be implemented in the content and delivery of driver education.

Mayhew and Simpson (2002) revisited the evidence in a later brief review, which also discussed other recent reviews. Similar conclusions resulted.

## **Australian Reviews of Driver Instruction**

Christie (2001) in Australia recently published an influential and detailed review of evaluations of various forms of driver instruction, including beginner driver education. He reviewed the

same studies as Mayhew and Simpson, as well as later ones, including one from Pennsylvania that will be discussed in the section on individual evaluations (McKenna et al. 2000). Christie provided an analysis of a Tasmanian retrospective records-based study that was unavailable to the present review (Langford 1997, cited in Christie 2001). The Langford study apparently showed positive initial effects of a limited high school course, but later re-analysis of the same cohorts with fuller reporting of records indicated no differences.

Apparently less impressed than Mayhew and Simpson with the limited positive impacts of driver education found in the literature, Christie concluded that no evidence shows beneficial effects of beginner driver education. He reiterated the view that driver education is harmful because it induces earlier licensing. Christie summarized his view as follows:

New approaches to driver training may eventually prove to be useful in reducing casualty accident risk/involvement, but much research and development work remains to be done before one could say that driver training is an effective crash countermeasure. In the interim, other approaches such as increased supervision and graduated licensing for novice drivers and traffic law enforcement for all drivers are likely to make greater and more lasting contributions to road safety. (Christie 2001, 43)

In a later review of road safety education, Christie (2002) also addressed public information and advertising programs for road safety. He concluded they too are ineffective except as adjuncts to legislative and enforcement programs. Again, he suggests that other kinds of programs, including graduated licensing, enforcement, crashworthy vehicles, and “black spot” roadway hazard correction should be given priority. This broader view of road safety management as a context for driver education in all forms is well considered and consistent with earlier recommendations by Lonero et al. (1995).

Christie’s analyses, however, seem to stop short of applying the same standard of effectiveness to other kinds of countermeasures. Using standards of proof applied to driver education, most other behavioral safety measures cannot be shown to be effective for preventing crashes (Lonero et al. 1994). As the present authors have concluded from reviewing the full range of safety programs aimed at road user behavior, very few behavioral interventions work well in isolation. Similar conclusions regarding the need for coordinated programs of behavioral influence have also been drawn in other fields, such as health promotion, that attempt to change behavior (e.g., Green and Kreuter 1999; Lonero and Clinton 1998).

Planned and coordinated combinations of influences seem to work, but single-technique approaches typically do not. Unplanned and uncoordinated combinations of influences may even add up to changes in culture and behavior over time. Examples might include the changes in seat belt use and impaired driving, which have occurred despite individual educational and enforcement interventions often showing no effect or short-lived effects. Little or no research has yet addressed these broader areas of safety behavior change, which Christie correctly attempts to bring to the driver education discussion.

Another Australian review (Woolley 2000) concluded that non-skills factors are key to resolving road safety problems and that no conclusive link exists between skills-based training and crash involvement. Rather motivation and risk-taking propensity are more important than any type of skills-based training, and driver education should be developed to address these critical factors.

Difficult choices are presented when the benefit of a critical program, such as driver education, does not appear as expected. Should resources be shifted to other kinds of programs, as suggested by Christie? Or should additional resources be employed to improve the program to make it effective in the ways expected?

## **Swedish Review of High School Driver Education**

A brief recent review appeared as a chapter by Inger Engström and colleagues in a wider Swedish review of licensing and other behavioral measures for novice drivers (Engström et al. 2003). This review covered some of the same evaluations as DeKalb and its reanalyses, as well as Australian and New Zealand studies. It also covered some U.S. state program studies that are more like qualitative program reviews than the more typical quantitative evaluation studies. In addition, these researchers reviewed an unavailable small evaluation of a Swedish 30-hour high school pre-driver education course, and reported that survey results showed improved knowledge of traffic risks and laws.

This review briefly touched on evaluation methodology issues, pointing out the controversy over appropriate measures for assessing driver education effects. Engström et al. did, however, maintain the distinction between safety measures, as typically defined, using only intermediate criteria, such as behavior or attitude measures. Alternative ways of looking for safety effects seem to exist, beyond those that have been used so far in driver education evaluation. These researchers also pointed out the common problem of unacceptably small sample sizes, which has plagued the driver education evaluation field from its inception and continues in recent studies.

The authors concluded:

The current approach, as far as evaluated, has not shown the benefits that could be expected from education and training in high schools. It is thus of the highest importance to develop new strategies since the high school environment provides good possibilities to reach youngsters with safety issues that are not normally covered by driving schools or parents. (Engström et al. 2003, 93)

Engström et al. did not provide an explanation of why they think these possibilities are more likely to succeed than driver education itself, but reiterated the prescription of Woolley (2000) that driver education should have a supportive role in a holistic approach to road safety in schools.

## Norwegian Meta-Analysis of Evaluation Studies

Elvik and Vaa (2004) reported the results of an extensive meta-analysis of driver education evaluations from around the world. This approach, which combined the data findings of 16 individual studies, overcomes some of the difficulties with the individual evaluations, such as the common problem of too-small sample sizes. The meta-analytic approach recognizes that no single evaluation is likely to be definitive, and this astute study can be seen as the most definitive summary of the past evaluations. Unfortunately, the only report on the meta-analysis is a very brief one in the 2004 edition of the Norwegian *Handbook of Road Safety Measures*.

The combined data of all 16 studies indicated that driver education graduates have 1.9% fewer crashes per driver (confidence interval, -3.8%; 0%). The overall difference appears to have been nearly statistically significant, as the combined results had the power to detect a significant difference ( $\alpha = .05$ ) if the result would have been 2% (rather than 1.9%). Results reported per kilometer driven was a 4% lower crash rate for graduates (-6%; -2%), but statistical significance for this difference is not reported.

When the combined results were limited to the experimental studies, however, a different picture emerged. No difference per driver surfaced (+/- 4%). Per kilometer driven, driver education graduates had 11% more crashes (+8%; +15%), again the significance of which was not reported. The authors concluded that the combined evaluation results do not indicate that driver education reduces crashes over the first couple of years of driving.

Elvik and Vaa also examined, briefly, four possible explanations for the generally disappointing findings among driver education evaluations. The first explanation was that the evaluation research is too poor to detect the real effects of driver education. This is refuted by indicating that the research overall is actually somewhat better than the evaluation research typical in most road safety programs. They also suggested that only the poorest studies have found any positive effect, although significant positive effects have appeared even among the favored experimental studies (Dreyer and Janke 1979).

The second possible explanation for lack of positive findings was that programs evaluated are not good enough. The authors see this as unlikely, however, because the best programs are probably those that have been evaluated. This is plausible, except that most evaluations took place in the distant past, and most current programs have not been evaluated. The third explanation was that crashes are too insensitive a measure to detect training effects. The authors are able to refute this possible explanation of no effects, as they indicate that the combination of results across of all the 16 evaluations raises the statistical power enough to detect even a 2% crash difference.

The final potential explanation, favored by the authors, was behavioral adaptation—less-skilled drivers taking more care and better-skilled drivers taking less. While the meta-analysis results do not directly address this explanation, the authors cited the negative effects found in evaluations of skid training and Gregersen's (1996) study, which showed skid training could raise confidence

without actually increasing skill. Many researchers have been skeptical of behavioral adaptation, and even if this explanation is accepted, it might raise the question of whether the best current or future driver education can be “good enough” to help overcome such motivational difficulties and, given good enough evaluation, clearly demonstrate an effect on crash rates.

## **INDIVIDUAL EVALUATIONS OF DRIVER EDUCATION PROGRAMS**

The great majority of driver education programs have never been formally evaluated, and most existing evaluations are severely limited in scope, power, and scientific rigor. In this section, individual evaluations of driver education programs are described, which represent either fairly recent work in the field, or older studies of special historical importance. As indicated earlier, the three basic types of studies are:

- Experimental studies—students are assigned to different training conditions;
- Quasi-experimental studies—naturally occurring groups are compared; and
- Ecological studies—assessment of changes in driver training requirements or program differences across different jurisdictions.

These categories are used to group the studies to be reviewed.

### **Experimental Studies of the Impact of Driver Education**

#### *DeKalb County, Georgia*

The U.S. DeKalb County Driver Education Project was the most comprehensive experiment in beginner driver education, based on the typical delivery of U.S. driver education in public secondary schools. The DeKalb Project is best known for its impressive efforts to provide improved training and well-controlled experimental evaluation of subsequent crashes (Lund, Williams, and Zador 1986; Ray et al. 1980; Smith 1983; Stock et al. 1983).

Volunteer high school students were assigned to one of three groups that received different driver instruction conditions. The random group assignment was intended to eliminate self-selection bias, which had troubled earlier attempts to evaluate driver education (Vernick et al. 1999). That is, in the normal course of events, beginner drivers who take driver education courses are different from those who do not, in other ways as well, and these other differences bias any attempt to compare their subsequent driving records. Random assignment to treatments in theory eliminates this bias, even if it introduces some other problems in practice.

A special educational program called the Safe Performance Curriculum (SPC) was developed for the DeKalb study. The new curriculum was based on a comprehensive driving task analysis, and was intended to represent the 1970s state-of-the-art driver education, both in terms of content and methods. The SPC, which was pilot tested and improved before implementation, was a much longer program and more carefully developed than typical driver education curricula of its time.

It was strongly oriented toward improving drivers' hazard perception skills, and it de-emphasized the motivational "scare tactics" that had been prevalent earlier.

The SPC consisted of 32 hours of classroom instruction, 16 hours of simulation instruction, 16 hours of driving range instruction, 3 hours of collision evasion instruction, and 3.3 hours of on-road, behind-the-wheel instruction, including 20 minutes at night (Lund, Williams, and Zador 1986).

The effect of SPC training on crashes was compared to two other training treatment groups. A minimal training course called PDL (Pre-Driver License) was intended to represent the minimal training required to pass the licensing test. It was essentially a standard driver education course of 30 classroom hours and 6 driving hours. The other comparison group was a "no-treatment" control group. The young people assigned to this condition were denied the opportunity to receive driver education in the public schools, leaving them to find their own instruction from family members or private instructors. The training the members of the control group actually received was apparently not known.

To overcome an initial resistance and induce those assigned to the control group to actually participate in the experiment, insurance discounts were offered equivalent to those available to the training group participants. These discounts were available to control group members if they passed a special road test that was used for the other training groups (Stock et al. 1983). The number of the supposedly untrained group who prepared themselves for this rigorous test is unknown, but such thorough preparation would presumably reduce the contrast between the two training groups and the supposed no-treatment control group. This is a clear illustration of the difficulty of applying the randomized controlled trial method to beginner driver education. Unlike the treatment assignment in an ideal RCT, subjects were able and encouraged to respond to group assignment in ways that could defeat the purpose of this type of assignment.

Intermediate learning outcomes were also assessed to some degree. Testing was carried out, and records were followed for the young volunteers in each of the three groups. SPC-trained drivers showed better on-road skills than the other groups. SPC graduates that received drivers' licenses also experienced fewer collisions per licensed driver over their first six months of licensed driving. When the data were reanalyzed, however, to examine net benefits for the entire group assigned to the SPC (whether or not they actually took the course), no benefit in collisions was found. The reduction of collisions per licensed driver seen in SPC drivers was offset by earlier licensing, and therefore more exposure to risk, for the young drivers assigned to SPC compared to the groups assigned to the minimal formal training and to no formal training.

After six months, even collisions per licensed driver were no longer better for the SPC group. In a long-term follow-up study of the records of the DeKalb students over six years, both the SPC and minimal curriculum males were found to have significantly fewer convictions, and both males and females in the minimal curriculum group had fewer crashes (6%) than the untrained controls (Smith and Blatt 1987; Weaver 1987). Over the long term, these studies indicated that minimal training led to slightly lower crash rates than did the more extensive and intensive SPC training. This finding was somewhat puzzling, as it seemed to show a delayed effect of training

(R. Peck, personal communication). A later reanalysis of the DeKalb data by Davis (1990) found no differences in the crash rates of the three training groups after the first year following training. Davis also strongly questioned the technical adequacy of the statistical methods used by some of the earlier analyses of the DeKalb data.

Mayhew and Simpson (1997) conducted a detailed review of the original DeKalb study findings and the subsequent analyses by other researchers, and concluded:

Thus despite significant effort, the DeKalb evaluation produced findings that failed to provide evidence of the beneficial effects of formal instruction. Not surprisingly, the equivocal nature of the results has led to substantial controversy that has had a profound impact on driver education/training. (Mayhew and Simpson 1997, 20)

Disappointment with the findings for the SPC graduates led to withdrawing support and subsequent stasis and decline of driver education in North America for many years (Nichols 2003). Smith (1983) viewed the issue more as one of specific training effectiveness and less as one of engineering safety on a broad societal scale. He contended that collision measures are not the appropriate criteria to assess a program that has a main objective to ensure proper and safe driving performance because collisions are not common occurrences. They are valid measures of driver performance only in conjunction with measures of exposure, and they do not reflect the full range of driving ability. He recommended adopting an intermediate criterion developed for the DeKalb project's improved curriculum. This measure was based on observed behavior in selected traffic situations. According to Smith, it measures:

... both cognitive and non-cognitive behaviors, observes actual behavior patterns in relation to real-life traffic, and records interrelationships of driver behaviors to changing traffic conditions. ... Such a criterion yields immediate results, is accumulated in a short period of time, identifies proficiencies/ deficiencies in response to real world situations without waiting for people to injure or kill themselves. (Smith 1983, 26)

The DeKalb experiment has numerous implications for the field. It suggests that even carefully planned RCTs can have difficulty in achieving and maintaining assigned treatment groups, particularly in a no-treatment control group. It raises endless possibilities for debate over proper comparisons. If measurable effects of a program exist, how long should they be expected to last before we consider them real or important? Should assignment to the target treatment group be considered the treatment, so that whole assigned groups are compared? If so, aren't the differential effects of the compared programs being diluted by those who are assigned to, but do not actually attend or complete the program?

From a broader program management perspective, the DeKalb experience also points out the limitations of the program development theories prevalent at the time. Progress in programs is rarely made by monolithic developments. More modern theory suggests incremental, continuous improvement approaches are more likely lead to success. Finally, the policy response to the disappointing results was probably not the right one—in retrospect, reducing research and program development resources did not kill driver education but rather delayed its development.

## *Australia and New Zealand*

Wynne-Jones and Hurst (1984) conducted an experimental trial evaluation of the Automobile Association (AA) high school driver education program in New Zealand, using random assignment to conditions to eliminate self-selection bias. One group (561 students) was offered a course consisting of 15 hours in the classroom and 8 hours of on-road driving. The other group (227 students) was not offered high school driver education, leaving them to family or private instruction.

For the 788 participants, a comparison of driving records after licensing was conducted using both self-report information and crash records. Overall, no statistically significant reductions in collisions or convictions were found for AA students. Females in the experimental group had significantly more collisions than those in the control group. Students who were offered high school driver education obtained their licenses earlier, in this case by substantial margins (6 months for males and 10 months for females). The authors suggested that the AA program must be seen as:

... a means to expedite the obtaining of a driver's licence with the advantage of known expert instruction. It should not be regarded as a means to guarantee training in safer driving, nor can one conclude that any habits of safe driving transmitted by this training would outlast the training period. (Wynne-Jones and Hurst 1984, 11)

A similar randomly assigned controlled trial also took place in Australia around the same time, with no significant differences being found in driver records between driver education students and a control group (Strang et al. 1982).

These experimental studies used such small numbers of students that their statistical power was very low. That is, the small number of cases made it unlikely that even a moderate difference between the crash records of the treatment and control groups would be found to be statistically significant.

## *Sweden*

In 1994, Gregersen reported a modest but elegant Swedish experimental trial that produced surprising results. The educational treatment was a specially developed cooperative program that combined home instruction for the theory component of driver education and coordinated professional in-car instruction (11 hours). Beginner drivers were assigned to either treatment or control groups from among teenagers who reported in a previous survey that they were planning to have only private driving instruction from family or friends. Group assignment was approximately random. The treatment group (about 850) was offered nearly-free professional instruction for the trial. Both groups were followed up with surveys over two years.

The educational treatment improved some measures of performance and attitude (reduced reported speed and less overconfidence). The surprising finding was the treatment group was

significantly worse in the first year in self-reported crashes per kilometer. In the second year, that group was significantly better in crashes. In looking for an explanation for the unexpected pattern of group differences, Gregersen (1994) speculated that cognitive overload might have prevented any benefit of the improved performance appearing during the first year. This suggests that training effects might be more complex than generally assumed, and that careful and prolonged follow-up is essential in a thorough evaluation. These surveys are important, in that they allow measurement of possible behavioral differences between groups, such as the amount of driving.

### *California*

Dreyer and Janke (1979) conducted an early prospective experimental evaluation in California, and found a substantial benefit in reduced crashes. Structured as a randomized controlled trial, the study compared results for about 2,000 students randomly assigned to driver education programs with and without in-car practice on a special off-road driving range (as opposed to on-road practice only). The total amount of driving time in the two programs was equivalent, but differed in where it took place. The classroom components of the two programs were the same. Unlike the DeKalb study, no attempt was made to include a no-treatment or minimal treatment control condition. Intermediate measures were taken around the time of training, but apparently no attempt was made to acquire data from the students during the follow-up year.

The students who took the assigned range and non-range programs were compared and found to be similar in a number of measures, such as licensing test scores and time to licensing. The non-range students were significantly better in a knowledge test and simulator scores, but the range students were better on a number of driver record measures over the first year of licensed driving. The range students' advantage in total crashes was large (33% lower) and the difference was statistically significant. Other record measures, such as injury crashes and violations were better for range students, but the differences did not reach statistical significance. It is not clear why the range students should have had so many fewer crashes during their first year, since they were not typically better on the limited intermediate criterion measures that were taken. There were no measures during or after the follow-up period, so differences in the amount of driving exposure and other possible differences between the groups are unknown.

A recent study by Masten and Chapman (2003; 2004), also a prospective experimental trial, addressed only intermediate student outcomes using test score results and did not include crash impacts. The study provides a refined modern experimental extension of the historical line of mainly retrospective, quasi-experimental evaluation research that attempted to find differences in outcomes of different forms of driver education and training (see Nichols 2003).

The driver education program under evaluation consisted of only a classroom component, without in-car training. The study's intent was to see whether the diverse delivery methods available and approved in California led to different educational outcomes. These outcomes were measured by

scores on specially-constructed, proctored exit tests of knowledge and attitude, as well as the state's licensing knowledge test. The attitude measure was based on the psychological concept "locus of control," which has a modest empirical basis as a correlate or predictor of safe driving (Lonerio et al. 1995). A total sample of 1,300 was divided into four groups for assigning to different training conditions. Use of criteria other than crashes permitted this study to use much smaller samples and still retain reasonable statistical power.

The sample of potential new drivers was randomly divided into treatment groups and assigned to one of four instructional settings: classroom, CD-ROM, workbook, or internet/workbook home study. All methods were intended to deliver the same standard driver education curriculum. Group differences on the special exit test and attitude scale were small, but the differences significantly favored the CD-ROM and internet/workbook groups. These methods also had substantially and significantly higher rates of course completion. The Department of Motor Vehicle (DMV) knowledge test scores were higher for the classroom group, but this may have been because classroom instructors taught directly to the test items, apparently a common practice.

The Masten and Chapman study is essentially a product evaluation that is narrowly focused on training methods measured against specific intermediate measures of knowledge and attitudes. It is a well-planned and controlled design, however, using reasonably well-tested psychometric measures of intermediate criteria, which in turn have some plausible, although not proven, relation to safe driving.

The study provides useful guidance in developing alternative high-tech delivery of, primarily, the basic knowledge component of beginner driver education. It does not, of course, give direct information about the safety implications of the alternative training delivery methods or of the content delivered. It does, however, raise interesting questions of how such findings should be applied to program development and how they should be followed up with further research. Should the field move toward the more effective methods suggested? Or should those methods be researched further to replicate the findings and further refine the effectiveness of the methods?

## **Quasi-Experimental Evaluation Studies**

### *United Kingdom*

An example of a typical quasi-experimental approach is found in a U.K. follow-up survey after licensing. In this approach, the relationship between learning to drive and subsequent accidents was measured by a longitudinal three-year survey study of a cohort of newly licensed drivers (Forsyth, Maycock, and Sexton 1995; Maycock 1995; Maycock and Forsyth 1997). About 15,000 new drivers were surveyed by mail questionnaire three times, after one, two, and three years of driving experience. About half returned the first questionnaire with moderate declines in response after the second and third year. Results of this study highlighted differences between

males and females. The length of time spent learning to drive, for example, was not related to self-reported accidents for females. For males, however, longer times were associated with fewer subsequent crashes. On average, the U.K. new drivers took about 14 months to learn.

Instruction in this study was limited to in-car lessons. Nearly all respondents had received some professional instruction (males 97%, females 99%). Surprisingly, more instruction was associated with more crashes. In females, where the effect was more clearly significant, the effect seemed to result from a small number of licensing candidates who: 1) required much instruction before taking the state driving test; 2) were less likely to pass the first time; and 3) crashed more after eventually passing.

Interestingly, only 21% of men and 30% of women reported reading the government driving manual, but this reading was also not significantly related to subsequent crashes. Clearly, this naturalistic type of survey research is limited in its ability to establish causal relations between instruction and crashes due to the self-selected subject population, extraneous, and possibly confounding differences between the groups other than the training received.

### *Pennsylvania*

A similar approach has also been conducted in Pennsylvania (McKenna et al. 2000). A random sample of 1,188 16- to 18-year-old drivers was selected to be part of a telephone survey that asked respondents to provide information about their driving records and personal demographics. The subjects consisted of high school driver education students (57%), commercial driver education students (13%), and those who reported no formal driver education (34%). Unlike the early quasi-experiments, studies such as this one gathered additional information about the characteristics of the drivers in the driver education and comparison groups, permitting use of multi-variate statistical techniques to partially compensate for the lack of random assignment to the groups.

In the Pennsylvania study, logistic regression was used to simultaneously assess the effects of 16 variables on the outcome of crash or no crash. Crash rates were lower for students with high grades and higher for those who made car payments. No evidence existed of fewer crashes, fewer convictions, increased seat belt use, or lower crash severity for the driver education group.

### *Manitoba*

Manitoba's driver education situation is unusual compared to most other North American jurisdictions. It has a centrally organized high school driver education program supported by Manitoba Public Insurance (MPI), the province's public automobile insurer. About 70% of new drivers (~14,000) take the course each year.

In 2000-2001, Lonero et al. (2005) conducted a quasi-experimental survey and record study to assess the impacts of the program on students and identify areas for improvement. Drivers'

insurance records were used. Driver licensing record data were not available so that some details of licensing and collision information could not be included. Samples consisted of 4,000 each of program graduates and those who learned to drive some other way, as well as 2,000 each of parents and the general public. Multivariate statistical methods were used to analyze survey and record data to partially control for pre-existing differences between the driver groups. Mail surveys were conducted, and questionnaire completions ranged from 26% for the general public to 39% for graduates, considered moderate response rates. The validity of any survey is reduced by the loss of subjects—those who reply probably differ from those who do not in other ways as well. A brief follow-up survey provided data on non-respondents.

When the age difference between the groups was accounted for, no significant difference in self-reported crash rates existed between the groups, regardless of whether the comparison group drivers had any formal training. Nearly 20% of young drivers in both groups indicated they had at least one crash not reported to authorities.

Graduates appeared more health conscious and confident in their ability to avoid crashes, and they reported better avoidance of errors and violations. Their greater reported seat belt use showed the strongest difference of any of the self-report items. In recorded convictions, statistically significant differences existed between the groups, but the findings were contradictory. When the whole populations were modeled, the graduates had more convictions in a given year, but when the smaller group of survey respondents was modeled, the comparison group had more violations.

Statistical modeling showed the effects of age, months of experience, and gender were all statistically significant factors in determining crash rate. When all of these other variables were controlled for, there was no statistically significant difference between groups in terms of overall crash incidence.

### *Australia*

A recent Australian study used a quasi-experimental survey follow-up of drivers receiving different forms of “pre-driver” education in the state of Victoria (Haworth, Kowadlo, and Tingvall 2000). The researchers attempted to compare self-report crash experience of drivers who had taken a pre-driver course including an off-road in-car component, with drivers who had not taken such a course. The comparison group had either taken an in-class-only course or none at all. Multi-variate statistical methods controlled for some of the extraneous differences between the groups (age, sex, and residence location), resulting in an adjusted crash difference of 20% in favor of the drivers with the in-car course.

Because of the small number of subjects (total number was 687), however, the difference between the groups was not statistically significant. This seemingly predictable outcome results from a design flaw and is unfortunate in an otherwise rather clever study design. It re-emphasizes the need for careful sample size calculation and power analysis when planning an evaluation study. This study could best be seen as a pilot project.

## *British Columbia*

An ongoing program of evaluation research in British Columbia (Wiggins 2005) is directed primarily to the province's Graduated Licensing Program (GLP) but also addresses the effects of driver education in the context of the time discount in GLP. In a record study that adjusted for age and gender, new drivers who used a driver education certificate to shorten their learner license period in GLP crashed 26% more over their first year of unsupervised driving than those who did not present a certificate. When adjusted for time spent in the learner period, the difference dropped to 13%. A case-control survey study identified how the new drivers learned to drive. Regardless of whether they used a certificate at licensing, those who took an approved course had about 26% more crashes, adjusted for age, gender, and frequency of driving during the first six months of unsupervised driving.

Wiggins suggests other reasons to believe that driver education in British Columbia may not yet operate at a level consistent with the new GLP standard, but she also points to similar findings appearing in the graduated licensing evaluations in Ontario and Nova Scotia. Factors accounting for the excess risk of driver education graduates in Canadian graduated licensing systems are unclear.

## *Ontario*

Zhao et al. (2006) surveyed 1,533 students in numerous Ontario high schools about their driving behavior and related factors, such as class of graduated license held, amount of driving, and crashes. Crash experience was compared for those who had or had not taken driver education, with a number of other factors accounted for by multi-variate statistical models. In this it resembles the Manitoba longitudinal study, although in the Ontario study, self-reported crashes were not supplemented with insurance or licensing records.

Results showed that, among drivers who held Ontario's first stage learner's license (G1), significantly fewer driver education graduates reported having crashes than those who had not taken driver education. In fact, among the G1 drivers, driver education was the only factor significantly associated with crashes. In contrast, among drivers with intermediate (G2) licenses, those with driver education had greater odds of reporting a crash, although the difference was not statistically significant. The findings suggest further study to identify reasons for the effect among the highly restricted learner drivers, while none is found for those at the intermediate licensing stage, in which independent driving is permitted with some restrictions.

## **Ecological Evaluation Studies**

The broader scale and less direct ecological approach is the principal alternative to experimental trials and quasi-experiments for evaluating the impact of driver education. In ecological studies, changes or differences in large-scale factors, such as laws or policies, are typically evaluated over time in one jurisdiction or between different jurisdictions.

## *Connecticut*

As a result of eliminating state subsidies for driver education in Connecticut high schools, nine school boards decided to drop the courses, while other communities continued to offer them. Robertson (1980) used this opportunity to investigate the relation between driver education and the age of first licensure and collision rates.

Obtaining a driver's license became more difficult and expensive in areas that dropped driver education courses. Robertson reported that the number of license-years of 16- and 17-year-old driver education graduates declined by 57% in the affected communities, compared with 9% in communities where driver education was retained. The affected communities showed a 63% decrease in the collisions of 16- and 17-year-old driver education graduates, while no change was present in the other communities.

With driver education no longer available in the affected communities, declines in licensing and crashes among driver education graduates are not surprising. The total licensing and crashes of all 16- and 17-year-olds in the affected communities declined much less (10-15%). Most young people apparently found other ways to learn to drive. From the data presented, there appear to be slightly fewer crashes per population of 16- and 17-year-olds in favor of the communities that dropped driver education, but only in the second follow-up year. Neither the actual rates for all 16- and 17-year-old drivers, nor any statistical inference regarding the significance of differences, however, are presented. The study results, therefore, do not strongly support Robertson's conclusions that eliminating driver education improved young driver crash experience significantly by delaying licensure.

Unfortunately, this severely flawed report is still often referenced in support of the suggestion that driver education has significant perverse effects on safety as a result of inducing earlier licensing. Based on their review of DeKalb and other results, Mayhew and Simpson (1995) concluded that students who take high school driver education are licensed earlier by about one month compared to students who would have taken the course had it been available to them. Also based on DeKalb data, however, they concluded that earlier-licensed driver education students drive less in a given period after licensing, at least partially offsetting the exposure increase that might result from earlier licensure.

## *Quebec*

Potvin, Champagne, and Laberge-Nadeau (1988) used a time-series design to evaluate the impact of introducing a mandatory driver-training requirement for all new drivers in Quebec. Previously, only new drivers under age 18 were required to be trained. The main effect of the mandatory training requirement was an increase in the total number of crashes, as more 16- and 17-year-old females became licensed, without any effect on crashes per licensed driver.

Prior to requiring formal training for all new drivers, there may have been a tendency for drivers to avoid the cost of driver training by waiting until the age of 18 to become licensed. The authors theorized that the increase in early licensure occurred because there was no longer any economic advantage to waiting until age 18 to be licensed. The effect was stronger in females, because it was mainly females who had waited until after age 18 to become licensed.

### *U.S. Fatal Crash Modeling Studies*

Additional ecological studies in the United States have also failed to find strong beneficial effects of driver education requirements, as concluded in the 1999 review by Vernick and colleagues. Levy (1990), however, conducted a 47-state econometric modeling study of the relationship of various safety factors to fatal crashes of teenaged drivers. He concluded that a mandatory driver education requirement had a small but significant association with fewer fatalities in 15-17-year-olds. An earlier modeling study by Robertson and Zador (1978) had failed to find a significant effect of the proportion of driver education graduates on fatalities per 10,000 licensed drivers.

These two studies are unusual in addressing fatal crashes, as nearly all other crash-based evaluations looked at new drivers' total crashes. Fatal crashes are so rare that only these very broad modeling studies, covering large populations of drivers, are able to use them as criteria. Preventing fatal and serious injuries is the main concern of road safety in general, and of young drivers' safety, but nearly all research and evaluation necessarily addresses total crashes. These have been used almost exclusively because of easier availability of large enough numbers of crashes to make moderate differences detectable among the groups being studied. Total crashes consist mainly of minor, property-damage-only crashes, which would not matter if total crashes were a good substitute indicator measure for serious crashes. Theoretical and empirical bases exist, however, for thinking that serious crashes are different from routine minor crashes.

On generalizing between fatal crashes and all crashes, Robertson and Zador wrote:

Since the characteristics of motor vehicle crashes involving serious injuries are generally similar to those of fatal crashes, it seems likely that these crashes are similarly affected. Run-of-the-mill crashes involving only property damage or minor injuries tend to have quite different characteristics, however, and it is not safe to assume that the conclusions apply to such crashes. (Robertson and Zador 1978, 965)

While any crash study requires large sample sizes, fatal or serious injury crashes can only be considered when huge, national-scale data sets are available. Greater use of large-scale modeling techniques has been recommended as an important direction for safety R&D and policy support

(Lonerio et al. 2005). These approaches are complex and dependent on statistical correction for many factors, but studies like Levy's approach using national data are worth pursuing further to see whether the apparent positive benefit of mandatory driver education provisions in the 1980s can be verified, replicated, or extended in the modern environment.

### *Denmark*

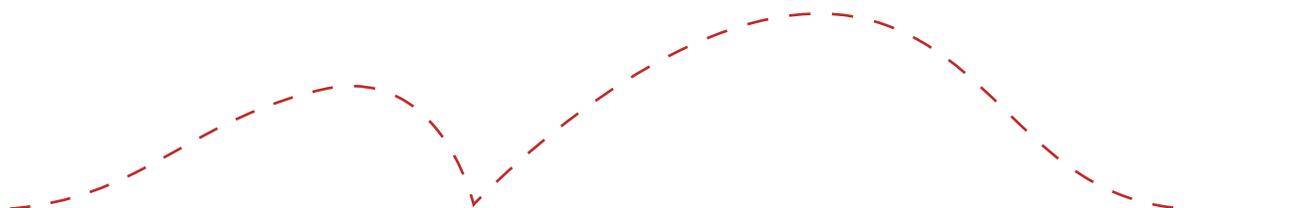
Carstensen's (2002) ecological study examined the effect of a new mandatory driver education program in Denmark consisting of classroom and in-car training that differed from traditional courses by more strongly emphasizing defensive driving and including motorway driving and night driving. Records of 18- and 19-year-old drivers (the youngest eligible to be licensed) were followed before and after the new training requirement and compared to older drivers' records. While both age groups' crash rates declined after 1986, the young drivers' crash records declined significantly more.

Driver education can affect subsequent driving exposure, and driving exposure is closely related to crash experience. Unlike many other studies, Carstensen checked for differential changes in driving exposure. Such changes were ruled out as explanations of the new program's greater crash reduction among new drivers.

Other factors in addition to the driver education requirement, however, could have contributed to a decline in young driver crashes. The new training requirement, for example, made becoming licensed more expensive, which may have prevented or delayed licensing for some potential new drivers, perhaps resulting in a higher percentage of older, and lower-risk, new drivers. Specific data on driver age and crashes per driver, however, were not reported. This ecological type of study cannot control for the many potential external forces that could influence the results (Mayhew and Simpson 1997). Adding periodic surveys to supplement record data would add considerable strength to this type of evaluation.

The Danish experience and the later introduction of second-stage training requirements elsewhere in Scandinavia (Keskinen, Hatakka, and Katila 1998) point to one potentially fruitful direction for further investigation: multi-stage instruction.

The following table provides a summary of findings of each of the individual studies discussed previously.



## Summary of Driver Education Evaluation Results

Reference	Design	Results	Methodological Strengths/Limitations
<b>Experimental Studies</b>			
Dreyer and Janke 1979 <i>California</i>	<ul style="list-style-type: none"> <li>2,057 students randomly assigned to two training conditions (off-road range training vs. on-road)</li> </ul>	<ul style="list-style-type: none"> <li>Those receiving range practice had fewer recorded crashes, but tests scores were no different</li> </ul>	<ul style="list-style-type: none"> <li>Randomized controlled trial</li> <li>Intermediate measures</li> <li>No follow-up survey for exposure and behavioral measures</li> </ul>
Ray et al. 1980 Stock et al. 1983 <i>DeKalb County, Georgia</i>	<ul style="list-style-type: none"> <li>Intensive, minimal, and no driver education groups</li> <li>About 6,000 students randomly assigned to each group</li> </ul>	<ul style="list-style-type: none"> <li>Intensive training (SPC) drivers had better skills and fewer crashes during first 6 months of driving, but not beyond</li> <li>Effects were complex (see text for explanation)</li> </ul>	<ul style="list-style-type: none"> <li>Comprehensive randomized controlled trial</li> <li>Long follow-up—6 years</li> <li>Formative evaluations and intermediate outcome measures</li> </ul>
Wynne-Jones and Hurst 1984 <i>New Zealand</i>	<ul style="list-style-type: none"> <li>788 students, 561 received course, 227 family or friend taught</li> <li>Random assignment</li> </ul>	<ul style="list-style-type: none"> <li>No reduction in collisions for driver education group</li> </ul>	<ul style="list-style-type: none"> <li>Adequate design</li> <li>Small control group</li> <li>No formative evaluation or intermediate outcome measures</li> </ul>
Gregersen 1994 <i>Sweden</i>	<ul style="list-style-type: none"> <li>850 students received driver education course compared to controls</li> <li>Random assignment</li> </ul>	<ul style="list-style-type: none"> <li>Driver education group crashes significantly worse first year, significantly better in second year</li> </ul>	<ul style="list-style-type: none"> <li>Longer follow-up—2 years</li> <li>Reasonable sample size</li> </ul>
Masten and Chapman 2003; 2004 <i>California</i>	<ul style="list-style-type: none"> <li>1,300 students randomly assigned to one of four instructional settings</li> </ul>	<ul style="list-style-type: none"> <li>Home-based methods better for 1 knowledge and attitude test, classroom better for DMV knowledge test</li> </ul>	<ul style="list-style-type: none"> <li>Sample size adequate</li> <li>Well planned and controlled</li> <li>Intermediate (psychometric) measures only</li> </ul>

Reference	Design	Results	Methodological Strengths/Limitations
<b><i>Quasi-Experimental Follow-Up Studies</i></b>			
Forsyth et al. 1995 <i>United Kingdom</i>	<ul style="list-style-type: none"> <li>Survey of 15,000 new drivers</li> </ul>	<ul style="list-style-type: none"> <li>Longer time learning to drive associated with fewer crashes for males</li> <li>More driving instruction was associated with more crashes</li> </ul>	<ul style="list-style-type: none"> <li>Several follow-ups over time</li> <li>Self-selection bias</li> <li>Self-reported data only</li> </ul>
Howarth et al. 2000 <i>Australia</i>	<ul style="list-style-type: none"> <li>Self-report crash effects for in-car training effects</li> </ul>	<ul style="list-style-type: none"> <li>Substantial crash differences in favor of in-car training condition, but not statistically significant</li> </ul>	<ul style="list-style-type: none"> <li>Sample size too small</li> </ul>
McKenna et al. 2000 <i>Pennsylvania</i>	<ul style="list-style-type: none"> <li>Survey and crash records</li> <li>Random sampling for survey</li> </ul>	<ul style="list-style-type: none"> <li>Driver education not associated with lower crashes or convictions</li> </ul>	<ul style="list-style-type: none"> <li>Multi-variate statistical analysis used to control for confounding variables</li> <li>SES missing from control variables</li> </ul>
Lonero et al. 2005 <i>Manitoba</i>	<ul style="list-style-type: none"> <li>Survey and crash records</li> <li>Random sampling for survey</li> </ul>	<ul style="list-style-type: none"> <li>Driver education not associated with lower crashes or convictions</li> </ul>	<ul style="list-style-type: none"> <li>Multi-variate statistical analysis used to control for confounding variables</li> </ul>
Wiggins 2005 <i>British Columbia</i>	<ul style="list-style-type: none"> <li>Cohort record study</li> <li>Case control study with survey and records</li> </ul>	<ul style="list-style-type: none"> <li>New graduated license holders who took driver education had 26% more crashes</li> </ul>	<ul style="list-style-type: none"> <li>Multi-variate statistical analysis used to control for confounding variables</li> </ul>
Zhao et al. 2006 <i>Ontario</i>	<ul style="list-style-type: none"> <li>Self-report survey of high school students</li> </ul>	<ul style="list-style-type: none"> <li>Driver education associated with fewer crashes for learner license holders</li> </ul>	<ul style="list-style-type: none"> <li>Multi-variate statistical analysis used to control for confounding variables</li> </ul>

Reference	Design	Results	Methodological Strengths/Limitations
<b>Ecological Studies</b>			
Robertson and Zador 1978 <i>27 States USA</i>	<ul style="list-style-type: none"> <li>Modeling study of driver education and fatal crash rates</li> </ul>	<ul style="list-style-type: none"> <li>No relation between proportion taking driver education and fatality rates</li> </ul>	<ul style="list-style-type: none"> <li>Not program specific</li> </ul>
Robertson 1980 <i>Connecticut</i>	<ul style="list-style-type: none"> <li>School boards with and without driver education</li> </ul>	<ul style="list-style-type: none"> <li>For school boards without driver education, total licensing and crashes of 16- and 17-year-olds decreased by 10-15%</li> </ul>	<ul style="list-style-type: none"> <li>Not enough data analysis presented</li> </ul>
Potvin et al. 1988 <i>Quebec</i>	<ul style="list-style-type: none"> <li>Mandatory driver education introduced in Quebec for all (formerly just 16-17-year olds)</li> </ul>	<ul style="list-style-type: none"> <li>Increased number of young driver crashes due to increased number of licensed females aged 16-17</li> </ul>	<ul style="list-style-type: none"> <li>Large sample size</li> <li>Different timeframes for treatment and control groups</li> </ul>
Levy 1988; 1990 <i>47 States USA</i>	<ul style="list-style-type: none"> <li>Large-scale modeling study of effects of mandatory driver education</li> </ul>	<ul style="list-style-type: none"> <li>Small but significant beneficial effect on fatal crashes</li> </ul>	<ul style="list-style-type: none"> <li>Not program specific</li> </ul>
Carstensen 1994; 2002 <i>Denmark</i>	<ul style="list-style-type: none"> <li>Mandatory driver education, new curriculum</li> </ul>	<ul style="list-style-type: none"> <li>Reduced crashes</li> </ul>	<ul style="list-style-type: none"> <li>Large sample size</li> <li>No control of confounding variables</li> </ul>

## IMPLICATIONS FOR DRIVER EDUCATION EVALUATION

A major concern for all kinds of evaluation is the criterion of effectiveness. What exactly do we expect the program to achieve? The traditional criterion for effectiveness of driver education has been its ultimate goal—to reduce crashes among graduates relative to those who learned to drive in other ways.

When considering safety effects, the specific measure of crash experience is important and can be controversial. Different ways of measuring crashes can provide quite different results. If crashes are measured through self-report, government records, or insurance records, differences arise as a result of variations in completeness of reporting and different timing for records to work their way through bureaucratic systems.

Crash experience is most clearly reported in terms of rates, and the denominator chosen for the rate has important implications. As seen in the various DeKalb analyses, crashes per licensed

driver can give quite different results than crashes per assigned subject. Crashes per distance driven can give different results again.

The rate used needs to match the goals of the program and the evaluation. Is the success criterion for driver education to be safer mobility or the broader public health goal of a safer youth population? Preference for mobility-based rates, such as crashes per mile traveled versus preference for population-based measures, such as crashes per age-group population, reflect fundamental theoretical differences and need to be resolved early in evaluation planning.

Ultimate safety measures are the most important success indicators, but they are not the only important educational objective for driver education, or any form of safety education. This is particularly clear where the safety education is sequenced and coordinated with other influences, which is a condition now thought to be critical to success. Donelson and Mayhew's (1987) extensive review of driver improvement programs emphasized that concentrating on intermediate outcomes was critical to more effective driver improvement interventions. These, like driver education, are usually intended to have educationally driven behavior change effects on crash propensity. In either case, focusing only on the safety effects makes it unlikely that programs can be improved enough to actually achieve the desired safety effects.

In the mid-1980s, the Road Transport Research Programme of the Organisation of Economic Co-operation and Development (OECD) assessed the efficacy of road safety education programs and provided strategies for program development, implementation, and evaluation (OECD 1986). This report outlined issues of effectiveness for program planning and implementation, but its primary focus was evaluation. It was suggested that a program should be seen as effective if it does what it was intended to do, and that it is, therefore, very important to be explicit about educational objectives, which should include measures of intermediate effects and not be solely focused on collisions.

Clear objectives addressing drivers' task requirements are crucial to program effectiveness, but they are problematical because of the lack of empirical knowledge available on many driver tasks. Intermediate objectives, such as better knowledge, skills, and attitudes are currently difficult because of lack of research to link them to collisions, and research should be undertaken to establish these linkages and validate such measures.

The content and structure of instruction were also addressed by the OECD report. Internal consistency between content and objectives is critical, but little was then known about the relationship between structures of instruction (where, how much, how often, etc.) and effectiveness. With respect to the process of instruction, three sets of critical variables were identified: learner variables, instructional variables, and social variables. Attitudes and a host of other learner factors must be considered when designing an effective program. Instructional variables are somewhat easier to clarify. Instructors must be skilled and motivated, and the report suggested that teachers are usually not well qualified to teach safety education on either of these criteria. Recent research has started to meet this need.

Critical social variables included socioeconomic and cultural differences, and language. The OECD report suggested that acceptance of driver education and its importance by teachers and other potential delivery agents is so difficult in many cases, particularly in secondary schools, that it would be better to train special staff to approach students directly through broadcast media and closed TV networks. While this may have seemed a gloomy forecast for driver education at the time it was made, recent trends toward computer- and web-based instruction could be seen as validating the OECD prognostication of driver education going directly to the student.

The OECD committee saw evaluation as the key to successful safety education. The report identified three evaluation levels. To structure the OECD model, the committee used the *formative and summative* evaluation concepts originally suggested by Michael Scriven (Stufflebeam 2001).

In the OECD model, two types of formative level evaluation were identified: 1) process evaluation—how a program is used and received; and 2) product evaluation—impacts on skills, knowledge, attitudes, or behavior. The third type was summative (or outcome) evaluation, featuring two kinds of measures—cost/benefit, and the driver education program fitting with the education system at large.

Evaluation is especially important in education programs, where the interventions seem as if they ought to be effective and seem so obviously desirable. Aside from lack of positive effects, the possibility of negative effects was recognized by the OECD, which pointed out that some kinds of advanced skills training can make some drivers less safe (OECD Scientific Expert Group 1990). This makes both formative and summative evaluation especially critical for education programs. Any potentially effective behavioral technology may be ineffective, or even harmful, depending on how it is applied.

A related weakness inherent in quasi-experimental studies involves attempts to make clean, unbiased comparisons using multi-variate statistical methods to partially control for extraneous factors that might bias the comparison. Identifying and controlling all likely biases in the characteristics of non-randomly assigned groups are difficult, and unlikely to be perfect in any one study. The evaluation of Pennsylvania's driver education program (McKenna et al. 2000), for example, identified 16 control variables but ignored socioeconomic status, typically an important factor in young drivers' risk differences.

With notable exceptions, such as the DeKalb experiment, Dreyer and Janke's 1979 experimental study, and Gregersen's 1994 survey study, most evaluations have failed to look at intermediate measures. Lacking information on what the students have or have not learned, directions for program improvement are left unclear. Most existing evaluations leave many unanswered questions regarding:

- Logical links between curricula and young drivers' needs;
- Theories explaining how a program is expected to achieve safer driving;
- The quality, comprehensibility, and usability of curriculum products;

- How well and how consistently the instructional processes actually deliver the intended learning; and
- Which learning and behavioral outcomes result, or fail to result, from the training.

Tracking of learning outcomes is an area where programs could easily build in ongoing evaluation. Knowledge and attitude measures at the beginning and end of the course and at later intervals would help keep contact with graduates and provide feedback for continuous improvement of curriculum and delivery.

The Dreyer and Janke (1979) study shows that intermediate measures may not be enough to help explain crash results. In that case, the few differences found in intermediate measures seemed to favor the group that subsequently had more crashes. This study did not include surveying the new drivers during the follow-up period, however, so possible differences in driving behavior or in amount of driving were not measured. Such information could help explain how the apparently effective training program actually had its effects.

Further, as reflected in Dreyer and Janke's study, exposure differences are too often ignored in driver education evaluation. In general, simple differences in the amount of driving (exposure to risk) account for much of the differences in crash rates between groups of drivers (Chipman 1982). Apart from the simple amount of driving, qualitative exposure differences, such as time of day, presence of passengers, geographic areas, and different trip purposes also represent different levels of collision risk, especially for young drivers (Preusser 1996; Preusser, Ferguson, and Williams 1997). Since different methods of learning to drive may be related to when the beginner chooses to be first licensed and subsequently exposed to risk, exposure information is important in any attempt to evaluate driver education programs. Otherwise exposure can be an important source of confounding or bias in the evaluation results.

The relative scarcity of evaluation work in driver education following the OECD report is unfortunate. As indicated earlier, a serious lack of statistical and research design sophistication was evident in the early studies. The later studies used progressively better designs and statistical methods. No perfect study exists. As discussed in the following section, even the random controlled trials have had weakness in terms of maintaining clean assignment of students to training conditions and clear comparisons between assigned groups.

## **Randomized Controlled Trials: The Comparison Problem**

Some of the liveliest discussions during the development of the Guidelines involved experimental designs in evaluation—particularly RCTs. In this powerful research method, people are randomly assigned to either a treatment group (the members of which receive some specific treatment), or to a control group (the members of which do not receive treatment). This allows researchers to conclude with a reasonable and calculable level of confidence that any between-group differences in outcomes are actually caused by the different treatments received. Since the groups are assigned randomly, other differences among the people should be distributed evenly between the two groups.

In this way, RCTs reduce the effects of confounding factors, such as differences between the groups, aside from the actual treatments received, that might cause different results. These could include, for example, differences between people who choose to take driver education and those who do not. In non-experimental designs, one cannot be sure that such confounding differences between the groups have not influenced the results. Even with random assignment, differences may still arise by chance, but statistical methods can calculate a good estimate of the probability that the observed differences could have resulted from chance.

Because of the “clean” comparison the RCT permits, it is considered the “gold standard” for establishing causal relationships in the experimental sciences. Indeed the RCT experimental paradigm is held in such high esteem that non-RCT studies are sometimes regarded as non-scientific and are disregarded. In Roberts and Kwan’s (2004) systematic review of driver education, for example, only RCT studies were considered. On the other hand, education evaluators are less favorable to experimental methods and RCTs.

What do these different views for the driver education evaluation imply? Can we simply accept that RCTs are more scientific than all other methods? Since experimental methods are hardly used at all in some sciences, such as astrophysics, economics, and epidemiology, this simple scientific/non-scientific distinction seems unsupportable. Although RCTs are a good way to structure many kinds of research, they are not the only scientific way.

In evaluating beginner driver education, basic practical problems surface with RCTs. To see this clearly, we need to look in detail at RCTs applied to beginner driver education. Numerous evaluations in the field have been RCTs, including the benchmark DeKalb study (Stock et al. 1982) and the contemporary California study (Dreyer and Janke 1979). Both of these evaluations randomly assigned subjects who wanted to take driver education to receive some form of treatment. The DeKalb study assigned subjects either to one of two groups receiving different high school driver education programs or to a third “no-treatment” control group, which was not supposed to receive formal driver education. The California RCT assigned subjects to two different training formats (training on a closed-course driving range vs. on-road training). A no-treatment control group was not included.

In beginner driver education evaluation, the RCT paradigm runs into difficulty when choosing a comparison condition against which to assess the results of the target program. Indeed, thoughtful researchers have suggested that, for evaluation of beginner driver education, the most suitable comparison is with informal driver training by parents. Comparing driver education to no training at all isn’t possible, of course, since all new drivers must somehow learn the basic driving skills.

Assigning students who apply for driver education to even an informal training control group is difficult in practical terms. As DeKalb showed, effectively denying formal driver education to students who want it is not easy, at least in part because it may mean student has to forego

an insurance premium discount. To work around the problem in DeKalb, the usual insurance discount was offered to the informal training control students if they passed a special road test. This probably compromised the control group, but it is not known how the group prepared for taking the special road test or how many prepared and took the test.

The difficulty in creating and maintaining a clean control group comparison seems to be an inevitable problem with beginner driver education, and it should not be seen as an avoidable error in the DeKalb study. If comparison with no formal training is viewed as absolutely necessary, then the closest approximation would be complex quasi-experimental designs, which try to statistically compensate for confounding differences between the groups. This conundrum seems unique to evaluating beginner driver education. Other forms of driver instruction, such as driver improvement programs, can be withheld from a randomly assigned no-treatment group.

Do alternative approaches to RCTs exist for beginner driver education evaluation that get around the comparison problem? The California RCT study simply compared two different forms of driver education, without an informal training control group (Dreyer and Janke 1979). Assuming the two programs are not so different that they provoke rebellion among those assigned to one of them, this can provide an opportunity for clean random assignment and maintenance of equivalent groups. Unfortunately, most reviewers have ignored Dreyer and Janke's study. It is not clear why, but it may be in part because of the lack of a no-treatment control group (Peck, personal communication). This leads to a potential problem of interpretation—uncertainty of whether or not the “better” treatment had a real positive impact. It is possible that the apparently better treatment had no effect, whereas the “worse” treatment yielded a negative effect, or even that the “better” treatment was actually the less detrimental of two treatments that both had negative safety impacts. Although these possibilities may seem remote, they cannot be dismissed, as some driver education evaluations have found negative impacts on some measures.

A study in Sweden found another way around the comparison problem. It selected its comparison groups from young people who were not planning to take formal driver education (Gregersen 1994). Some students were assigned to a formal training condition and induced to take the training offered. This could result in a clean comparison with informal training, since the comparison group will probably not seek training. This approach, however, leads to a problem of generalizability. Some might argue that the results of such an evaluation do not necessarily apply to most new drivers whose inclination is to take driver education.

A key implication of these difficulties is that there is no perfect method for evaluation of beginner driver education. RCTs can and should be used, but they should be part of a broader and more systematic evaluation approach, which includes other study designs, a wide range of output and outcome measures, and a variety of carefully planned comparisons. Perhaps it is time the “gold standard” title is passed along to comprehensive, systematic evaluation that includes various approaches to developing the whole picture of program effects and improving programs.

The final implication of past evaluations of driver education is how little they have contributed to developing and improving the programs. Earlier thinking, including that behind the DeKalb project, seems to reflect the idea that a single massive development effort could achieve driver education's safety goals in a one step. More recently, however, researchers and theorists have emphasized the importance of incrementally building knowledge gains and other systematic intermediate effects (Lonerio et al. 1994; Woolley 2000), as well as continuously developing and improving programs (Keskinen, Hatakka, and Katila 1998). Keskinen and colleagues wrote: "We have decided ... that the development of driver education will take place in short steps, with constant evaluations of the results and trying to avoid solutions which are thought to be final" (382).

## **SUMMARY AND CONCLUSIONS**

Issues of curriculum content and structure, standards and governance, instructor qualifications, and market incentives are foremost in the minds of those who develop, deliver, or regulate driver education. Nevertheless, however useful these developments may be, evaluation still seems to be more the exception than the rule in driver education, and whether the new developments are more effective than traditional approaches is not clear. Progress toward driver education of a type that clearly improves safety seems dependent on revising this situation.

Driver education is intended to be an effective behavior change agent leading to a significant reduction of traffic crashes in novice drivers. While fatal and other serious crashes are the main concerns of road safety, driver education has rarely been evaluated except on total crash rates, which consist mostly of minor crashes. Crashes are rare events with complex causation. Fatal and serious injury crashes are much rarer, and they may have different patterns of causation than minor crashes, which mainly result from simple, inadvertent errors and may be harder for novice drivers to control.

Reviewers of the evaluation literature have typically concluded that beginner driver education has yet to demonstrate clear success in improving the safety of new drivers who receive it, or of youth on a population basis. A few studies have shown such positive effects. Driver education could be seen as facilitating greater mobility at little or no additional cost in crashes and injuries, but even this is not yet clearly established.

If driver education is to achieve success in its demanding mission, it needs to be more firmly based on the body of research and theory concerning young driver skills, behavior, motivation, and risk. Program evaluation is critical for more effective program development in the future, but its limitations in the past must be recognized and corrected for a beneficial effect. Ongoing evaluation is critical for changes in the ultimate criterion of success—crash rates. A more comprehensive approach to evaluation is also needed to address theory, products, processes, and program management.

Evaluating effects on intermediate criteria is also required. These should include changes in behavior, knowledge, attitudes, and exposure to risk. Intermediate measures should continue

during the follow-up period if we are to have a clear picture of the effects of the program and the reasons for them. Meeting the ultimate goal of reducing novice drivers' serious crashes will also likely require evaluating and managing the context of driver education. This would involve assessing the linkage of driver education with parental and community influences, graduated licensing, and other behavioral influences such as incentives and disincentives.

The now aging DeKalb study has been considered to be the most extensive and rigorous driver education evaluation, but even this study had serious limitations, and its conclusions are still controversial. Most other evaluation studies were more limited in scope and scale. Experimental evaluations typically have found no statistically significant effects of driver education on crash records, but one California study did. Several quasi-experimental and ecological studies have been conducted. Two large-scale ecological evaluations showed positive effects of driver education, but one early study did not. No one study design is perfect, and progress will likely develop on a "weight-of-evidence" basis over numerous studies of different types.

A main concern with driver education in schools has been that it leads to earlier licensing and may lead to increased driving exposure. Earlier licensing of new drivers may occur for numerous reasons, some of which are related to driver education, for example, when parents consider their children well trained and, therefore, safe to drive. The effect of high school driver education on earlier licensing seems to be fairly consistent but modest in the amount of time of earlier licensing.

Aside from earlier licensing leading to more exposure, better-trained students may become overconfident, and this may offset the potential benefits of their superior skill and knowledge. Later research reinforces the idea that driver education can lead to overconfidence. These considerations lead to questions of suitable goals and expectations for driver education—whether it is supposed to support safe mobility, to enhance the public health of youth, or both. Such considerations have important implications for driver education evaluation and program development, requiring a much more intensive and comprehensive evaluation approach than has been typical in the past.

Thus, more systematic and comprehensive evaluation is essential to the rapidly evolving future of driver education. Past studies have demonstrated that assumptions about what is effective in reducing young driver crash risk are not always well founded. Evaluation of both existing and new programs, therefore, is critical. While evaluation is important to improving the effectiveness and efficiency of driver education, it is also important to recognize the limitations of evaluation. This recognition has been lacking in the past and has led to unfortunate policy decisions, particularly in the cutting of R&D and other support. Driver education evaluation, like driver education itself, is evolving and still far from its ultimate conclusion.

## APPENDIX B: Glossary of Terms

The following glossary of terms is a compilation of definitions from several evaluation sources, listed on pages 228-29.

### A

**Analysis:** The process of systematically applying statistical techniques or logic to interpret, compare, categorize, and summarize data collected to draw conclusions.

**Assumptions:** Hypotheses about conditions necessary to ensure that: (1) planned activities will produce expected results; and (2) the cause-effect relationship between different levels of program results will occur as expected. Achieving results depends on whether or not the assumptions made prove to be true. Incorrect assumptions at any stage of the results chain can become an obstacle to achieving the expected results.

**Attribution:** Causal link of one event with another. The extent to which observed effects can be ascribed to a specific intervention.

**Attrition (Bias):** Loss of subjects from a defined sample during a study.

**Auditing:** An independent, objective, systematic process that assesses the adequacy of an organization's internal controls and the effectiveness of its risk management and governance processes, in order to improve efficiency and overall performance. It verifies compliance with established rules, regulations, policies, and procedures, and validates the accuracy of financial reports.

### B

**Benchmark:** Reference point or standard against which program effects can be assessed. A benchmark refers to the performance achieved in the recent past by the same or other comparable organizations, or the performance that can be reasonably inferred to have been achieved in similar circumstances. A benchmark is a referenced behavior for comparing observed performance at a given level.

**Bias:** A constant error or any systematic influence on measures, judgments, or statistical results, irrelevant to the purpose of the evaluation. Statistical bias is inaccurate representation that produces systematic error in a research finding. Bias may result in overestimating or underestimating certain characteristics of the population. It may result from incomplete information or invalid data collection methods and may be intentional or unintentional.



**Case Study:** A research method that focuses on the characteristics, circumstances, and complexity of a single case, or a small number of cases, often using multiple methods. The case is viewed as being valued in its own right, and while the findings can raise awareness of general issues, the aim is not to generalize the findings to other cases. Also an intensive, detailed description and analysis of a single project, program, or instructional material in the context of its environment. In driver education, a case could also be a student, a class, or an instructor.

**Clinical Trial:** An experiment where the participants are patients, usually involving a comparison of a treatment group (that receives a treatment or intervention) and a control group that does not.

**Closed-ended Question:** A question followed by predetermined response choices, such as multiple choice, scales, or yes/no. Many closed-ended questions have “other” as the last alternative for respondents to specify their answer in words.

**Cluster Sample:** A probability sample for which groups or jurisdictions comprising groups were randomly selected.

**Coding:** The process of transforming data, evidence, information, judgments, notes, and responses to numeric and/or alphabetic codes for data analysis.

**Comparability:** The similarity of phenomena, such as attributes, performances, assessments, and data sources, being examined. The amount or degree of comparability is often used to determine the appropriateness of using one phenomenon in lieu of another, and to help ensure fairness.

**Confidence Interval:** The probability based on statistics, that a number will be between an upper and lower limit. The measure of the precision of an estimated value. The interval represents the range of values, consistent with the data, that is believed to encompass the “true” value with high probability (usually 95%). The confidence interval is expressed in the same units as the estimate. Wider intervals indicate lower precision; narrow intervals indicate greater precision.

**Confidentiality:** The obligation not to disclose the identity of respondents, or the obligation of persons to whom private information has been given not to use the information for any purpose other than that for which it was given.

**Consent:** Voluntary agreement of a person or group to participate in research. This should be obtained in conjunction with the person or group being given adequate information that has to be fully understood by the subjects; hence “informed consent.”

**Content Analysis:** A set of procedures for collecting and organizing non-structured information into a standardized format that allows one to make inferences about the characteristics and meaning of written and otherwise recorded material.

**Context, Input, Process, and Product (CIPP) Evaluation Model:** A comprehensive framework for guiding evaluations of programs, projects, personnel, products, institutions, and systems. The model’s core parts are context, input, process, and product evaluation. These four parts of an evaluation ask, “What needs to be done?” “How should it be done?” “Is it being done?” “Did it succeed?”

**Control Group:** A group as closely as possible equivalent to an experimental treatment group (one that is exposed to a program, project, or instructional material), and exposed to all the conditions of the investigation except the program, project, or instructional material being studied.

**Cost/Benefit Analysis:** A type of analysis that compares the costs and benefits of programs in money terms. If the benefits as expressed in monetary terms are greater than the money spent on the program, the program is considered to be of absolute benefit. Cost/benefit analysis can be used to compare interventions that have different outcomes, and comparisons are also possible across sectors.

**Cost Effectiveness Analysis (CEA):** A tool used to aid decisions about which interventions should be offered. It is a method of comparing the cost and effectiveness of two or more alternatives. The goal is to determine whether an intervention’s value justifies its cost. Cost effectiveness involves more than determining cost; it also involves assigning a value to the outcome. A more technical definition is that CEA is an analysis of two or more alternatives in order to identify the alternative with the highest input/output ratio; that is, to achieve the maximum output, or the result with the minimum input or costs.

**Covariate:** A variable occurring incidentally along with the variable of primary interest and measured to make informed adjustments on the variable of primary interest; for example, measuring pre-test performance of two groups to adjust their post-test scores so that differences between groups that existed prior to the treatment of one of the groups are taken into account.

**Critical Path Method (CPM):** CPM models a project’s activities and events as a network diagram. Activities are depicted as nodes or circles, and events signifying the beginning and end of activities are shown as lines or arcs between the nodes. CPM was developed for complex but fairly routine projects where minimal uncertainty exists in project completion times.

## D

**Data:** The information produced by or used in an evaluation. Data are numbers, words, pictures, ideas, or any type of information used.

**Data Analysis:** The process of organizing, summarizing, and interpreting numerical, narrative, or artifact data, so that the results can be validly interpreted.

**Data Source:** Identifies the origin of the information to be collected.

**Demographic Information:** Descriptive data that include race/ethnicity, gender, age, grade level, socioeconomic status, and similar kinds of information. This information can help in the analysis of program impact on different groups of participants, and in proving the audience targeted by the program was reached.

**Document Review:** Examining records or documents that reveal information about the context in which a program occurs, about people’s behavior, and about other conditions or events. Evaluators can use existing records, such as test results, or develop forms especially for an evaluation, such as participant journals and attendance sheets.

## E

**Ecological Study:** An investigation in which populations or groups of people, rather than individuals, are examined. Also, evaluation of a program or intervention as it is operationally implemented or of “naturally occurring” changes.

**Effectiveness:** A measure of the extent to which a program achieves its planned results (outputs, outcomes, and goals), or of how economically or optimally inputs (financial, human, technical, and material resources) are used to produce outputs.

**Epidemiology:** The study of the occurrence and causes of health effects in human populations. An epidemiological study often compares two groups of people who are alike except for one factor, such as exposure to a chemical or the presence of a health effect. The investigators try to determine whether any factor is associated with the health effect. Other definitions include the study of the distribution and determinants of health-related states or events in populations and the application of this study to the control of health problems, and the study of determinants (risk factors) and distribution of disease among populations.

**Evaluability:** The extent to which an activity or a program can be evaluated in a reliable and credible fashion.

**Evaluation:** A time-bound exercise that attempts to assess systematically and objectively the relevance, performance, and success, or lack thereof, of ongoing and completed programs. Evaluation is undertaken selectively to answer specific questions to guide decision makers and program managers. It provides information on whether underlying theories and assumptions used in program development were valid, what worked, what did not work, and why. Evaluation commonly aims to determine a program's relevance, validity of design, efficiency, effectiveness, impact, and sustainability.

**Evaluation Design:** A blueprint, strategy, or outline developed to answer questions about a program. It includes a clear statement about the purpose and plans for gathering, processing, and interpreting the information needed to answer the questions. More specifically, it represents the set of decisions that determine how an evaluation is to be conducted, including identifying purposes and use of information, developing or selecting assessment methods, collecting assessment information, judging, scoring, summarizing and interpreting results, reporting evaluation findings, and following up on evaluation results.

**Evaluation Methods:** Data collection options and strategies selected to match or fit the overall design and answer the evaluation questions. Methods depend on knowing who the information is for, how it will be used, which types of information are needed and when, and the resources available.

**Experimental Design:** The plan of an experiment, including selecting subjects; determining the order of administering the experimental treatment, the kind of treatment, and the procedures by which it is administered; as well as recording data (with special reference to the particular statistical analyses to be performed).

**External/Independent Evaluation:** An evaluation conducted by individuals or entities free of control by those responsible for the design and implementation of the program to be evaluated.

# F

**Face Validity:** The quality of a measure that makes it seem a reasonable measure of a variable. Face validity is concerned with how a measure or procedure appears. Does it seem like a reasonable way to gain the information? Does it seem well designed? Does it seem as though it will work reliably?

**Feasibility:** The coherence and quality of a program strategy that makes successful implementation likely. The extent to which resources allow an evaluation to be conducted.

**Feedback:** Transmission of findings of monitoring and evaluation activities organized and presented in an appropriate form for dissemination to users in order to improve program management, decision making, and organizational learning. Feedback may include findings, conclusions, recommendations, and lessons learned from experience.

**Finding:** A factual statement about a program based on empirical evidence gathered through monitoring and evaluation activities.

**Focus Group:** A qualitative technique developed by social and market researchers in which 6-12 individuals are brought together to talk about their views and impressions on a specified topic. This can include sharing insights and observations, obtaining perceptions or opinions, suggesting ideas, or recommending actions on a topic of concern. Focus groups are often homogeneous with members being generally of the same age, gender, and status to encourage participation. This method provides in-depth and insightful information from a relatively small number of people.

**Formative Evaluation:** A type of evaluation undertaken during program implementation to provide information that will guide program improvement. A formative evaluation focuses on collecting data on program operations so that changes or modifications can be made to the program in its early stages. Formative evaluations are used to provide feedback to program managers and other personnel about aspects of the program that are working and those that need to be changed.

# G

**Generalizability:** The extent to which information about a program, project, or instructional material collected in one setting can be used to reach a valid judgment about how it will perform

in other settings. The appropriateness of using results from one context or purpose in another context or for another purpose.

**Goal:** A higher order objective to which a program or intervention is intended to contribute.

**Grounded Theory:** Theory where the researcher develops conceptual categories from the data and then makes new observations upon which to create the theory. Hypotheses are derived directly from the data, and may be tested against it. All conclusions must be grounded in and supported by the data.

## I

**Impact:** Positive and negative long-term effects on identifiable population groups produced by a program intervention, directly or indirectly, intended or unintended.

**Impact Evaluation:** A type of outcome evaluation that focuses on the broad, longer-term impact or results of a program.

**Indicator:** A specific, measurable item of information that specifies progress toward achieving a result. More specifically, a quantitative or qualitative measure of program performance used to demonstrate change and which details the extent to which program results are being or have been achieved. For indicators to be useful for monitoring and evaluating program results, it is important to identify indicators that are direct, objective, practical, and adequate and to regularly update them.

**Informed Consent:** A written or verbal agreement in which potential participants agree to participate in a study after receiving adequate information about the study to make a reasoned decision.

**Inputs:** The resources used to conduct a program.

**Instrument:** A tool used to measure or study a person, event, or other object of interest. Examples are topic guides for focus groups (qualitative instrument) and questionnaires for surveys (quantitative instrument).

**Intermediate Measures and Outcomes:** Tests or instruments used to assess program outcomes, that is, measurements of things (outcomes) that are intermediate between the program and its impacts.

**Internal Evaluation:** Evaluation conducted by a staff member or unit from within the organization being studied.

**Interview:** A series of orally delivered questions designed to elicit responses concerning attitudes, information, interests, knowledge, and opinions. Interviews may be conducted in person or by telephone, and with an individual or a group. The three major types of interviews are: (1) structured, where all questions to be asked by the interviewer are specified in advance; (2) semi-structured, where the interviewer can ask other questions and prompts in addition to the specified questions; and (3) unstructured or open-ended, where the interviewer has a list of topics (topic guide) with no or few specified questions, and respondents can shape the direction of the interview by being encouraged to express their own story from their own perspective.

**ISO 9000 Standards:** Standards primarily concerned with quality management. This refers to what an organization does to fulfill customers' quality requirements, and applicable regulatory requirements, while aiming to enhance customer satisfaction, and achieve continual performance improvement in pursuit of these objectives. Most ISO standards are specific to a particular product or process, but the ISO generic management system standards are the most widely used. They can be applied to any organization, large or small, whether its product is a service, in any sector, and whether it is a business enterprise, a public administration, or a government department. Management systems refer to the organization's structure for managing its processes or activities that transform resource inputs into a product or service that meets the organization's objectives, and in some cases, legal requirements.

## K

**Knowledge Construction:** A methodological approach that assumes knowledge is not available, and therefore needs to be built or constructed as well as acquired. Knowledge construction can be contrasted with knowledge acquisition.

## L

**Learning Outcomes:** Products of instruction or exposure to new knowledge or skills. Examples include the mastery of a new skill or successful completion of a training program.

**Logic Model:** A systematic and visual way to present the perceived relationships among resources available to operate the program, planned activities, and the changes or results to be achieved. This planning and evaluation tool most often takes the form of a graphic repre-

sentation (flowchart, diagram, or table) that depicts the linkages among program assumptions, goals, objectives, activities, target and stakeholder groups, and outcomes.

**Longitudinal Study:** A quasi-experimental study in which repeated measurements are obtained prior to, during, and following the introduction of an intervention or treatment in order to reach conclusions about the effect of the intervention. It can be either repeated measures or time-series study.

## M

**Measure:** An instrument or device that provides data on the quantity or quality of the aspect of performance being evaluated.

**Merit:** The excellence of an object as assessed by its intrinsic qualities or performance. Quality is usually independent of context or cost.

**Metaevaluation:** Assessing an evaluation to judge its quality and/or assess the performance of the evaluators.

**Methodology:** A description of how something will be done. A set of analytical methods, procedures, and techniques used to collect and analyze information appropriate for evaluating the particular program, component, or activity.

**Modeling:** Creating a numerical representation of a program or process for purposes of statistical analysis.

**Monitoring:** A continuous management function that aims primarily at providing program managers and key stakeholders with regular feedback and early indications of progress, or lack thereof, in achieving intended results. Monitoring tracks actual performance against what was planned or expected according to predetermined standards. It generally involves collecting and analyzing data on program processes and results and recommending corrective measures. It involves checking on a process or a person to verify that progress is being made, required activities are occurring, assessment and evaluation procedures are being implemented, suggested practices are being tried, prior information is still applicable, earlier decisions can still be justified, and standards are being met.



**Objectives:** Specific desired program outcomes.

**Observation:** A research method in which the investigator systematically watches, listens to, and records the phenomenon of interest.

**Open-ended Question:** A question in a semi-structured questionnaire or topic guide that allows respondents to respond in their own words. Occasionally open-ended questions may appear in a structured interview using a “closed question” instrument. This is not that common, however, due to difficulties of analyzing these quantitatively.

**Outcome:** The intended or achieved short- and medium-term effects of an intervention’s outputs. Outcomes represent changes in conditions occurring between the completion of program outputs and the achievement of the program’s impact.

**Outcome Evaluation:** An examination of a related set of programs, components, and strategies intended to achieve a specific outcome. An outcome evaluation gauges the extent of success in achieving the outcome; assesses the underlying reasons for achievement or nonachievement; validates the contributions of a specific organization to the outcome; and identifies key lessons learned and recommendations to improve performance.

**Outputs:** Products and services that result from completing activities within a program or intervention.

## P

**Participant Observation:** A research method involving direct participation of the researcher in the events being studied and observation of the interactions between the researcher and subjects to achieve a greater understanding.

**Pedagogy:** The study of the methods and application of educational theory to create learning contexts and environments—the art or science of teaching and the methods used to teach.

**Pilot Study/Testing:** A small, preliminary test, dress rehearsal, or trial run. This should be a mirror image of the research evaluation to be done, only on a much smaller scale. Interviews,

questionnaires, sampling, and initial analysis should all be considered. The results of the pilot are used to improve the program or evaluation procedures being piloted before they are used on a larger scale.

**Population:** The whole group about which the evaluator wants to draw conclusions. All the members of a population are potential subjects. Usually we cannot gather information from everyone in a population, so a sample needs to be created or drawn. A sample is a subgroup taken from the population that is often meant to be representative of the population.

**Post-test:** A test to determine performance after administering a program, project, or instructional material.

**Pre-test:** A test or measurement to determine performance before administering a program, project, or instructional material, which is then compared with the results of a post-test; a preliminary test administered to determine baseline data; or the advance testing of something, such as a questionnaire, product, or idea (i.e., a pilot test).

**Process Evaluation:** A type of formative evaluation that assesses ongoing program processes to determine whether a program is operating as intended. A process evaluation helps program managers identify changes needed in design, strategies, and operations to improve performance.

**Product Evaluation:** Similar to process evaluation, with a focus on products rather than processes, a product evaluation helps program managers and consumers identify changes needed in the design and utility of their products to improve performance.

**Program Evaluation and Review Technique (PERT):** A project model that depicts activities and events in a similar manner as the CPM model, but also allows for randomness in activity completion times.

**Program Logic:** Synonymous with program theory. Usually used when program theory is very simple or straightforward, or when theory is lacking.

**Program Theory:** An approach for planning and evaluating programs or interventions. Program theory entails systematic and cumulative study of the links between inputs, activities, outputs, outcomes, impacts, and contexts of interventions. It specifies how activities will lead to outputs, outcomes, and longer-term impact, and identifies contextual conditions that may affect the achievement of results. A program theory describes and explains the relationships among the various parts of a program and its objectives—how the program is expected to achieve its goals.

**Prospective Longitudinal Study:** A study design that involves collecting data at different points in time, following individuals over a period of time from the present into the future.

## Q

**Qualitative Data:** Information gathered from evaluation methods such as personal interviews, focus groups, observations, diaries, and documents such as case histories, correspondence, and records. This type of data can include detailed descriptions of situations, events, people, interactions, observed behaviors, and people's own thoughts about their experiences, attitudes, beliefs, and behaviors.

**Qualitative Evaluation:** A type of evaluation that is primarily descriptive and interpretative, and may or may not lend itself to quantification.

**Qualitative Research:** The approach advocated as a means to understanding social phenomena. Generally viewed as any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification, and includes in-depth interviews, observations, and participant observation.

**Quantitative Data:** Information presented and/or summarized in numerical form.

**Quantitative Evaluation:** A type of evaluation involving numerical measurement and data analysis based on statistical methods.

**Quantitative Research:** A research approach that measures social phenomena and obtains numerical values that can be analyzed statistically.

**Quasi-experiment:** A research method that compares naturally occurring or other groups that are not randomly assigned. Carefully matching treatment and control groups greatly reduces or may eliminate the likelihood that the groups were different in important ways at the outset.

**Questionnaire:** An instrument consisting of a series of questions and statements used to collect data and information.

## R

**Random Sampling:** Selecting a number of individuals from a larger group or population, so that all individuals in the population have the same chance of being selected.

**Randomized Controlled Trial (RCT):** A research method in which comparisons are made between treatment and control groups established by random assignment of individuals from the same population.

**Reliability:** The extent to which a measure is consistent and accurate, or the degree to which an instrument consistently measures an attribute. The questions “Are we measuring consistently?” and “How stable is our measure?” reflect concerns of reliability. There are several types of reliability, for example: 1. Intra-rater—the degree to which the measure yields consistent results over different administrations; 2. Inter-rater—the degree to which the measure yields similar results for more than one assessor; 3. Internal consistency—the degree to which individual observations or items consistently measure the same attribute; and 4. Test-retest—the degree to which the measure produces consistent results over several administrations assessing the same attribute of an individual.

**Research:** The general field of disciplined investigation.

**Result:** The output, outcome, or impact (intended or unintended, positive or negative) derived from a cause-and-effect relationship set in motion by a program or intervention.

**Retrospective Longitudinal Study:** A study which assesses changes over time by identifying individuals and assessing what has happened to them in the past.

## S

**Sample:** A subset of people, documents, or things that has similar characteristics to the larger group from which it is selected.

**Sample Size:** The number of individuals selected or drawn from a population for research purposes.

**Sampling:** Techniques used to obtain a subset of a population. This includes “probability sampling” where each subject has a known statistical chance of selection (often used in quantitative studies), and “non-probability” sampling where subjects do not have a known statistical chance of selection (used for qualitative sampling).

**Self-Selection Bias:** The ways in which individuals who choose to expose themselves to a program or intervention differ from those who do not.

**Stakeholders:** People, groups, or entities that have a role and interest in the aims and implementation of a program. They include: the community whose situation the program seeks to change; field staff who implement activities; program managers who oversee implementation; donors and other decision-makers who influence or decide the course of action related to the program; and supporters, critics, and other persons who influence the program environment. These are the individuals or groups who may affect or be affected by a program evaluation.

**Statistical Significance:** Results that are determined to have no more than a small, known probability of occurring by chance, according to appropriate inferential statistical methods.

**Strategies:** Approaches and modalities to deploy human, material, and financial resources and implement activities to achieve results.

**Successful Outcome:** A favorable program result that is assessed in terms of effectiveness, impact, and sustainability.

**Summative Evaluation:** Evaluation designed to present conclusions about the merit or worth of an object and recommendations about whether it should be retained, altered, or eliminated. It includes outcome and impact evaluation that assesses a program's overall effectiveness.

**Survey:** A method of collecting information from a sample of the population of interest. This is usually a quantitative method that allows statistical inferences to be drawn from the sample about the population.

**Sustainability:** Durability of a program or results after terminating the external support for the program.

## T

**Target Group:** A program's main stakeholders who are expected to gain from the results of that program. Sectors of the population that a program aims to reach in order to address their needs.

**Task Analysis:** Identifying the critical components of a complex task, job, or performance requirement.

**Theory:** Generally, a set of logically interrelated propositions and their implications.

**Theory-based/driven evaluation:** An evaluation that uses program theory or logic as its guiding framework, ideally investigating the links among variables in the logic model to check whether

its underlying assumptions are sound. It is an evaluation approach based on an explicit theory or model of how the program is expected to reach its goals.

**Transparency:** Carefully describing and sharing information, rationale, assumptions, and procedures as the basis for value judgments and decisions.

**Transportability:** The appropriateness of extending the use of a policy, program, instrument, assessment procedure, or evaluation system across groups, subject areas, instructional approaches, learning activities, school settings, states, etc.

## U

**Utility:** The value of something to someone or to an institution. The extent to which evaluations meet the information needs of their users.

## V

**Validity:** The extent to which a measure captures the dimension of interest. It is the soundness of the use and interpretation of a measure. The question: “Are we measuring what we’re supposed to be measuring?” reflects the issue of validity. Related to the purposes of the evaluation, the degree to which inferences drawn about a subject’s knowledge, skills, attitudes, and behaviors from the results of research methods are correct, trustworthy, and appropriate for making decisions.

**Variable:** An indicator assumed to represent a related underlying construct or concept.

**Voluntary:** Free of coercion, including any sanctions for not participating.

## W

**Worth:** The value of an object in relationship to a purpose, combining merit, context, and cost.

## Sources

Bond, S. L., S. E. Boyd, & K. A. Rapp. 1997. *Taking stock: A practical guide to evaluating your own programs*, Horizon Research.

<http://www.horizon-research.com/publications/stock.pdf>

*CDC Introduction to Program Evaluation for Comprehensive Tobacco Control Programs. Glossary*

[http://www.cdc.gov/tobacco/evaluation\\_manual/glossary.html](http://www.cdc.gov/tobacco/evaluation_manual/glossary.html)

Center for Program Evaluation, Bureau of Justice Assistance

<http://www.ojp.usdoj.gov/BJA/evaluation/glossary/index.htm>

*CIPP Evaluation Model Checklist*, The Evaluation Center, Western Michigan University

<http://www.wmich.edu/evalctr/checklists/cippchecklist.htm>

Davidson, E. J. 2004. *Evaluation methodology basics*. Thousand Oaks, CA: Sage Publications.

International Organization for Standardization

<http://www.iso.org/iso/en/aboutiso/introduction/index.html>

Joint Committee on Standards for Educational Evaluation. 1994. *The Program Evaluation Standards*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Joint Committee on Standards for Educational Evaluation. 2003. *The Student Evaluation Standards*. Thousand Oaks, CA: Corwin Press.

*Planning and Evaluation Wizard*, South Australian Community Health Research Unit.

[www.sachru.sa.gov.au](http://www.sachru.sa.gov.au)

Porteous, N. L., B. J. Sheldrick, and P. J. Stewart. 1997. *A program evaluation tool kit—A blueprint for public health management*. Ottawa, ON: Public Health Research, Education and Development Program, Ottawa-Carleton Health Department.

Rossi, P. H., M. W. Lipsey, & H. E. Freeman. 2004. *Evaluation. A systematic approach*. 7th ed. Thousand Oaks, CA: Sage Publications.

Stemler, S. 2001. An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Available from <http://PAREonline.net/getvn.asp?v=7&n=17>

The Center for the Advancement of Community Based Public Health. 2000. *An evaluation framework for community health programs*. Durham, NC: CBPH.

The Health Communication Unit. 2006. *Evaluating health promotion programs*. Toronto, ON: Centre for Health Promotion, University of Toronto.

United Nations Population Fund. *Program Manager's Planning, Monitoring, and Evaluation Toolkit—Tool No. 1—Glossary of Planning, Monitoring, and Evaluation Terms*  
[http://www.unfpa.org/monitoring/toolkit/tool1\\_glossary.pdf](http://www.unfpa.org/monitoring/toolkit/tool1_glossary.pdf)

Wheeler, P., G. D. Haertel, and M. Scriven. 1992. *Teacher evaluation glossary*. Kalamazoo, MI: CREATE Project, The Evaluation Center, Western Michigan University.



# **APPENDIX C: Program Evaluation Standards Definitions**

## **Utility Standards**

The following utility standards ensure that an evaluation will serve the information needs of intended users.

### *U1 Stakeholder Identification*

Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

### *U2 Evaluator Credibility*

The persons conducting the evaluation should be both trustworthy and competent in performing the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

### *U3 Information Scope and Selection*

Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

### *U4 Values Identification*

The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

### *U5 Report Clarity*

Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

### *U6 Report Timeliness and Dissemination*

Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

### *U7 Evaluation Impact*

Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

## **Feasibility Standards**

The following feasibility standards ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

### *F1 Practical Procedures*

The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

### *F2 Political Viability*

The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

### *F3 Cost Effectiveness*

The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

## **Propriety Standards**

The following propriety standards ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

### *P1 Service Orientation*

Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

### *P2 Formal Agreements*

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

### *P3 Rights of Human Subjects*

Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.

### *P4 Human Interactions*

Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

### *P5 Complete and Fair Assessment*

The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and the problem areas addressed.

### *P6 Disclosure of Findings*

The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.

### *P7 Conflict of Interest*

Conflict of interest should be handled openly and honestly, so that it does not compromise the evaluation processes and results.

### *P8 Fiscal Responsibility*

The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accountable and appropriate.

## **Accuracy Standards**

The following accuracy standards ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

### *A1 Program Documentation*

The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

### *A2 Context Analysis*

The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

### *A3 Described Purposes and Procedures*

The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

### *A4 Defensible Information Sources*

The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

### *A5 Valid Information*

The information-gathering procedures should be chosen or developed and then implemented, so that they will assure that the interpretation arrived at is valid for the intended use.

### *A6 Reliable Information*

The information-gathering procedures should be chosen or developed and then implemented, so that they will assure that the information obtained is sufficiently reliable for the intended use.

### *A7 Systematic Information*

The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.

### *A8 Analysis of Quantitative Information*

Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

### *A9 Analysis of Qualitative Information*

Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

### *A10 Justified Conclusions*

The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

### *A11 Impartial Reporting*

Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

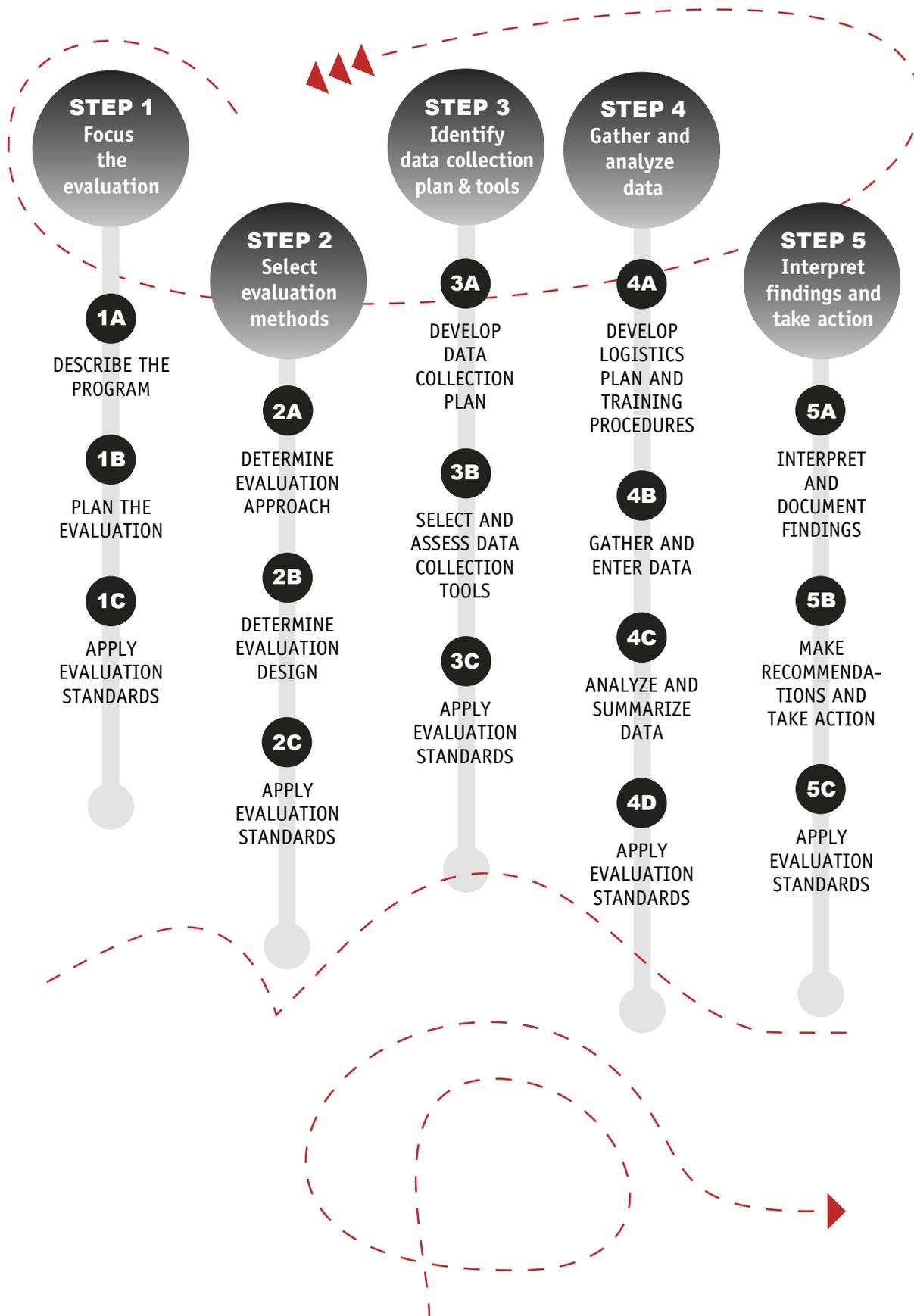
### *A12 Metaevaluation*

The evaluation should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

---

Reproduced from *The Program Evaluation Standards*. 2nd ed. Joint Committee on Standards for Educational Evaluation. 1994.

## APPENDIX D: Overview of Evaluation Steps



# STEP 1

## FOCUS THE EVALUATION

1A DESCRIBE THE PROGRAM	1B PLAN THE EVALUATION	1C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Identify stakeholders, and user and program needs</li> <li>➤ Identify the program's vision, goals, and objectives</li> <li>➤ Identify and document program activities, resources, and context</li> <li>➤ Develop a program logic model</li> <li>➤ Assess program readiness to be evaluated</li> </ul>	<ul style="list-style-type: none"> <li>➤ Identify the purpose of the evaluation</li> <li>➤ Identify knowledge from driver education evaluations</li> <li>➤ Identify potential users and uses of the evaluation</li> <li>➤ Identify key evaluation questions and targets</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

# STEP 2

## SELECT EVALUATION METHODS

2A DETERMINE EVALUATION APPROACH	2B DETERMINE EVALUATION DESIGN	2C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Identify evaluation approach options</li> <li>➤ Determine evaluation level</li> <li>➤ Select research methods</li> </ul>	<ul style="list-style-type: none"> <li>➤ Develop research design</li> <li>➤ Determine sample sizes</li> <li>➤ Develop ethics and rights of human subjects procedures</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

# STEP 3

## DEVELOP DATA COLLECTION PLAN AND SELECT DATA COLLECTION TOOLS

3A DEVELOP DATA COLLECTION PLAN	3B SELECT AND ASSESS DATA COLLECTION TOOLS	3C APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Determine appropriate data types and data gathering methods</li> <li>➤ Specify data and sources</li> <li>➤ Identify indicators for program success</li> <li>➤ Assess feasibility of data collection plan</li> </ul>	<ul style="list-style-type: none"> <li>➤ Select, modify or develop tools</li> <li>➤ Conduct quality assessment of tools and revise</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>

# STEP 4

## GATHER, ANALYZE, AND SUMMARIZE DATA

4A DEVELOP LOGISTICS PLAN AND TRAINING PROCEDURES	4B GATHER AND ENTER DATA	4C ANALYZE AND SUMMARIZE DATA	4D APPLY EVALUATION STANDARDS
<ul style="list-style-type: none"> <li>➤ Develop data collection logistics plan</li> <li>➤ Develop procedures to train data collection personnel and conduct training</li> </ul>	<ul style="list-style-type: none"> <li>➤ Ensure timely and consistent data collection</li> <li>➤ Enter data and ensure accuracy</li> <li>➤ Ensure confidentiality and security of data</li> </ul>	<ul style="list-style-type: none"> <li>➤ Identify data analysis procedures and conduct data analysis</li> <li>➤ Assess, synthesize, and summarize data analysis results</li> </ul>	<ul style="list-style-type: none"> <li>➤ Apply relevant standards</li> </ul>



**STEP  
5**

## INTERPRET AND ACT UPON THE EVALUATION FINDINGS

<b>5A</b> <b>INTERPRET AND DOCUMENT FINDINGS</b>	<b>5B</b> <b>MAKE RECOMMENDATIONS AND TAKE ACTION</b>	<b>5C</b> <b>APPLY EVALUATION STANDARDS</b>
<ul style="list-style-type: none"> <li>➤ <i>Interpret findings</i></li> <li>➤ <i>Prepare conclusions and make judgments</i></li> <li>➤ <i>Document evaluation process and findings in evaluation report</i></li> <li>➤ <i>Undertake peer review</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Prepare recommendations</i></li> <li>➤ <i>Ensure feedback, follow-up, and dissemination of evaluation results</i></li> <li>➤ <i>Undertake actions to ensure use of evaluation and share lessons learned</i></li> <li>➤ <i>Determine changes to implement in next evaluation cycle and prepare action plan</i></li> </ul>	<ul style="list-style-type: none"> <li>➤ <i>Apply relevant standards</i></li> </ul>

# APPENDIX E: Worksheets

## Worksheet #1: Organizing Program Information

Program Goal:		
Objective:		
Expectations	Activities	Resources
1.	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>
2.	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>
3.	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>
4.	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>
5.	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>	<ul style="list-style-type: none"><li>•</li><li>•</li><li>•</li><li>•</li></ul>



(Worksheet #2 continued)

continued from page 240

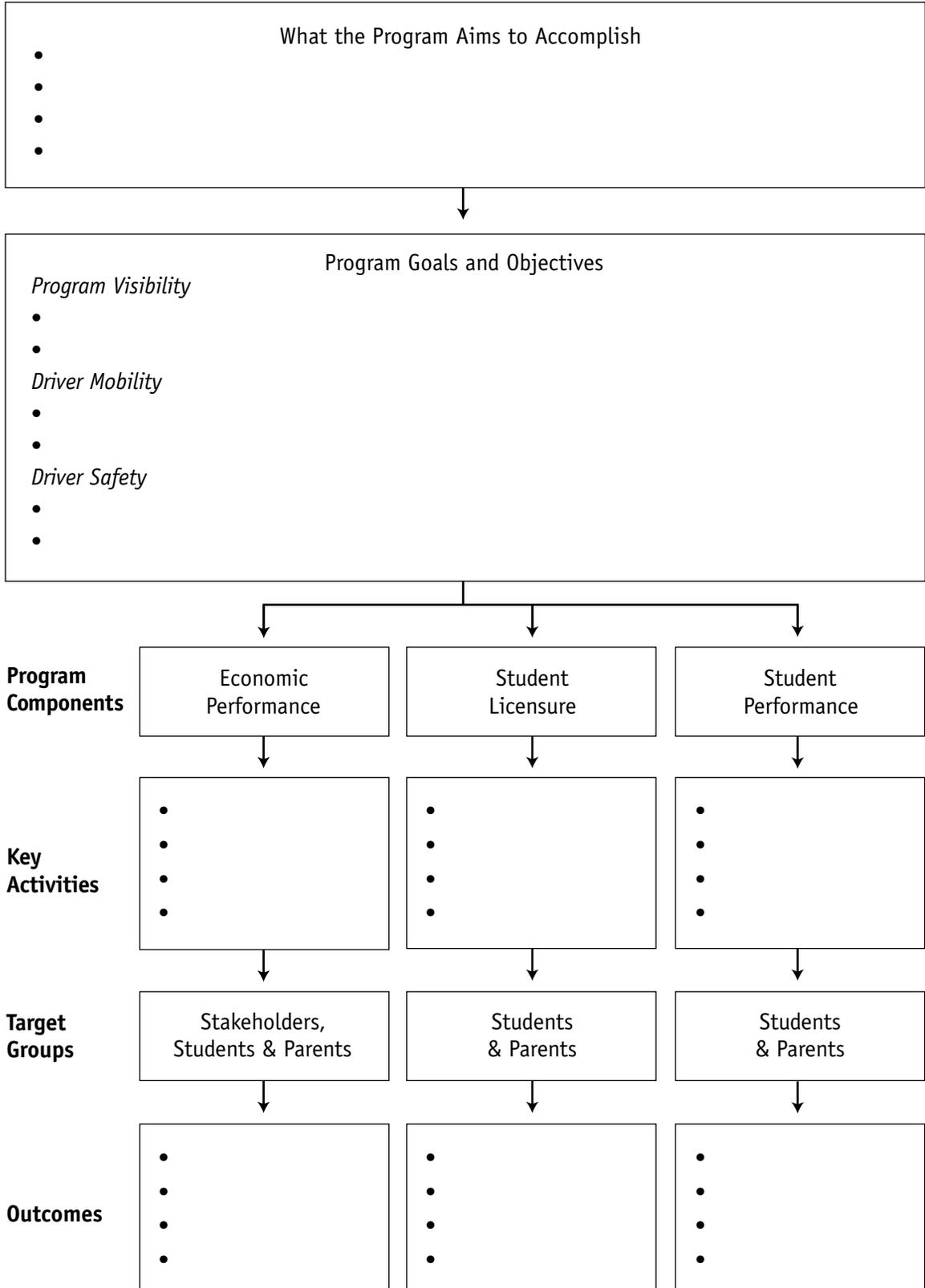


OTHER EXPENSES			TIME		FEASIBILITY
Equipment, Supplies, Space	Cost	Funds	Date Required	Meet Deadline?	Yes/No

**Worksheet #3A: Driver Education Program Logic Model**

Program Goals and Objectives	Program Processes and Activities	Outcomes	Target Groups
<b>PROGRAM VIABILITY</b>			
Economic competitiveness			
<b>DRIVER MOBILITY</b>			
Starting independent driving career			
<b>DRIVER SAFETY</b>			
Performance capability			

## Worksheet #3B: Program Logic Model



Adapted from *A Program Evaluation Tool Kit*, Porteous, Sheldrick, and Stewart 1997.



<p><b>Outcomes</b></p> <p><i>Think about which outcomes are crucial. Which outcomes are the most difficult to achieve?</i></p>	<p><b>How important are the answers to these questions for this evaluation?</b></p> <p><i>High Medium Low</i></p>		
Have the program's outcomes listed below been achieved?			
•			
•			
•			
•			
•			
•			
•			
Have the program's impacts listed below been achieved?			
•			
•			
•			
•			
•			
•			
How can the program be improved? List the aspects of the program that are to be evaluated.			
•			
•			
•			
•			
•			
•			
•			
•			

Adapted from *A Program Evaluation Tool Kit*, Porteous, Sheldrick, and Stewart 1997.

### Worksheet #5: Evaluation Targets

Program Areas	Possible Evaluation Targets	Important Questions to Ask*	Answers	Evaluation Target? Yes or No
Logic or Theory	User needs	What is known about the users of my program? What is known about their needs?		
	Logic model	Has a logic model been developed for the program?		
	Evaluability	Is the program ready to be evaluated?		
Program Context	Stakeholder expectations	How much is known about the program's stakeholders and their expectations of the program?		
	Regulatory environment	Are there any current concerns or issues related to regulation?		
	Contractual environment	Are there any current concerns or issues related to contractual arrangements?		
Business Processes	Quality management and control	Is enough known about quality control of the program delivery and operations?		
	Operations Management	Are there problems with program operations and facilities?		
	Marketing	Are marketing activities working well?		
	Customer service	Is enough known about customer service and satisfaction?		
	Sustainability	Are revenues, consumer support, and growth at acceptable levels?		

Program Areas	Possible Evaluation Targets	Important Questions to Ask*	Answers	Evaluation Target? Yes or No
Program Standards	Program benchmarking and certification	Is the program benchmarked or certified against industry standards?		
	Transportability of program	Is enough known about how transportable the program is, (e.g., into other jurisdictions, or franchising)?		
Instructional Products	Curriculum materials	Is enough known about how effective the classroom materials are in increasing students' knowledge?		
	Tests and measurement	Do we want to measure changes in knowledge?		
Instructional Processes	Instructor preparation	Do we want to evaluate a change in how instructors are trained?		
	Curriculum delivery; in-car practice	Do we want to evaluate the effectiveness of how the curriculum is delivered?		
	Instructional facilities	Do we want to assess the adequacy of program facilities?		
Student Outcomes <i>(continued)</i>	Knowledge outcomes	Do we want to know about overall knowledge change as a result of the program?		
	Skill outcomes	Do we want to know about driving skills gained as a result of the program?		

Program Areas	Possible Evaluation Targets	Important Questions to Ask*	Answers	Evaluation Target? Yes or No
Student Outcomes (continued)	Motivation outcomes	Do we want to know about students' motivations and attitudes toward safe and responsible driving?		
	Mobility outcomes	Do we want to know how many of our students become licensed?		
	Behavioral outcomes	Do we want to measure our students' driving performance after licensure, driving choices, amount of driving, and habit formation?		
Social Impacts	Crash and injury reduction impacts	Do we want to measure our students' crash and injury rates after licensure?		
	Socioeconomic impacts	Do we want to know what a cost/benefit analysis of our program indicates?		
Meta-evaluation	Evaluation effectiveness	Do we want to know how effective our evaluation is and how to improve it?		

\*These questions build on the evaluation questions from Worksheet 4, and their answers will help determine which evaluation targets are most important for the evaluation.





## APPENDIX F: Data Collection Tools: Basic Evaluation Scenario



- Quality Control Interview Guide
- Student Focus Group Guide
- Instructor's Classroom Observation Log Book
- Instructor's In-Car Observation Log Book
- Parent/Guardian Satisfaction Survey
- Student Knowledge Test
- Student Exit Survey
- Parent Feedback Sessions
- Oregon Driver and Traffic Safety Education Self-Study Assessment Tool for Driver Education Program Coordinators
- Oregon Driver Education Program Evaluation Forms

# Quality Control Interview Guide

## Introduction

- Explain the reason for the interview with instructor.
- Indicate the type of information that is being gathered.
- Indicate that the information will be kept confidential, and no individual comments will be identified without permission.
- If applicable, request permission to record or videotape the interview.

## Discussion Areas Related to Quality Control of Program Materials and Delivery Processes

- Identify and describe the specific program processes for which information is being gathered, and ask about how each process is implemented.
  - Course outline
  - Session/class outlines and content
  - Textbook
  - Audiovisual aids such as videos, tapes, overheads, slides
  - Handouts
  - Exercises
  - Tests
  - Log books
  - Instructor's manual
  - Record keeping
- For each process, ask the instructor to talk briefly about how consistent its use is, and if not consistent, to explain the obstacles or barriers to consistency.
- Ask the instructor for his or her views on the consistency of curriculum delivery and methods.
- Ask the instructor for views on quality of materials and delivery processes.
- Ask the instructor to identify areas where quality and consistency need the most improvement.
- Ask the instructor for suggestions on how to improve identified areas.
- Assess information against program standards.

# Student Focus Group Guide

## Introduction

- Provide introduction and explain the purpose of the focus group, e.g., to obtain input from students on the quality of instruction and materials, both classroom and in-car. Emphasis is on what they have learned and how it helps them become good drivers.
- Explain confidentiality of information provided, no individual comments are identified with the student.
- Describe the process: informal discussion, guided by facilitator, everyone encouraged to participate, no censoring of individual comments, will take about 1.5 hours.
- Request permission to record or videotape the session.
- Express appreciation of students' willingness to participate.
- Have students introduce themselves.

## Discussion Guide

### 1. Classroom Course

- What were your expectations about what you would learn at the beginning of the course?
- Which sections were most informative and useful? Why?
- Which sections were most interesting? Why?
- Which sections were most important? Why?
- What could be improved and how? What would you like to see done differently?
  - Materials including textbook, handouts, audiovisuals, charts, pictures, logs
  - Instruction methods, for example, lectures, group work, projects, group discussion, role playing, guest speakers, student presentations, length of classes
  - Time allotted to sections of course
  - Subject areas covered
  - Pace of instruction- Too fast? Too slow?
- What was missing?
- What should the course spend more or less time on?
- What didn't you like and why?

- What was least useful? Most informative?
- Will the course help you to be a safe driver? Responsible driver?
- Insights from the course?
- What was the single most important thing you learned? Peak moments?

## **2. In-Car Sessions**

- What were your expectations about what you would learn from the in-car lessons?
- Which lessons were most informative, useful? Why?
- Which lessons were most important? Why?
- What could be improved and how? What would you like to see done differently?
- What was least useful? Most informative?
- What didn't you like and why?
- Pace of instruction -Too fast? Too slow?
- Reaction to having other students in car (where appropriate)?
- Will the in-car lessons help you to be a safe driver? Responsible driver?
- What was the single most important thing you learned?
- How much practice driving did you do between lessons and with whom? Provide time frames to help students answer, such as none, an hour a week, 2-3 hours, 4-5 hours, more than 5 hours?
- How confident are you in your driving ability as a result of taking this course?

## **3. User Needs**

- Why did you decide to take driver education and this course in particular?
- What were your parents' expectations for you taking this course?

## **4. Attitudes**

- Do you think it is important to be a safe driver?
- How important is it to be responsible to others when you are driving?
- How confident are you in your ability to drive? How good a driver are you?
- Do you think it is OK to take risks when you are driving?
- Do your friends influence how you drive?
- Do you think how you drive will help you avoid crashes?

## Instructor's Classroom Observation Log Book

During every class, instructors can keep track of things that work and don't work; of areas where students have more difficulty understanding and learning, taking note of approaches and processes that are effective or less so; and of materials that catch students' attention and those that appear boring or outdated. This log provides instructors a way to regularly capture important information about the classroom lessons.

This log should be maintained on an as-needed basis but regularly referred to so that important areas for attention are not forgotten or overlooked. For the first few courses in which the log is used, the instructors can be asked to complete one sheet for each classroom session. This will help instructors look for and note areas that need attention and possibly improvement.

The instructors' entries can be discussed at monthly meetings, and solutions and actions identified. Not taking on too many changes simultaneously is important. If several things need addressing, their relative importance should be determined and a priority list developed. Changes should also be checked against other evaluation activities for consistency and possible duplication and overlap before being implemented.

### Log Book Format

Date and Session	Priority Tracking Areas	Rating
Session #: Date:	Information concepts	<ul style="list-style-type: none"><li>• Easy, about right, difficult.</li><li>• Identify the most difficult concepts.</li></ul>
	Materials used	<ul style="list-style-type: none"><li>• List and identify which worked well and which need improvement.</li><li>• Identify improvements that can be done on own and those that require discussion with management.</li></ul>
	Instructional processes used	<ul style="list-style-type: none"><li>• List and identify which worked well and which need improvement.</li><li>• Identify improvements that can be done on own and those that require discussion with management.</li></ul>
	Students' needs	<ul style="list-style-type: none"><li>• Identify what works and what doesn't for different students, (e.g., fast vs. slow learners, different learning styles, males vs. females).</li></ul>
	Students' reactions	<ul style="list-style-type: none"><li>• Identify those aspects of session that were most and least interesting.</li><li>• Identify ways to improve interest levels.</li></ul>
	Instructor's needs	<ul style="list-style-type: none"><li>• Identify aspects of session that need improvement from instructor's perspective, (e.g., behavior control, time allocations, resources available).</li></ul>

## Instructor's In-Car Observation Log Book

Instructors can also keep track of things that work and don't work during the in-car lessons. Maneuvers that students have most difficulty executing, the amount of practice students get outside of lessons, and teaching methods that are more or less effective are examples of important information that can be kept track of by using a log book. It is similar to the one created for instructors to note important classroom information.

The log should similarly be maintained on an as-needed basis but regularly referred to so that important areas for attention are not forgotten or overlooked. For the first few courses in which the log is used, instructors can be asked to complete a sheet for each in-car lesson.

The instructors' entries can be discussed at monthly meetings, and solutions and actions identified. Again, too many changes should not be taken on simultaneously. Develop a priority list, check for consistency and avoid duplication.

### Log Book Format

Date and Session	Priority Tracking Areas	Rating
Session #: Date:	Information concepts	<ul style="list-style-type: none"><li>• Easy, about right, difficult.</li><li>• Identify the most difficult concepts.</li></ul>
	Instructional methods used	<ul style="list-style-type: none"><li>• List and identify which worked well and which need improvement.</li><li>• Identify improvements that can be done on own and those that require discussion with management.</li></ul>
	Students' needs	<ul style="list-style-type: none"><li>• Identify what works and what doesn't for different students, (e.g., fast vs. slow learners, different learning style, males vs. females, different levels of experience with vehicles).</li></ul>
	Students' reactions	<ul style="list-style-type: none"><li>• Identify those aspects of lesson that were most and least interesting.</li><li>• Identify ways to improve interest levels.</li></ul>
	Instructor's needs	<ul style="list-style-type: none"><li>• Identify aspects of each lesson that need improvement from instructor's perspective, (e.g., behavior control, time allocations, scheduling).</li></ul>
	Licensing information	<ul style="list-style-type: none"><li>• When instructor accompanies students to the license test, outcomes are to be recorded.</li></ul>

# Parent/Guardian Satisfaction Survey

## Introduction

Provide a brief introduction at the beginning of the survey to explain its purpose, provide assurance of confidentiality, and encourage parents to complete it.

For example:

*Knowing what you think of the driver education course your son or daughter has just finished is important to us. We would appreciate your time to complete this questionnaire. Your answers will be kept confidential and will never be associated with you directly. We will use your input to improve our course. Thank you.*

## Questionnaire Items

The questionnaire is divided into four parts, as follows:

### Part 1: Demographics

1. Gender
2. Highest level reached in school
3. Number of vehicles in family
4. Number of family members who are licensed to drive

### Part 2: Needs and Reasons for Son/Daughter Taking the Course

1. What are your needs relative to your son or daughter taking our course?

Choose as many answers as appropriate from the following list:

- Excellence in teaching students how to drive in order to pass the licensing test
- Excellence in teaching students how to be safe and good drivers
- Convenience of location
- Convenience of classroom schedule
- Convenience of in-car schedule
- Qualification of instructors
- Quality of instruction methods
- Affordability
- Other (specify)

2. Why did you decide that your son or daughter should take our course?

Choose as many answers as appropriate from the following list:

- Price
- Location

- Word of mouth recommendation
- Marketing
- Reputation of excellence
- Qualifications of instructors
- Timing/schedule of classes
- Other member of family has taken the course
- Insurance discount
- Other (Specify)

### Part 3: Overall Opinions of Course

1. Overall, the course has met my expectations.  Yes  No  
Please explain your answer.

2. Ask parents their opinions about the course using agree-disagree questions, each with a scale from 1-5 where 1 represents complete dissatisfaction and 5 represents complete satisfaction, as follows:

Put an **X** in the box that best describes **how much you disagree or agree** with each statement:

- |                          |                          |                          |                          |                          |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> |
| 1 Completely<br>Disagree | 2 Somewhat<br>Disagree   | 3 Undecided              | 4 Somewhat<br>Agree      | 5 Completely<br>Agree    |

#### Questionnaire Items\*

- I think the program is valuable for training new drivers.
- I believe my teenager thinks the course is valuable for training new drivers.
- I think that young drivers who take the course are more skilled than those who do not take the course.
- I think that if my son or daughter did not take the course, he or she would have more accidents once he or she gets licensed.
- If I knew a high school student who was planning to get a driver's license soon, I would recommend he or she take the course.
- The course has increased my confidence in my son or daughter's driving.
- The course will help my son or daughter be a more cautious driver.
- I think the course is better than lessons from another driving school.
- I think the course has been a good preparation for my teenager taking his or her driver's license test.

---

\*Some of these items are taken from questionnaires developed for the study, *A Longitudinal Analysis of Manitoba Public Insurance Driver Education Program*, Lonero et al. 2005.

#### Part 4: Input on Specific Aspects of Course

1. Ask parents to rate each aspect of the course on a scale from 1-5 where 1 represents complete dissatisfaction and 5 represents complete satisfaction, using the following format:

Put an **X** in the box that best describes **how much you are dissatisfied or satisfied** with each program component:

- |                          |              |                          |              |                          |           |                          |           |                          |            |
|--------------------------|--------------|--------------------------|--------------|--------------------------|-----------|--------------------------|-----------|--------------------------|------------|
| <input type="checkbox"/> | Completely   | <input type="checkbox"/> | Somewhat     | <input type="checkbox"/> | Undecided | <input type="checkbox"/> | Somewhat  | <input type="checkbox"/> | Completely |
| 1                        | Dissatisfied | 2                        | Dissatisfied | 3                        |           | 4                        | Satisfied | 5                        | Satisfied  |

#### Questionnaire Items

- Classroom instruction
- Course materials such as textbook and handouts
- Instructional processes that they are aware of, such as lectures, group discussions, group work, role playing, videos, and guest presentations
- In-car instruction
- In-car practice log
- Work load, assignments, quizzes, and tests
- Parent involvement and participation

2. Provide specific comments you have about any of these aspects of the course.

#### Part 5: Things to Change

1. What are the three most important things about the course that you would like to see changed and why?
2. Do you have any suggestions on how to make these changes?

# Student Knowledge Test

## Pre-Post Knowledge Test

Students will be given this test at the beginning of the first class of the course and then again in the last class. The subject areas for each program may differ, as the test will be directly based on the knowledge areas of the program's curriculum.

General subject areas for questions\* include:

### Introduction

- State traffic laws
- Vehicle familiarization
- Driver readiness
- Vehicle control
- Establishing vehicle position

### Traffic Entry Skills

- Basic vehicle maneuvering tasks
- Roadway characteristics
- Roadway signs and signals
- Roadway markings
- Basic vehicle control tasks

### Space Management Skills

- Space management system development
- Turnabouts
- Speed management
- Lane changes
- Perpendicular, angle, and parallel parking

### Developing Space Management Skills

- Traffic flow situations

---

\*Taken from *Traffic Safety Education Life Long Learning Processes*, Highway Safety Center, Indiana University of Pennsylvania 2002.

- Space management situations
- Intersection entry
- Curve entry/exit
- Passing on multiple lanes

#### **Dealing with Complex Environments**

- Traffic flow situations up to maximum speed limit
- Space management situations to maximum speed limit
- Merging/driving on/exiting limited access highway
- Passing
- Passing on multiple lanes

#### **Affecting Driver Performance**

- Driver fitness
- Chemical use/abuse information

#### **Adverse Conditions**

- Adverse conditions preparation
- Occupant protection
- Traffic flow situations under limited conditions of visibility/traction
- Space management assessment

#### **Vehicle functions/malfunctions**

- Vehicle functions/malfunctions

# Student Exit Survey

## Introduction

Provide a brief introduction at the beginning of the survey, explaining its purpose, providing assurance of confidentiality, and encouraging students to complete it.

For example:

*Knowing what you think of the driver education course you have just finished is important to us. We would appreciate your time to complete this questionnaire. Your answers will be kept confidential and will never be associated with you directly. We will use your input to improve our course. Thank you.*

## Questionnaire Items

The questionnaire is divided into four parts, as follows:

### Part 1: Demographics

1. Gender
2. Birth date
3. Highest level reached in school

### Part 2: Reasons for Taking the Course

2. Why did you and your parents decide that you should take our course?

Choose as many answers as appropriate from the following list:

- Price
- Location
- Word of mouth recommendation
- Marketing
- Reputation of excellence
- Qualifications of instructors
- Timing/schedule of classes
- Other member of family has taken the course
- Insurance discount
- Other (Specify)

### Part 3: Overall Opinions of the Course

1. Ask students for their opinions about the course, using an agree-disagree scale for each item, where 1 represents completely disagree and 5, completely agree, as follows:

Put an **X** in the box that best describes **how much you disagree or agree** with each statement:

- <sup>1</sup> Completely Disagree     <sup>2</sup> Somewhat Disagree     <sup>3</sup> Undecided     <sup>4</sup> Somewhat Agree     <sup>5</sup> Completely Agree

#### Questionnaire Items\*

- I think the program is valuable for training new drivers.
- I would be a good driver even if I hadn't taken the course.
- I think that young drivers who take the course are more skilled than those who do not take the course.
- If I hadn't taken the course, I think I would have more accidents once I get my license.
- If I knew a high school student who was planning to get a driver's license soon, I would recommend he or she take the course.
- The course has increased my confidence in my driving.
- The course will help me be a more cautious driver.
- I think the course is better than lessons from another driving school.
- I think the course has been a good preparation for taking my driver's license test.

#### Part 4: Input on Specific Aspects of Course

1. Ask students to rate each aspect of the course on a scale from 1-5, where 1 represents complete dissatisfaction and 5 represents complete satisfaction, using the following format:

Put an **X** in the box that best describes **how much you are dissatisfied or satisfied** with each program component:

- <sup>1</sup> Completely Dissatisfied     <sup>2</sup> Somewhat Dissatisfied     <sup>3</sup> Undecided     <sup>4</sup> Somewhat Satisfied     <sup>5</sup> Completely Satisfied

#### Questionnaire Items

- Textbook
- Course handouts
- Lectures
- Group work
- Presentations

- Audiovisual materials (specify)
- In-car instruction
- In-car log
- Work load, assignments etc.

2. Provide specific comments you have about any of these aspects of the course.

### Part 5: Things to Change

1. What are the three most important things about the course that you would like to see changed and why?
2. Do you have any suggestions on how to make these changes?

### Part 6: Attitudes

1. Ask students to indicate how much they agree or disagree with the following statements\*, using an agree-disagree scale for each item, where 1 represents completely disagree and 5, completely agree, as follows:

Put an **X** in the box that best describes **how much you disagree or agree** with each statement:

- |   |   |   |  |  |
|---|---|---|--|--|
| <input type="checkbox"/> Completely<br>1 Disagree | <input type="checkbox"/> Somewhat<br>2 Disagree | <input type="checkbox"/> Undecided<br>3 | <input type="checkbox"/> Somewhat<br>4 Agree | <input type="checkbox"/> Completely<br>5 Agree |
|---|---|---|--|--|

- I am confident that I know all the rules of the road.
- I feel like the one place where I am totally in control is in my car.
- I live my life for today rather than worrying about the future.
- Even with all the thousands of cars on the roads, there's a lot I can do by myself to avoid a crash.
- I don't mind taking risks. Otherwise, life is too boring.
- If friends told me to drive faster, I would probably not do so.
- Lots of drivers are careless, and I can't do anything about it if they crash into me.
- It doesn't really matter if I drive recklessly, because I'm still better than most drivers.
- I guess I take more driving risks when I am with my friends, but who doesn't?

---

\*Some of these items are taken from questionnaires developed for the study, *A Longitudinal Analysis of Manitoba Public Insurance Driver Education Program*, Lonero et al. 2005.

- If I was a more cautious driver, some of my friends would laugh at the way I drive.
- Today's cars are built safe and most have airbags, so going faster or cornering fast is OK.

### Part 7: Practice Driving

1. How much practice driving did you do per week while you were taking the course?  
Choices: none, 1 hour/week, 2-3 hours/week, 4-5 hours/week, more than 5 hours/week?
2. With whom did you practice?  
Mother, father, stepmother or stepfather, sibling, relative, friend?

### Part 8: Getting Licensed

1. Please complete the attached postcard (with pre-paid postage), and return it to us after you have completed your driver's license test.

Sample Postcard

<p>To: (Name of School)</p> <p>From: (Name of Student)</p> <p>Address:</p> <p>I took my driver's license test on _____ (D/M/Y)</p> <p>Circle the appropriate answers for the following questions:  The result of my test was:    Passed    Failed  If you failed the test, do you intend to retake it?    Yes    No</p>
---

## Parent Feedback Sessions

At the end of each course, parents are invited to attend a feedback session to provide feedback on the program. These sessions also give program staff an opportunity to reinforce the important role that parents play in the driving experiences of their teenagers, even after they get licensed. The evaluation team also sees these sessions as an opportunity to obtain information from parents about their views on what their teenagers learned during the course.

Discussion areas that are added to the agenda for these sessions include:

- How much did your teenagers know about driving before they started the course?
- How much did they learn from the classroom component of the program?
- Can you think of some specific knowledge areas?

# Oregon Driver and Traffic Safety Education Self-Study Assessment Tool for Driver Education Program Coordinators

This tool was developed by the Oregon Department of Transportation, Transportation Safety Division, Driver Education Office to assist program coordinators self-assess their driver education programs. (Reproduced with permission.)

The initial Driver and Traffic Safety Education (DTSE) self-study process shall begin by the start of a semester, and be completed by the school (public/private) DTSE coordinator by the end of the semester. The self-study process shall emphasize the following areas:

1. The participation of staff, parents, community members, and students where appropriate.
2. A comprehensive assessment of the instructional program, staff services, learning resources, student activities, and facilities.
3. The development of a plan for program improvement.

## Curriculum

The curriculum must provide learning experiences, which equip students with knowledge, thought processes, insights, and motivations needed to become safe and efficient drivers. These qualities are instilled through classroom and laboratory learning activities, which are guided by measurable objectives. The best results are obtained when student experiences in the classroom and behind-the-wheel experiences are closely associated in philosophy, content, methods, and scheduling.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Does your school (public/private) have a local curriculum guide?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Has the local guide been revised in the last five years?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Does every teacher and instructor in the program have a copy?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Does the teaching staff use and follow the guide?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	5. Does it include all concepts listed in OAR's (Oregon Administrative Rules)?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	6. Does it include performance objectives appropriate for all concepts?	

- Yes  No 7. Does it include student activities that enable the student to accomplish objectives?
- Yes  No 8. Does it include the level of competency students are to reach for each objective?
- Yes  No 9. Does it include evaluation criteria for classroom and laboratory instruction?

## Teacher-Instructor

The teacher is the most important factor in a quality driver and traffic safety education program. The teacher's responsibility is to set an example for students and to instill in them the concepts of safe and efficient driving. Additionally, the teacher must create learning situations in which the students acquire skills and knowledge they need to develop a responsible attitude toward driving. The closeness of the student-teacher relationship developed during the driver education experience will determine, to a great extent, the overall quality of the program, the levels of skill and knowledge gained, and the attitude of the young driver toward safety.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Is the instructional staff (hereafter called staff) certified by ODOT-Transportation Safety Division?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Does staff have satisfactory driving records, and are they checked annually?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Do any of the staff members have a disability that would limit teaching their assigned areas of traffic safety education, and are physical examinations required every two years?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Do supervisory personnel evaluate staff at least once a year?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	5. Does staff keep parents informed of program activities and student performance?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	6. Is staff given opportunities to keep abreast of new developments in driver education through inservice, professional workshops, and regional, state, and national conferences?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	7. Does staff possess specific knowledge of dual control car instruction, simulation systems, multimedia systems, and related literature?	

- Yes  No 8. Does staff supervise the care and maintenance of vehicles, simulators, test equipment, and other instructional aids?
- Yes  No 9. Does the curriculum include evaluation tests for the objectives?
- Yes  No 10. Is traffic safety education considered an integral part of the school curriculum?
- Yes  No 11. Do teachers of other subjects integrate traffic safety concepts into their classes?
- Yes  No 12. Are classroom and lab instruction integrated and presented concurrently?
- Yes  No 13. Are modes of instruction coordinated into an integrated, sequential, orderly pattern of learning experience?
- Yes  No 14. Is in-car instruction flexible, allowing for individual differences, abilities, and limitations?
- Yes  No 15. Are parents encouraged to provide supervised practice driving?
- Yes  No 16. Is parent involvement encouraged, parent participation guide provided, and time given for parent-teacher interaction?

## Instructional Materials

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Are up-to-date textbooks and basic reference materials available?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Are supplementary teaching materials related to driver and traffic safety education available?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Are commercially sponsored supplementary teaching materials critically reviewed before use?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Are multisensory materials used in light of the objectives in the curriculum?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	5. Are a variety of quality instructional materials available to help students achieve the objectives?	

## Administration

Quality programs rarely exist by chance. They are largely dependent on the interest and capability of program managers and teachers. Such programs are products of an organizational formula that features continuous planning, administrative attention, and supervision based on sound policies and practices. They usually enjoy active support by administrations, which are directly interested in and concerned with development. The most successful programs are understood and supported by parents and community groups.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Do school board members and administrators or owners actively support traffic safety education?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Do teachers, supervisors and administrators/owners cooperatively plan the program?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Are teachers selected on the basis of academic and practical preparation, experience, and their professional competency in traffic safety education?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Does the provider (public/private) provide adequate funds for instructional materials, equipment, and in-service for teachers?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	5. Are goals and objectives of driver education coordinated with the goals of the school and district?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	6. Is adequate insurance provided for traffic safety education vehicles and occupants?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	7. Is someone in the district designated as coordinator of the program?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	8. Does the district have a written board policy on fee collection for completed, withdrawn, dropped, transferred, and repeating students, and consequences for students getting their license before the end of the class?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	9. Does the school have records on students who enroll and are claimed for state reimbursement (3-7 years recommended depending on a public school or community college)?	

- Yes  No 10. Are the records organized by fiscal year (July 1–June 30) and kept according to district retention policy?
- Yes  No 11. Is the program offered at the age when most students are eligible and have their permit?
- Yes  No 12. Is appropriate instruction made available to students with special needs, including those who are handicapped or disabled?
- Yes  No 13. Are academic standards and credit maintained on par with those of other courses?
- Yes  No 14. Is credit toward graduation awarded for successful completion of the course?

## Scheduling

Time frames in this section are recommended from the National Institute for Driver Behavior and the American Driver and Traffic Safety Education Association.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Is the course scheduled for at least 9 weeks but not over 18 weeks in length?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Is the summer course scheduled to be at least five weeks in length?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Are classroom lessons scheduled for two hours or less?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Are behind-the-wheel lessons scheduled for no more than one hour per student per 24-hour period?	

## Evaluation

Evaluation of program effectiveness, an indispensable tool for improving instruction, should include all program functions to assure effective and efficient instruction.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. At the course onset, are written criteria for successful completion given to all students?	

- Yes  No 2. Are students evaluated frequently to determine the value of various instructional units and the sequence in which they're presented?
- Yes  No 3. Is the program evaluated annually by administrators and the instructional staff?
- Yes  No 4. Are student performance records maintained as a guide for program evaluations and to indicate student achievement?
- Yes  No 5. Does the coordinator evaluate the program yearly and make recommendations for content and financial improvements?
- Yes  No 6. Do qualified supervisors or knowledgeable administrators make regular class visitations and objectively evaluate teaching, as one means of trying to improve instructional quality?
- Yes  No 7. Are curriculum guides and instructional materials evaluated and revised annually?

## Facilities

Quality programs are characterized by proper selection and use of instructional facilities, vehicles, equipment, and materials. If the program is to accomplish its goals and objectives, these elements must be adequate for the enrollment, properly maintained, and compatible with instructional intent and requirements.

Response	Statement	Improvement Plan
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Are adequate practice vehicles available that are properly maintained and equipped for all phases of the behind-the-wheel experience?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Are vehicles that are loaned from dealers for driver education used solely for instructional purposes within that program?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Are the responsibilities of school officials and automobile dealers relating to use of vehicles set forth in written agreements?	

- Yes  No 4. Are all vehicles used for driver and traffic safety education inspected yearly?
- Yes  No 5. Are driver education vehicles equipped with dual controls and other equipment such as identification signs, rear view mirrors, a safety belt for each occupant, and emergency equipment as described in the OAR's?
- Yes  No 6. Is a clearly established policy in existence and followed for reporting crashes and damage involving driver education vehicles?
- Yes  No 7. If simulation instruction is provided, is equipment maintained in good working order?
- Yes  No 8. Is an up-to-date driver education textbook, consistent with course content and objectives, readily available to each student throughout the course?
- Yes  No 9. Is adequate audiovisual equipment available when and where it is needed?
- Yes  No 10. Are audiovisual/technology materials used to reinforce, supplement, and improve teacher presentations for both individualized/group instruction?
- Yes  No 11. Are supplementary instructional materials, consistent with program objectives and course content, provided to students when appropriate?
- Yes  No 12. Are practice vehicles and simulation equipment suitably designated/outfitted to meet the needs of each disabled student?

## Support

Community support and media relations enable the school to achieve active public backing not only for driver education but also for the entire school and community safety program. After everything administratively feasible is done to assure that the best instructional program is being provided, students, parents, civic clubs, governmental agencies, community leaders, and news media professionals can effectively aid in publicizing the program.

<b>Response</b>	<b>Statement</b>	<b>Improvement Plan</b>
<input type="checkbox"/> Yes <input type="checkbox"/> No	1. Do teachers, administrators, and others appear before community groups to relate the goals, accomplishments, and needs of driver education?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	2. Is information about the education program provided to the entire family regularly (at least annually)?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	3. Are regular and special news items relating to students, teachers, or the program provided to the media?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	4. Does the school interpret the nature and purpose of driver education for the community?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	5. Do school personnel actively assist and encourage community groups working for program improvement?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	6. Are a variety of communication techniques used to inform parents and the community about the program?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	7. Do teachers ensure that safe driving practices are consistently exhibited on streets, highways, in off-street areas, and on special facility grounds?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	8. Is proper recognition and publicity provided for dealers who provide program vehicles?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	9. Is written information concerning all aspects of the program provided for all parents?	
<input type="checkbox"/> Yes <input type="checkbox"/> No	10. Does a cooperative relationship exist between the school and public agencies responsible for driver and traffic safety?	

# **Driver Education Program Evaluation Forms**

Developed by  
Oregon Department of Transportation  
Transportation Safety Division, Driver Education Office  
(Reproduced with permission)

## **Introduction**

The “Evaluation of Classroom Instructor,” “Evaluation of In-Car Instructor,” and “Evaluation of Driver Education Program” forms are crucial elements if your program is to meet the needs and expectations of the students and parents or guardians. These need to be completed and carefully reviewed at the conclusion of every Driver Education course.

## Evaluation of Classroom Instructor

Classroom Instructor: \_\_\_\_\_

The following statements refer to your classroom instructor. Read each statement and, using the scale below, indicate, using a number between 1 and 10, your feelings. If the statement definitely states your feelings, and you definitely agree with it, your response would be "10." If the statement does not express your feelings, and you definitely disagree with it, your response would be "1." If you are uncertain how you feel about the statement or are neutral, your response would be "5."

**The nearer your answer to "10," the more definite your "YES" answer.**

**The nearer your answer to "5," the more NEUTRAL/UNCERTAIN your answer.**

**The nearer your answer to "1," the more definite your "NO" answer.**

1	2	3	4	5	6	7	8	9	10
NO				MAYBE					YES

Statement	Response
1. a. The instructor seemed to be concerned whether the students learned the material.	_____
b. I enjoyed going to driver education.	_____
c. The instructor knew the subject matter.	_____
d. The instructor seemed to feel the driving aspect of driver education is more important than the classroom aspect.	_____
e. The instructor was prepared for class.	_____
f. The instructor recognized individual differences in the students' abilities.	_____
g. The instructor was generally too involved in lecturing to be aware of the class.	_____
h. The instructor satisfactorily answered the students' questions.	_____
i. The classroom instructor seemed to feel the classroom aspect of traffic safety education is more important than the driving aspect.	_____
j. The classroom instructor was a better-than-average teacher.	_____
k. The instructor provided a good combination of lecture and discussion.	_____
l. The instruction I received in the classroom helped make me a better driver.	_____

## Evaluation of In-Car, On-The-Road Instruction

Behind-The-Wheel Instructor: \_\_\_\_\_

The following statements refer to your in-car, BTW instructor. Please record your response number in the space provided.

**The nearer your answer to "10," the more definite your "YES" answer.**  
**The nearer your answer to "5," the more NEUTRAL/UNCERTAIN your answer.**  
**The nearer your answer to "1," the more definite your "NO" answer.**

1	2	3	4	5	6	7	8	9	10
NO				MAYBE					YES

Statement	Response
1. a. The instructor seemed to be concerned whether the students learned the material.	_____
b. I enjoyed going to this part of the driver education class.	_____
c. The teacher showed mastery of the subject matter.	_____
d. The instructor was prepared for class.	_____
e. The instructor recognized individual differences in the students' abilities.	_____
f. The instructor satisfactorily answered the students' questions.	_____
g. My in-car instructor was a better-than-average teacher.	_____
h. My instructor was genuinely interested in teaching us to drive.	_____
i. My in-car instructor caused emotional stress by asking me to perform a task(s) that had not first been demonstrated or explained.	_____
j. I feel that I am a better driver because I took the driver education course.	_____

2. What did you especially **LIKE** about the in-car, on-the-road instruction?

---



---



---



---



7. How worthwhile were the reports on your teenager's class and driving progress?

- A. Very worthwhile
- B. Somewhat worthwhile

- C. Not very worthwhile
- D. Not at all worthwhile

8. How worthwhile was the parent evening seminar?

- A. Very worthwhile
- B. Somewhat worthwhile

- C. Not very worthwhile
- D. Not at all worthwhile

9. How many times did your teenager take the road test before passing it? (Sometimes students may not test after completion of course until much later.)

- Once
- Twice
- Three or more times

10. If you were to grade the traffic safety education program your teenager received, the grade would be (check one):

- "A"
- "B"
- "C"
- "D"
- "F"

11. Comments about the course:

---

---

---

---

---

---

---

---

---

---



## APPENDIX G: Data Collection Tools: Advanced Evaluation Scenario



- Guide for Stakeholder Interviews
- New Driver Focus Group Guide for Graduate and Comparison Groups
- Student Knowledge Test
- Student Exit Survey
- Example of Mail-out Survey for Driver Education Program Graduates\*
- Example of Mail-out Survey for Newly Licensed Drivers\*
- Example of Mail-out Survey for Parents of Driver Education Program Graduates\*
- Example of Mail-out Survey of General Public Attitudes towards Driver Education\*
- Example of Non-Response Mail-out Survey for Newly Licensed Drivers\*

---

\*These tools were developed for the study, *A Longitudinal Analysis of Manitoba Public Insurance Driver Education Program*, Lonero et al. 2005.

# Guide for Stakeholder Interviews

## Introduction

- Explain the reason for the interview and length of time required.
- Indicate the type of information being gathered.
- Explain that the information will be kept confidential, and no individual comments will be identified without permission.

## Discussion Areas Related to Stakeholder Expectations of Program

- Identify the program aspects that you would like to talk about. A list of possible topics includes:
  - Overall course content and materials
  - Course availability and accessibility
  - Instructor qualifications
  - Customer service and satisfaction
  - Marketing and business processes
  - Program uptake
  - Program effectiveness
- Ask stakeholders for additional views on overall program issues they feel are important or need more attention.
- Ask stakeholders for suggestions on improvements.

Use this input to identify key issues that can be further addressed at a stakeholder workshop.

# **New Driver Focus Group Guide for Graduate and Comparison Groups**

## **Introduction**

- Provide an introduction and explain the purpose of the focus group, e.g., to obtain input from graduates on driving knowledge, skills, and attitudes. Emphasis is on what they think about their driving and what they learned from driver education, or whatever method they used to learn to drive for comparison focus groups.
- Indicate confidentiality of information provided and that no individual comments will be identified with the participant.
- Describe the process: informal discussion, guided by facilitator, everyone encouraged to participate, no censoring of individual comments, will take about 1.5 hours.
- Request permission to record or videotape the session.
- Express appreciation of participants' willingness to participate.
- Have participants introduce themselves.

## **Discussion Guide**

### **1. Knowledge and Skills**

- What were the most important things you learned about safe driving from the driver education course (or whatever method used to learn to drive for comparison focus groups)?
- Did the in-car lessons (or whatever method used to learn to drive for comparison focus groups) help you to be a safe driver? a responsible driver?
- How confident are you in your driving ability as a result of taking the course (or whatever method used to learn to drive for comparison focus groups)?

### **2. Attitudes**

- Do you think being a safe driver is important? Why?
- How important is it to be responsible to others when you are driving?
- How confident are you in your ability to drive? How good a driver are you?
- Do you think that taking risks when you are driving is OK?
- Do your friends influence how you drive?
- Do you think how you drive will help you avoid crashes?
- What motivates you when you are driving?

### 3. Behavior

- What type of trips do you take? What are their purposes?
- Do you have any specific patterns of regular trips that you make?
- What times of day and night do you typically drive?
- How much driving do you do each week?
- Are you usually alone, or do you have other people in the vehicle? Who?
- What types of risks do you take when you are driving?
- Do these risks change depending who is in the vehicle with you?
- How good a driver do you think you are?
- Have you had any crashes since you got your license? What was the outcome?
- Did you learn anything from those experiences?

# Student Knowledge Test

## Pre-Post Knowledge Test

Students will be given this test at the beginning of the first class of the course and then again in the last class. The subject areas for each program may differ, as the test will be directly based on the knowledge areas of the program's curriculum.

General subject areas for questions\* include:

### Introduction

- State traffic laws
- Vehicle familiarization
- Driver readiness
- Vehicle control
- Establishing vehicle position

### Traffic Entry Skills

- Basic vehicle maneuvering tasks
- Roadway characteristics
- Roadway signs and signals
- Roadway markings
- Basic vehicle control tasks

### Space Management Skills

- Space management system development
- Turnabouts
- Speed management
- Lane changes
- Perpendicular, angle, and parallel parking

### Developing Space Management Skills

- Traffic flow situations

---

\*Taken from *Traffic Safety Education Life Long Learning Processes*, Highway Safety Center, Indiana University of Pennsylvania 2002.

- Space management situations
- Intersection entry
- Curve entry/exit
- Passing on multiple lanes

### **Dealing with Complex Environments**

- Traffic flow situations up to maximum speed limit
- Space management situations to maximum speed limit
- Merging/driving on/exiting limited access highway
- Passing
- Passing on multiple lanes

### **Affecting Driver Performance**

- Driver fitness
- Chemical use/abuse information

### **Adverse Conditions**

- Adverse conditions preparation
- Occupant protection
- Traffic flow situations under limited conditions of visibility/traction
- Space management assessment

### **Vehicle functions/malfunctions**

- Vehicle functions/malfunctions

# Student Exit Survey

## Introduction

Provide a brief introduction at the beginning of the survey, explaining its purpose, providing assurance of confidentiality, and encouraging students to complete it.

For example:

*Knowing what you think of the driver education course you have just finished is important to us. We would appreciate your time to complete this questionnaire. Your answers will be kept confidential and will never be associated with you directly. We will use your input to improve our course. Thank you.*

## Questionnaire Items

The questionnaire is divided into four parts, as follows:

### Part 1: Demographics

1. Gender
2. Birth date
3. Highest level reached in school

### Part 2: Reasons for Taking the Course

1. Why did you and your parents decide that you should take our course?

Choose as many answers as appropriate from the following list:

- Price
- Location
- Word of mouth recommendation
- Marketing
- Reputation of excellence
- Qualifications of instructors
- Timing/schedule of classes
- Other member of family has taken the course
- Insurance discount
- Other (Specify)

### Part 3: Overall Opinions of the Course

1. Ask students for their opinions about the course, using an agree-disagree scale for each item, where 1 represents completely disagree and 5, completely agree, as follows:

Put an **X** in the box that best describes **how much you disagree or agree** with each statement:

Completely  
1 Disagree

Somewhat  
2 Disagree

Undecided  
3

Somewhat  
4 Agree

Completely  
5 Agree

#### Questionnaire Items\*

- I think the program is valuable for training new drivers.
- I would be a good driver even if I hadn't taken the course.
- I think that young drivers who take the course are more skilled than those who do not take the course.
- If I hadn't taken the course, I think I would have more accidents once I get my license.
- If I knew a high school student who was planning to get a driver's license soon, I would recommend he or she take the course.
- The course has increased my confidence in my driving.
- The course will help me be a more cautious driver.
- I think the course is better than lessons from another driving school.
- I think the course has been a good preparation for taking my driver's license test.

#### Part 4: Input on Specific Aspects of Course

1. Ask students to rate each aspect of the course on a scale from 1-5, where 1 represents complete dissatisfaction and 5 represents complete satisfaction, using the following format:

Put an **X** in the box that best describes **how much you are dissatisfied or satisfied** with each program component:

Completely  
1 Dissatisfied

Somewhat  
2 Dissatisfied

Undecided  
3

Somewhat  
4 Satisfied

Completely  
5 Satisfied

#### Questionnaire Items

- Textbook
- Course handouts
- Lectures
- Group work
- Presentations

- Audiovisual materials (specify)
- In-car instruction
- In-car log
- Work load, assignments etc.

2. Provide specific comments you have about any of these aspects of the course.

### Part 5: Things to Change

1. What are the three most important things about the course that you would like to see changed and why?
2. Do you have any suggestions on how to make these changes?

### Part 6: Attitudes

1. Ask students to indicate how much they agree or disagree with the following statements\*, using an agree-disagree scale for each item, where 1 represents completely disagree and 5, completely agree, as follows:

Put an **X** in the box that best describes **how much you disagree or agree** with each statement:

- |   |   |   |  |  |
|---|---|---|--|--|
| <input type="checkbox"/> Completely<br>1 Disagree | <input type="checkbox"/> Somewhat<br>2 Disagree | <input type="checkbox"/> Undecided<br>3 | <input type="checkbox"/> Somewhat<br>4 Agree | <input type="checkbox"/> Completely<br>5 Agree |
|---|---|---|--|--|

- I am confident that I know all the rules of the road.
- I feel like the one place where I am totally in control is in my car.
- I live my life for today rather than worrying about the future.
- Even with all the thousands of cars on the roads, there's a lot I can do by myself to avoid a crash.
- I don't mind taking risks. Otherwise, life is too boring.
- If friends told me to drive faster, I would probably not do so.
- Lots of drivers are careless, and I can't do anything about it if they crash into me.
- It doesn't really matter if I drive recklessly, because I'm still better than most drivers.
- I guess I take more driving risks when I am with my friends, but who doesn't?

---

\*Some of these items are taken from questionnaires developed for the study, *A Longitudinal Analysis of Manitoba Public Insurance Driver Education Program*, Lonero et al. 2005.

- If I was a more cautious driver, some of my friends would laugh at the way I drive.
- Today's cars are built safe and most have airbags, so going faster or cornering fast is OK.

### Part 7: Practice Driving

1. How much practice driving did you do per week while you were taking the course?  
Choices: none, 1 hour/week, 2-3 hours/week, 4-5 hours/week, more than 5 hours/week?
2. With whom did you practice?  
Mother, father, stepmother or stepfather, sibling, relative, friend?

### Part 8: Getting Licensed

1. Please complete the attached postcard (with pre-paid postage), and return it to us after you have completed your driver's license test.

Sample Postcard

<p>To: (Name of School)</p> <p>From: (Name of Student)</p> <p>Address:</p> <p>I took my driver's license test on _____ (D/M/Y)</p> <p>Circle the appropriate answers for the following questions:  The result of my test was:    Passed    Failed  If you failed the test, do you intend to retake it?    Yes    No</p>
---

## Example of Mail-out Survey for Driver Education Program Graduates

### Driving Questionnaire for Graduates of Manitoba's High School Driver Education Program (DEP)

Your answers on this survey are strictly confidential. Under no circumstances will any law enforcement, insurance, driver licensing, or other organization be able to identify which answers or other information are yours.

#### Section 1—General Questions

- 1.1 What is your sex?    1 Female    2 Male  
*(circle the code number of the answer that is correct for you)*
- 1.2 Your age: \_\_\_\_\_ years
- 1.3 Where do you do **most** of your driving? *(circle just one code number)*  
1.....Winnipeg  
2.....Brandon  
3.....Another city with a population **over** 5,000  
4.....Rural area or a town with a population of 5,000 or **less**
- 1.4 How many times did you take the provincial road test before receiving your driver's licence?  
*(circle the number of times)*    1   2   3   4   5 or more
- 1.5 In what **year** did you take the Driver Education Program?    19\_\_\_\_
- 1.6 How **old** were you when you **first** got your driver's licence (not learner's permit)?  
\_\_\_\_\_ years
- 1.7 What is the **highest** level of school **you** have reached?  
*(circle the code number next to the best answer for you)*  
1.....Some high school or less  
2.....High school graduation  
3.....Some university or college courses  
4.....Community college graduate  
5.....University graduate

**1.8** What is the **highest** level reached by your **parent** (or guardian) who went the furthest in school?

*(circle the code number next to the best answer for your parent/guardian)*

- 1.....Some high school or less
- 2.....High school graduate
- 3.....Some university or college courses
- 4.....Community college graduate
- 5.....University graduate

## Section 2—Frequency and Purpose of Driving

We would like to know how much, and why, you drove in your **first year as a licensed driver (2.1)**, and **how much, and why, you drive now (2.2)**.

For each *Purpose of Trip and Driving Environment* listed below, circle the code number in the box **under the words** that best show **how often** you would make a trip where that was the main reason or main environment.

**2.1** In your first year of driving, how often did you drive? *(circle one code number in each row)*

	Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
<b><u>Purpose of Trip</u></b>					
1. To and from school & activities?	5	4	3	2	1
2. To and from work?	5	4	3	2	1
3. As <b>part of</b> your work?	5	4	3	2	1
4. For errands or shopping?	5	4	3	2	1
5. Just for fun or something to do?	5	4	3	2	1
<b><u>Driving Environment</u></b>					
6. During rush hours?	5	4	3	2	1
7. In a town?	5	4	3	2	1
8. In a city?	5	4	3	2	1
9. On gravel roads?	5	4	3	2	1
10. At night after 10 p.m.?	5	4	3	2	1
11. With passengers in your vehicle?	5	4	3	2	1
12. On major highways?	5	4	3	2	1

- 2.2** If you have been licensed for more than 1 year, **how often** do you drive **now**?  
(circle one code number in each row)

	Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
<b><u>Purpose of Trip</u></b>					
1. To and from school & activities?	5	4	3	2	1
2. To and from work?	5	4	3	2	1
3. As <b>part of</b> your work?	5	4	3	2	1
4. For errands or shopping?	5	4	3	2	1
5. Just for fun or something to do?	5	4	3	2	1
<b><u>Driving Environment</u></b>					
6. During rush hours?	5	4	3	2	1
7. In a town?	5	4	3	2	1
8. In a city?	5	4	3	2	1
9. On gravel roads?	5	4	3	2	1
10. At night after 10 p.m?	5	4	3	2	1
11. With passengers in your vehicle?	5	4	3	2	1
12. On major highways?	5	4	3	2	1

### Section 3—Accidents and Traffic Tickets

We would like to know a bit about any accidents and traffic tickets that you may have had.

- 3.1** How many accidents have you been involved in as a driver since you received your licence? Put the number of accidents in the row that matches the damage that occurred. (For example, an accident that caused injury would be counted in the first row, even though there might have been vehicle damage as well. Enter 0 if you had no accidents of that type.)

Type of accident	# of accidents
1. Accidents with injury to one or more people requiring medical attention (may also involve vehicle damage)	
2. Single vehicle accident with no injuries but damage to vehicle	
3. Multi-Vehicle accident with no injuries but damage to vehicles	

**3.2** Whose fault was your most recent accident (where **fault** has been determined)?

*(circle the code number next to the best answer)*

7.....All mine

6.....Mostly mine

5.....About 50/50

4.....Mostly the other driver's

3.....All the other driver's

2.....Don't know, or fault not determined

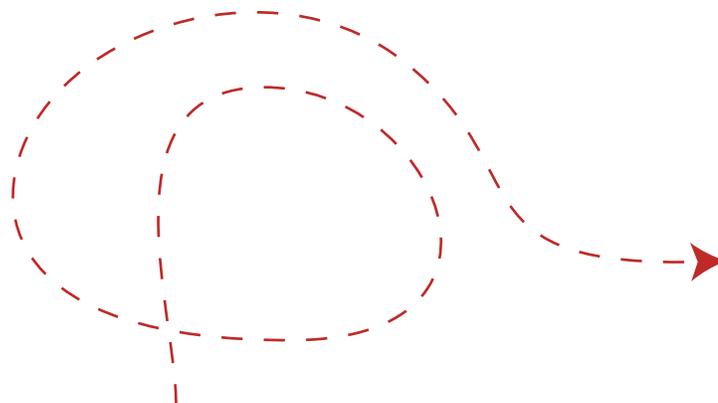
1.....No accidents

**3.3** How many accidents, even minor ones, have you had that were not reported to Autopac?

*(circle the number of unreported accidents) 0 1 2 3 4 or more*

**3.4** **How many times** have you been stopped or ticketed by the police for each of the following kinds of traffic violations since you received your licence? *(Please enter the correct number in the box for each type of violation. If none enter 0.)*

Type of violation	# times
1. Speeding	
2. Failed to stop at stop sign	
3. Imprudent driving	
4. Failed to wear seat belt	
5. Ran red light	
6. Unsafe passing	
7. Impaired driving	
8. Failed to yield	
9. All Others	



## Section 4—Driver’s Opinion of the DEP Course

We would like to know what you think of Manitoba’s high school Driver Education Program (DEP).

- 4.1** Please indicate your feelings about the following statements about driver training. (circle the code number in the box under the heading that shows how much you **agree or disagree**)

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
1. I think Manitoba’s high school Driver Education Program (DEP) is valuable for training new drivers.	5	4	3	2	1
2. I would be as good a driver as I am even if I hadn’t taken the DEP course.	5	4	3	2	1
3. I think that young drivers who take the DEP course are more skilled than young drivers who do not take the course.	5	4	3	2	1
4. If I had not taken the DEP course, I think I would have had more accidents by now.	5	4	3	2	1
5. If I knew a high school student who was planning to get a driver’s licence soon, I would recommend he or she take the DEP course.	5	4	3	2	1
6. The DEP course increased my confidence in my driving.	5	4	3	2	1
7. The DEP course helped me be a more cautious driver.	5	4	3	2	1
8. I think the DEP course is better than lessons from a private driving school.	5	4	3	2	1

- 4.2** Now that you have been driving for some time, think back to the DEP course. What things do you think the course should have spent **more** time on? (circle code numbers for up to five items that you think would have been most helpful to you, and write in any others that you would like to mention)

- 1.....Anticipating road and traffic hazards
- 2.....Turning
- 3.....Lane changing
- 4.....Observing the actions of other drivers
- 5.....How to avoid driving after drinking
- 6.....How to stop friends driving after drinking
- 7.....Winter driving

- 8.....Highway driving
- 9.....Gravel road driving
- 10.....Driving in heavy traffic
- 11.....Laws and regulations
- 12.....What to expect from other drivers
- 13.....Handling a car in an emergency
- 14.....Other (specify below)

## Section 5—Attitudes

We would like to know a little about your attitudes and feelings towards driving and some other matters affecting your health and safety.

**5.1** Please indicate your feelings about the statements listed in the left column. (circle the code number in the box under the heading that shows how much you **agree or disagree**)

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
1. I am confident that I know all the rules of the road.	5	4	3	2	1
2. I feel like the one place where I am totally in control is in my car.	5	4	3	2	1
3. I live my life for today rather than worrying about the future.	5	4	3	2	1
4. Even with all the thousands of cars on the roads, there's a lot I can do by myself to avoid an accident.	5	4	3	2	1
5. I like the idea of volunteering my time to help other people.	5	4	3	2	1
6. I don't mind taking risks. Otherwise, life is too boring.	5	4	3	2	1
7. I think a lot about what I am going to do in the future.	5	4	3	2	1
8. If friends told me to drive faster, I would probably <b>not</b> do so.	5	4	3	2	1
9. I choose not to smoke.	5	4	3	2	1
10. Lots of drivers are careless, and I can't do anything about it if they crash into me.	5	4	3	2	1
11. It doesn't really matter if I drive recklessly, because I'm still better than most drivers.	5	4	3	2	1
12. If we just do whatever our bosses, parents, or authorities tell us, we will go through life like robots, never fully enjoying things.	5	4	3	2	1
13. I try to get some exercise a few times a week.	5	4	3	2	1
14. I guess I take more driving risks when I am with my friends, but who doesn't?	5	4	3	2	1
15. As an adult, I can make my own decisions, even if they involve risk.	5	4	3	2	1
16. If I were a more cautious driver, some of my friends would laugh at the way I drive.	5	4	3	2	1
17. Today's cars are built safe and most have air-bags, so going faster or cornering fast is OK.	5	4	3	2	1

## Section 6— Driver Skills

We would like to know how you feel about your driving skills, compared to the average driver’s skills.

**6.1** Please estimate **how well** you usually perform each of the driving skills listed in the left column. *(circle the code number in the box under the words that best describe you)*

Skill	Well Above Average	Somewhat Above Average	Average	Somewhat Below Average	Well Below Average
1. Anticipating hazards	5	4	3	2	1
2. Driving in a strange city	5	4	3	2	1
3. Obeying the traffic rules	5	4	3	2	1
4. Preventing a skid	5	4	3	2	1
5. Predicting traffic situations ahead	5	4	3	2	1
6. Driving cautiously	5	4	3	2	1
7. Smooth lane-changing in heavy traffic	5	4	3	2	1
8. Fast reactions	5	4	3	2	1
9. Paying attention to other road users	5	4	3	2	1
10. Driving fast, if necessary	5	4	3	2	1
11. Driving in the dark	5	4	3	2	1
12. Controlling the vehicle in a skid	5	4	3	2	1
13. Avoiding “tailgating” other vehicles	5	4	3	2	1
14. Adjusting your speed to the conditions	5	4	3	2	1
15. Passing on a 2-lane road	5	4	3	2	1
16. Giving other drivers the right-of-way	5	4	3	2	1
17. Obeying the speed limits	5	4	3	2	1
18. Avoiding unnecessary risks	5	4	3	2	1
19. Tolerating other drivers’ mistakes calmly	5	4	3	2	1

## Section 7—Driving Errors and Violations

No one is perfect; even the best drivers make mistakes. We would like to know how often, if ever, you have committed the following errors or violations during the last year.

**7.1** Please indicate how often you commit the errors and violations listed in the left column. *(circle the code number in the box under the heading that shows how much you **agree or disagree**)*

<b>Errors and Violations</b>	Never	Hardly Ever	Some-times	Quite Often	Fre- quently
1. In the wrong lane when approaching an intersection	5	4	3	2	1
2. Drive through a yellow traffic signal, even if you could safely stop	5	4	3	2	1
3. Miss a traffic control, such as a stop sign	5	4	3	2	1
4. Fail to notice pedestrians crossing when starting to turn	5	4	3	2	1
5. "Tailgate" the car in front to make its driver speed up or get out of your way	5	4	3	2	1
6. Drive 10 km/h or more above the speed limit	5	4	3	2	1
7. Accidentally hit something when backing up	5	4	3	2	1
8. Cross an intersection knowing that the traffic light has already turned red	5	4	3	2	1
9. Forget to cancel turn signals after changing lanes	5	4	3	2	1
10. Fail to check mirrors before moving into a new lane	5	4	3	2	1
11. Dislike a particular type of driver and do something to show your hostility	5	4	3	2	1
12. Underestimate the speed of an oncoming vehicle when turning or passing	5	4	3	2	1
13. Switch on one thing, such as the wipers, when you meant to switch on something else, such as turn signals	5	4	3	2	1
14. Drive when you may be over the legal blood-alcohol limit	5	4	3	2	1
15. Get involved in unofficial "races" with other drivers	5	4	3	2	1
16. Go after another driver to show your anger	5	4	3	2	1
17. Drive too fast for road conditions or weather conditions	5	4	3	2	1
18. Drive without wearing your seat belt	5	4	3	2	1
19. Leave too late, and find yourself speeding to get to your destination in time	5	4	3	2	1

**Thanks for helping to improve Manitoba's driver education!**

## Example of Mail-out Survey for Newly Licensed Drivers

### Driving Questionnaire for Manitoba Drivers

Your answers on this survey are strictly confidential. Under no circumstances will any law enforcement, insurance, driver licensing, or other organization be able to identify which answers or other information are yours.

#### Section 1—General Questions

- 1.1 What is your sex?    1 Female    2 Male  
*(circle the code number of the answer that is correct for you)*
- 1.2 Your age: \_\_\_\_\_ years
- 1.3 Where do you do **most** of your driving? *(circle just one code number, next to the best answer)*  
1.....Winnipeg  
2.....Brandon  
3.....Another city with a population **over** 5,000  
4.....Rural area or a town with a population of 5,000 or **less**
- 1.4 How many times did you take the provincial road test before receiving your driver's licence?  
*(circle the number of times)    1   2   3   4   5 or more*
- 1.5 How **old** were you when you **first** got your driver's licence (not learner's permit)? \_\_\_\_ years
- 1.6 What is the **highest** level of school **you** have reached?  
*(circle the code number next to the best answer for you)*  
1.....Some high school or less  
2.....High school graduation  
3.....Some university or college courses  
4.....Community college graduate  
5.....University graduate
- 1.7 What is the **highest** level reached by your **parent** (or guardian) who went the furthest in school?  
*(circle the code number next to the best answer for your parent/guardian)*  
1.....Some high school or less  
2.....High school graduation  
3.....Some university or college courses  
4.....Community college graduate  
5.....University graduate

- 1.8** Did you ever take any formal driver training with a professional instructor in Manitoba?  
*(circle the code number next to the best answer)*  
 1.....No (Go to Section 2)  
 2.....Yes, at a private driving school  
 3.....Yes, I took high school Driver Education but did not complete it

**1.9** In what year did you take that driver training? 19\_\_\_\_\_

**1.10** During that training, about how many hours did you spend at the following activities, including test time. *(If none, enter 0.)*

In-car time with a professional instructor	hours
Practice driving with family or friends	hours
Classroom time	hours

## Section 2 – Frequency and Purpose of Driving

We would like to know how much, and why, you drove in your **first year as a licensed driver (2.1)**, and **how much, and why, you drive now (2.2)**.

For each Purpose of Trip and Driving Environment listed below, circle the code number in the box **under the words** that best show **how often** you would make a trip where that was the main reason or main environment.

**2.1** In your **first year** of driving, how often did you drive? *(circle one code number in each row)*

	Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
<b><u>Purpose of Trip</u></b>					
1. To and from school & activities?	5	4	3	2	1
2. To and from work?	5	4	3	2	1
3. As <b>part of</b> your work?	5	4	3	2	1
4. For errands or shopping?	5	4	3	2	1
5. Just for fun or something to do?	5	4	3	2	1
<b><u>Driving Environment</u></b>					
6. During rush hours?	5	4	3	2	1
7. In a town?	5	4	3	2	1

	Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
8. In a city?	5	4	3	2	1
9. On gravel roads?	5	4	3	2	1
10. At night after 10 p.m?	5	4	3	2	1
11. With passengers in your vehicle?	5	4	3	2	1
12. On major highways?	5	4	3	2	1

**2.2** If you have been licensed for more than 1 year, **how often** do you drive **now**?  
(circle one code number in each row)

	Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
<b><u>Purpose of Trip</u></b>					
1. To and from school & activities?	5	4	3	2	1
2. To and from work?	5	4	3	2	1
3. As <b>part of</b> your work?	5	4	3	2	1
4. For errands or shopping?	5	4	3	2	1
5. Just for fun or something to do?	5	4	3	2	1
<b><u>Driving Environment</u></b>					
6. During rush hours?	5	4	3	2	1
7. In a town?	5	4	3	2	1
8. In a city?	5	4	3	2	1
9. On gravel roads?	5	4	3	2	1
10. At night after 10 p.m?	5	4	3	2	1
11. With passengers in your vehicle?	5	4	3	2	1
12. On major highways?	5	4	3	2	1

### Section 3 – Accidents and Traffic Tickets

We would like to know a bit about any accidents and traffic tickets that you may have had.

- 3.1** How many accidents have you been involved in as a driver since you received your licence? Put the number of accidents in the row that matches the damage that occurred. (For example, an accident that caused injury would be counted in the first row, even though there might have been vehicle damage as well. Enter 0 if you had no accidents of that type.)

Type of accident	# of accidents
1. Accidents with injury to one or more people requiring medical attention (may also involve vehicle damage)	
2. Single vehicle accident with no injuries but damage to vehicle	
3. Multi-Vehicle accident with no injuries but damage to vehicles	

- 3.2** Whose **fault** was your most recent accident (where fault has been determined)?  
(circle the code number next to the best answer)

- 7.....All mine  
6.....Mostly mine  
5.....About 50/50  
4.....Mostly the other driver's  
3.....All the other driver's  
2.....Don't know, or fault not determined  
1.....No accidents

- 3.3** How many accidents, even minor ones, have you had that were not reported to Autopac?  
(circle the number of unreported accidents)    0   1   2   3   4 or more

- 3.4** **How many times** have you been stopped or ticketed by the police for each of the following kinds of traffic violations since you received your licence? (Please enter the correct number in the box for each type of violation. If none enter 0.)

Type of violation	# times
1. Speeding	
2. Failed to stop at stop sign	
3. Imprudent driving	
4. Failed to wear seat belt	
5. Ran red light	
6. Unsafe passing	
7. Impaired driving	
8. Failed to yield	
9. All Others	

## Section 4 – Driver’s Opinion of Driver Education Courses

We would like to know what you think of Manitoba’s high school Driver Education Program (DEP), and of private driver education courses. If you did not take any professional driver education or training yourself, please **skip to question 4.2**.

**4.1** Please indicate your feelings about the following statements about your driver training. *(circle the code number in the box under the heading that shows how much you agree or disagree)*

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
1. If I had not taken any driver training, I think I would have had more accidents.	5	4	3	2	1
2. I would be as good a driver as I am even if I hadn’t taken any training.	5	4	3	2	1
3. Driver training helped me be a more cautious driver.	5	4	3	2	1
4. Driver training increased my driving confidence.	5	4	3	2	1

**4.2** Whose decision was mainly responsible for **you not taking** the high school Driver Education Program before you got your driver’s licence? *(circle the code number next to the best answer)*

- 1.....It was mainly my parent or guardian’s decision.
- 2.....It was mainly my own decision.

**4.3** What was the main reason for not taking the Manitoba high school Driver Education Program? *(circle one or two main reasons, and add others on lines below)*

- 1.....Not in a Manitoba high school then
- 2.....Not convenient
- 3.....Couldn’t afford the cost
- 4.....Couldn’t afford the time
- 5.....Took other lessons
- 6.....Not available in my school
- 7.....Other (specify)

---



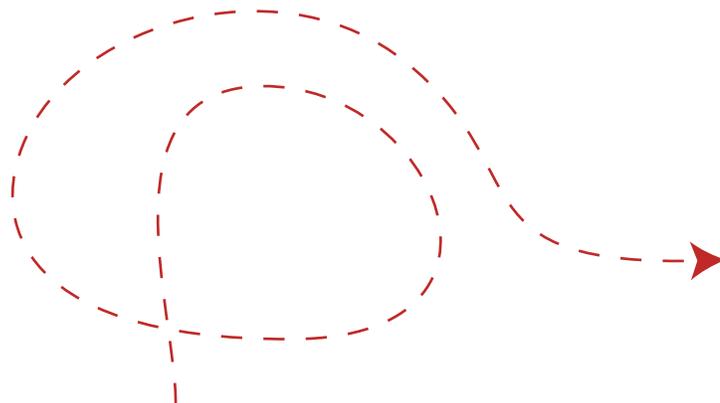
---



---

**4.4** Please indicate your feelings about Manitoba’s high school Driver Education Program (DEP).  
*(circle the code number in the box under the heading that shows how much you **agree or disagree**)*

<b>Statement</b>	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
1. I think Manitoba’s high school Driver Education Program (DEP) is valuable for training new drivers.	5	4	3	2	1
2. I think that young drivers who take the DEP course are more skilled than other young drivers who did not take the course.	5	4	3	2	1
3. If I knew a high school student who was planning to get a driver’s licence soon, I would recommend he or she take the DEP course.	5	4	3	2	1
4. I think the DEP course is better than lessons from a private driving school.	5	4	3	2	1
5. I wish that I had taken the DEP course.	5	4	3	2	1



## Section 5 – Attitudes

We would like to know a little about your attitudes and feelings towards driving and some other matters affecting your health and safety.

- 5.1** Please indicate your feelings about the statements listed in the left column.  
(circle the code number in the box under the heading that shows how much you **agree or disagree**)

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
1. I am confident that I know all the rules of the road.	5	4	3	2	1
2. I feel like the one place where I am totally in control is in my car.	5	4	3	2	1
3. I live my life for today rather than worrying about the future.	5	4	3	2	1
4. Even with all the thousands of cars on the roads, there's a lot I can do by myself to avoid an accident.	5	4	3	2	1
5. I like the idea of volunteering my time to help other people.	5	4	3	2	1
6. I don't mind taking risks. Otherwise, life is too boring.	5	4	3	2	1
7. I think a lot about what I am going to do in the future.	5	4	3	2	1
8. If friends told me to drive faster, I would probably <b>not</b> do so.	5	4	3	2	1
9. I choose not to smoke.	5	4	3	2	1
10. Lots of drivers are careless, and I can't do anything about it if they crash into me.	5	4	3	2	1
11. It doesn't really matter if I drive recklessly, because I'm still better than most drivers.	5	4	3	2	1
12. If we just do whatever our bosses, parents, or authorities tell us, we will go through life like robots, never fully enjoying things.	5	4	3	2	1
13. I try to get some exercise a few times a week.	5	4	3	2	1
14. I guess I take more driving risks when I am with my friends, but who doesn't?	5	4	3	2	1
15. As an adult, I can make my own decisions, even if they involve risk.	5	4	3	2	1
16. If I were a more cautious driver, some of my friends would laugh at the way I drive.	5	4	3	2	1
17. Today's cars are built safe and most have air-bags, so going faster or cornering fast is OK.	5	4	3	2	1

## Section 6 – Driver Skills

We would like to know how you feel about your driving skills, compared to the average driver’s skills.

**6.1** Please estimate **how well** you usually perform each of the driving skills listed in the left column. *(circle the code number in the box under the words that best describe you)*

Skill	Well Above Average	Somewhat Above Average	Average	Somewhat Below Average	Well Below Average
1. Anticipating hazards	5	4	3	2	1
2. Driving in a strange city	5	4	3	2	1
3. Obeying the traffic rules	5	4	3	2	1
4. Preventing a skid	5	4	3	2	1
5. Predicting traffic situations ahead	5	4	3	2	1
6. Driving cautiously	5	4	3	2	1
7. Smooth lane-changing in heavy traffic	5	4	3	2	1
8. Fast reactions	5	4	3	2	1
9. Paying attention to other road users	5	4	3	2	1
10. Driving fast, if necessary	5	4	3	2	1
11. Driving in the dark	5	4	3	2	1
12. Controlling the vehicle in a skid	5	4	3	2	1
13. Avoiding “tailgating” other vehicles	5	4	3	2	1
14. Adjusting your speed to the conditions	5	4	3	2	1
15. Passing on a 2-lane road	5	4	3	2	1
16. Giving other drivers the right-of-way	5	4	3	2	1
17. Obeying the speed limits	5	4	3	2	1
18. Avoiding unnecessary risks	5	4	3	2	1
19. Tolerating other drivers’ mistakes calmly	5	4	3	2	1

## Section 7 – Driving Errors and Violations

No one is perfect; even the best drivers make mistakes. We would like to know how often, if ever, you have committed the following errors or violations during the last year.

**7.1** Please indicate how often you commit the errors and violations listed in the left column. (circle the code number in the box under the heading that shows how much you **agree or disagree**)

<b>Errors and Violations</b>	Never	Hardly Ever	Some-times	Quite Often	Fre-quently
1. In the wrong lane when approaching an intersection	5	4	3	2	1
2. Drive through a yellow traffic signal, even if you could safely stop	5	4	3	2	1
3. Miss a traffic control, such as a stop sign	5	4	3	2	1
4. Fail to notice pedestrians crossing when starting to turn	5	4	3	2	1
5. "Tailgate" the car in front to make its driver speed up or get out of your way	5	4	3	2	1
6. Drive 10 km/h or more above the speed limit	5	4	3	2	1
7. Accidentally hit something when backing up	5	4	3	2	1
8. Cross an intersection knowing that the traffic light has already turned red	5	4	3	2	1
9. Forget to cancel turn signals after changing lanes	5	4	3	2	1
10. Fail to check mirrors before moving into a new lane	5	4	3	2	1
11. Dislike a particular type of driver and do something to show your hostility	5	4	3	2	1
12. Underestimate the speed of an oncoming vehicle when turning or passing	5	4	3	2	1
13. Switch on one thing, such as the wipers, when you meant to switch on something else, such as turn signals	5	4	3	2	1
14. Drive when you may be over the legal blood-alcohol limit	5	4	3	2	1
15. Get involved in unofficial "races" with other drivers	5	4	3	2	1
16. Go after another driver to show your anger	5	4	3	2	1
17. Drive too fast for road conditions or weather conditions	5	4	3	2	1
18. Drive without wearing your seat belt	5	4	3	2	1
19. Leave too late, and find yourself speeding to get to your destination in time	5	4	3	2	1

**Thanks for helping to improve Manitoba's driver education!**



## Example of Mail-out Survey for Parents of Driver Education Program Graduates

### MPI Survey of Parents of Driver Education Program Graduates

Your answers on this survey are strictly confidential. Under no circumstances will any law enforcement, insurance, driver licensing, or other organization be able to identify which answers or other information are yours.

1. What is your sex? 1 Female 2 Male (circle the code number next to the right answer)
2. Your age: \_\_\_\_\_ years
3. Whose decision was mainly responsible for your teenager taking the Manitoba high school Driver Education Program? (circle the code number next to the best answer)
  - 1.....It was primarily a parent’s decision.
  - 2.....It was primarily the teenager’s decision.
4. When your teenager was learning to drive, how much time would you say was spent practice driving with a parent, other family member, or friend? (circle the code number next to your best estimate)
  - 1.....None
  - 2.....1 to 10 hours
  - 3.....11 to 25 hours
  - 4.....26 to 50 hours
  - 5.....over 50 hours
5. Please indicate your feelings about Manitoba’s high school Driver Education Program (DEP). (circle the code number in the box under the heading that shows how much you **agree or disagree**)

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
I think Manitoba’s high school Driver Education Program (DEP) is valuable for training new drivers.	5	4	3	2	1
I believe my child thinks the DEP course is valuable for training new drivers.	5	4	3	2	1
I believe the general public thinks the DEP course is valuable for training new drivers.	5	4	3	2	1
I think that young drivers who take the DEP course are more skilled than young drivers who do not take the course.	5	4	3	2	1
If I knew a high school student who was planning to get a driver’s licence soon, I would recommend he or she take the DEP course.	5	4	3	2	1

# Example of Mail-out Survey of General Public Attitudes Towards Driver Education

## MPI Survey of Public Attitudes to Driver Education

Your answers on this survey are strictly confidential. Under no circumstances will any law enforcement, insurance, driver licensing, or other organization be able to identify which answers or other information are yours.

1. What is your sex? 1 Female 2 Male (circle the code number next to the right answer)
  
2. Your age: \_\_\_\_\_ years
  
3. Did you ever take any formal driver training with a professional teacher or instructor?  
(circle the code number next to the best answer)
  - 1.....No
  - 2.....Yes, at a private driving school
  - 3.....Yes, I took high school driver education
  
4. Please indicate your feelings about Manitoba's high school Driver Education Program (DEP).  
(circle the code number in the box under the heading that shows how much you **agree or disagree**):

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
I think Manitoba's high school Driver Education Program (DEP) is valuable for training new drivers.	5	4	3	2	1
I believe the general public thinks the DEP course is valuable for training new drivers.	5	4	3	2	1
I think that young drivers who take the DEP course are more skilled than young drivers who do not take the course.	5	4	3	2	1
If I knew a high school student who was planning to get a driver's licence soon, I would recommend he or she take the DEP course.	5	4	3	2	1



## Example of Non-Response Mail-out Survey for Newly Licensed Drivers

### Driving Questionnaire for Manitoba Drivers (Reminder)

**Dear Driver:**

We recently sent you a questionnaire about your driving experience and opinions. If you have already sent it in, many thanks and please disregard this. If you have not filled out and returned a questionnaire yet, it is not too late to help this important research! We would greatly appreciate the time and effort you would spend filling out the questionnaire and dropping it in a mailbox. If you do **not** have the original questionnaire at hand, please help out by answering the following few questions and returning this questionnaire in the postage-paid envelope provided. Your information is very important to our research and will be very much appreciated. If you wish to verify that this survey is a legitimate MPI project, please call MPI Strategic Research at 204-985-7594.

Sincerely,

**All of your answers on this survey are strictly confidential. Under no circumstances will any law enforcement, insurance, driver licensing, or other organization be able to identify which answers or other information are yours.**

Please **circle** the *number* of the answer that is correct for you.

1. What is your sex?    1 Female    2 Male
  
2. Your age: \_\_\_\_\_ years
  
3. What type of driver's licence do you have now?  
1 None    2 Learner's Permit    3 Full Licence (Class 1-6)
  
4. Did you ever take any formal driver training with a professional instructor in Manitoba?  
(circle the code number next to the best answer)  
1.....No  
2.....Yes, at a private driving school  
3.....Yes, I completed high school Driver Education  
4.....Yes, I took high school Driver Education but did not complete it
  
5. Where do you do most of your driving? (circle just one number)  
1.....Winnipeg  
2.....Brandon  
3.....Another city with a population **of more than 5,000**  
4.....Rural area or a town with a population of 5,000 or **fewer**

**6. How often do you drive?**

Please circle the number under the words that best indicate how often you usually drive.

Never, or almost never	Less than once a month	Once a month to a few times a month	Once a week to a few times each week	At least every workday
5	4	3	2	1

**7. How many accidents have you been involved in as a driver since you received your licence?**

Put the number of accidents in the row that matches the damage that occurred. (For example, an accident in which someone was injured would be counted in the first row, even though one or more of the vehicles may have been damaged as well. If none, enter 0)

Type of accident	# of accidents
1. Accidents with injury to one or more people requiring medical attention (may also involve vehicle damage)	
2. Single vehicle accident with no injuries but damage to vehicle	
3. Multi-Vehicle accident with no injuries but damage to vehicles	

**8. How many accidents, even minor ones, have you had that were **not** reported to Autopac?**

Circle the number: 0 1 2 3 4 or more

**9. We would like to know what you think of Manitoba’s high school Driver Education Program (DEP), regardless of whether you took it yourself or simply heard about it from friends or relatives.**

Please circle the number in the box under the words that show how much you **agree or disagree** with the following statements about the Driver Education Program (DEP).

Statement	Strongly Agree	Agree	Un-decided	Disagree	Strongly Disagree
I think Manitoba’s high school Driver Education Program (DEP) is valuable for training new drivers.	5	4	3	2	1
I think the DEP course is better than lessons from a private driving school.	5	4	3	2	1

**10. Finally, to help us improve future questionnaires and other research, please tell us why you did not complete the original, full version of the questionnaire we sent you.**

- Please circle the answer that most closely applies.
- |                                      |  |
|--------------------------------------|--|
| 1 ..... I forgot about it            | 5 ..... I did not trust the confidentiality of the whole process |
| 2 ..... I dislike all surveys        | 6 ..... I meant to do it, but never got to it                    |
| 3 ..... I misplaced it               | 7 ..... Other (please specify)                                   |
| 4 ..... It looked too time consuming |  |

**Thank you very much for helping to improve Manitoba’s driver education!**



# APPENDIX H: Attitude and Risk-Taking Questionnaires



## 1. Driving Style Questionnaire

Adapted by the Traffic Injury Research Foundation (TIRF) from:

Donovan, D. M. 1980. *Drinking behavior, personality factors and high-risk driving*. Unpublished doctoral dissertation, University of Washington.

## 2. Personal Style: Thrill and Adventure Seeking Questionnaire

Zuckerman, M. 1979. *Sensation seeking. Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.

## 3. A Modified Version of Jessor's Risky Driving Scale

Adapted by TIRF from:

Jessor, R. 1987. Risky driving and adolescent problem behavior: An extension of problem-behavior theory. *Alcohol, Drugs and Driving* 3(3/4), 1-11.



## Attitude and Risk-Taking Questionnaires

### 1. Driving Style

For each of the following, indicate whether it is TRUE or generally characteristic of you, or FALSE, if it is not characteristic of you.

- T F 1. I get annoyed if the traffic lights change to red as I approach them.
- T F 2. I find driving a form of relaxation when I feel tense.
- T F 3. I swear under my breath at other drivers.
- T F 4. If the driver behind me has his lights shining in my mirror, I pay him back in some way.
- T F 5. I sometimes swear out loud at other drivers.
- T F 6. I have never given chase to a driver who has annoyed me.
- T F 7. When I am upset, driving helps soothe my nerves.
- T F 8. I find it difficult to control my temper when driving.
- T F 9. It's fun to beat other drivers at the getaway.
- T F 10. I often use my horn when I get annoyed at other drivers.
- T F 11. It's fun to pass other cars on the highway even if you are not in a hurry.
- T F 12. I am easily provoked or angered when driving.
- T F 13. During the past few months, at least once I have gone driving to "blow off steam" after an argument.
- T F 14. Driving at high speeds gives one a thrilling sense of power.
- T F 15. Driving at high speeds is exciting.
- T F 16. It's fun to outwit other drivers.
- T F 17. Driving helps me to forget about pressures.
- T F 18. Sometimes I take a risk when driving just for the sake of the risk.
- T F 19. At times, I've felt I could gladly kill another driver.
- T F 20. It's a thrill to beat other drivers at the getaway.
- T F 21. There is something about being behind the wheel of a car that makes one feel bigger.
- T F 22. It's fun to maneuver and weave through traffic.
- T F 23. I often make rude signs at other motorists who annoy me.
- T F 24. I lose my temper when another driver does something stupid.
- T F 25. I have been known to flash my car lights at others in anger.

#### Scoring:

Driving Aggression: #s 1, 3, 4, 5, 6 (False), 8, 10, 12, 18, 19, 23, 24, 25.

Competitive Speed: #s 9, 11, 15, 16, 20, 22.

Driving for Tension

Reduction: #s 2, 7, 13, 14, 17, 21.

## 2. Personal Style: Thrill and Adventure Seeking

This questionnaire is about your interests and things you like. Each item below contains two choices, A and B. Please circle either A or B which best describes **your likes** or the way **you feel**. In some cases you may find items in which both choices describe your likes or the way you feel. In these cases you are still to choose only **one**—the one which better describes your likes or feelings. In some cases you may find items in which you do not like either choice. Again, choose only one—choose the one you dislike the least.

It is important that you respond to every item with **only one choice**, A or B. We are interested only in **your** likes or feelings, not in how other feel about these things or how one is supposed to feel. There are no right or wrong answers.

1. A. I often wish I could be a mountain climber.  
B. I can't understand people who risk their necks climbing mountains.
2. A. A sensible person avoids activities that are dangerous.  
B. I sometimes like to do things that are a little frightening.
3. A. I would like to take up the sport of water-skiing.  
B. I would not like to take up water-skiing.
4. A. I would like to try surf-board riding.  
B. I would not like try surf-board riding.
5. A. I would not like to learn to fly an airplane.  
B. I would like to learn to fly an airplane.
6. A. I prefer the surface of the water to the depths.  
B. I would like to go scuba diving.
7. A. I would like to try parachute jumping.  
B. I would never want to try jumping out of a plane with or without a parachute.
8. A. I like to dive off the high board.  
B. I don't like the feeling I get standing on the high board (or I don't go near it at all).
9. A. Sailing long distances in small sailing crafts is foolhardy.  
B. I would like to sail a long distance in a small but seaworthy sailing craft.
10. A. Skiing fast down a high mountain slope is a good way to end up on crutches.  
B. I think I would enjoy the sensations of skiing very fast down a high mountain slope.

**Score:** \_\_\_\_\_

1A            5B            9B  
2B            6B            10B  
3A            7A  
4A            8A

### **3. A Modified Version of Jessor’s Risky Driving Scale**

In the past 12 months, how often have you ...

(1 Never    2 Rarely    3 Occasionally    4 Fairly often    5 Often)

1. Had to steer or brake sharply to avoid a collision? \_\_\_\_
2. Taken some risks while driving because it makes driving more fun? \_\_\_\_
3. Driven after having one or two drinks? \_\_\_\_
4. Driven after having more than two drinks? \_\_\_\_
5. Driven when you thought you had “too much to drink”? \_\_\_\_
6. Driven after using marijuana or other drugs? \_\_\_\_

## **APPENDIX I: Benchmark Program Standards**

1. American Driver and Traffic Safety Education Association (ADTSEA) Standards for:
  - Classroom and In-car Content Segment I and Segment II
  - Delivery of Driver Education
  - Driver Development Outcomes

Located at: <http://adtsea.iup.edu/adtsea/resources/NationalDriverDevProgram.aspx>

2. National Institute for Driver Behavior (NIDB) Minimum Standards: Driving Behaviors for Risk Prevention

Located at: <http://www.nidb.org/drivingstandardsfrp.html>

3. DSAA Process for Curriculum Review, presented next.



# Road Safety Educators' Association

## PROCESS FOR CURRICULUM REVIEW FOR DRIVING SCHOOL ASSOCIATION OF THE AMERICAS, INC.

### TABLE OF CONTENTS

#### Process for Curriculum Review

Stage One

Stage Two

#### Standards and Criteria for Curriculum Development

Standards for Submission

Criteria

Validating Content

Evaluation

About RSEA

#### Application Form

Reproduced with permission

# Process for Curriculum Review

The review process is in two stages.

## Stage One

1. Submit a proposal for course development. The proposal should include any time requirements, a brief synopsis of the desired learning outcomes, from what research and what orientation.  
Application fee is \$50.00.
2. Approval of the proposal will be negotiated by the registrar, in conjunction with the review committee. At this time, if there are courses that are apparently very close or similar, action will be taken to resolve this issue. Action will take the form of one or more of the following:
  - a) The applicant will be made aware of the possible conflict and be given the choice to:
    - i) meet with the other party developing materials in the area or,
    - ii) make a presentation to the review committee to justify the originality of the concept. (this is to protect course developers from plagiarism and reduce the risk of copyright infringement)
3. RSEA will supply, on a cost recovery basis, the Standards and Criteria in either 'Microsoft Word' or 'Word Perfect' in order to simplify the process for program developers.
4. Once the curriculum is developed it should be submitted with:
  - a) the layouts, handouts, materials and/or texts to be used.
  - b) the standards and criteria properly formatted.
  - c) all pertinent information clearly spelled out according to the kit called "RSEA Curriculum Review Kit".
5. The curriculum is submitted to the Registrar.
6. The curriculum is cleaned (all personal identifiers are removed).
7. The review committee decides which reviewers are appointed, one of whom must be content relevant and at least one approved as a RSEA/DSAA reviewer.
8. A time/cost estimate will be provided based on the curriculum submitted. If the review process is likely to exceed this estimate, the registrar will notify the developer of the approximate additional time involved as soon as it is identified.

The developer may at that time decide to continue with the review or retrieve/withdraw their curriculum.

Upon completion of the review, a report is submitted to the Registrar along with an invoice for services rendered by the reviewers. The reviewers also return to the Registrar, any materials provided by the developer for purposes of the review at this time. The fee for services is to be paid by the applicant prior to the release of the curriculum report.

9. The applicant is then informed of the results of the review and can:
  - a) proceed to administer the pilot upon a satisfactory review, or
  - b) revise and re-submit, or
  - c) ask for 'negotiation status'

NOTE: If the applicant requests 'negotiation status', the reviewers are notified in order to start the negotiation process. A process and fee are negotiated for making changes to the curriculum that reflect the recommendations of the reviewers and changes are enacted in concert between the developer and the reviewers.

## Stage Two

1. Pilot is field tested.
2. Test materials used in generating marks/scores submitted to RSEA are re-submitted.
3. A report on the feasibility of:
  - a) timelines
  - b) evaluation process is submitted. Marks/scores must resemble the normal distribution for the level of course offered. Marks/scores not resembling the normal distribution must be justified.
4. The review committee:
  - a) accredits the course, or
  - b) asks for further documentation.
5. All documentation concerning this process must be kept on file with RSEA/DSAA.
6. The review committee will assign an expiry date once the accreditation process is finalized.
7. Any additional materials added to the course during its life-span should be submitted to RSEA/DSAA to be added to its file.

# Road Safety Educators' Association Standards and Criteria for Curriculum Development

*Working Document authored for*

*The Board of Directors of the Road Safety Educators' Association*

*July, 1993, revisions January 1994, January 1997, October 1999*

## Standards for Submissions

Courses should be submitted with the following included:

1. a title that as clearly as possible represents the material to be included in the course
2. a clearly stated rationale (including a clear conception of the learner, society, and the subject matter as well as a statement of the educational goals)
3. the context of the course (level and subject matter)
4. the content of the course (validation information)
5. a comprehensive list of intended learning outcomes
6. time lines for completion of the course including time for evaluation and submission of the grades
7. personnel involved in the administration of the course and their qualifications for delivery of the course
8. location of where/how the course will be given
9. course format (marrying content with teaching style)
10. the number of proposed candidates (optimum and maximum)
11. how the candidates are to be evaluated, either a letter grade or a percent

For example:

A+	90 - 100
A	85 - 90
A-	80 - 85
B+	77 - 79
B	73 - 76
B-	70 - 72
C+	67 - 69
C	63 - 66
C-	60 - 62
D+	57 - 59
D	53 - 56
D-	50 - 52
F	failure to successfully complete

12. a description of the utility and value of the course for road safety
13. indication of the text or materials to be used
14. a reference section

The course will be evaluated for the coherence and clarity on the above fourteen measures and then compared with the following eighteen RSEA criteria:

1. **Direct Applicability:** the program must deal with issues directly applicable to road safety or road safety education and be supported by research.
2. **Integrative vs Sectoral:** the program must address all three areas of human behaviour relating to road safety, those being cognitive, behavioral and affective.
3. **Ecologically Representative:** the program must address issues that are within the ecological system of the individuals participating in the program.
4. **Temporally Extensive:** the program must have a direction that will allow for growth through the different stages in learning to become an autonomous road user or professional in the road safety educator system.
5. **Ipsative vs Normative:** the program must allow for the individual to measure their progress within themselves as well as against others in such a way so as to allow for this growth.
6. **Modular in Construction:** the program must be constructed in such a way as to allow new material to be added and redundant material to be deleted.
7. **Subjective Saliency:** the material must be important to the individuals taking the course in their understanding of their role as a road user or road safety educator within their community and within society as a whole.
8. **Systematic vs Discreet:** the program must reflect a systems approach utilizing all aspects and agencies in the system that interact to form comprehensive road safety education.
9. **Reflexive:** the program must reflect the problems that instructors/trainers have in the driving and/or teaching task as well as those that all drivers and/or teachers experience. Instructors must be careful to teach only those concepts and skills that they as advanced drivers and educators are both capable of and knowledgeable about.
10. **Responsible:** the program must reflect the most recent and contemporary views of traffic safety, teaching, and social change and incorporate these into any training or teaching elements of an educational program.
11. **Accountable:** the program and its' deliverers must be financially responsible to society and the agencies that support the education delivered. Individuals receiving instruction through this delivery system must have access for complaint and arbitration should the instruction be unsatisfactory.

12. **Evaluation:** the program should have a method of evaluating its effectiveness. Both in how it reaches its target audience and whether the material presented or explored has the desired outcome over time.
13. **Suitability:** the programs must meet the needs of the individuals at what ever cognitive level these individuals are operating. This must be accomplished for the benefit of driver, instructors and/or trainers that have not had sufficient time to expand their repertoire within the field of road safety education.
14. **Sufficient Length:** the program must be sufficient in length to allow time for this material to be absorbed, time for the necessary literature review and readings and time to complete any assignments necessary for evaluation.
15. **Molarity Levels:** the program must address issues at all levels of molarity from the molecular to the molar.
16. **Language:** the program must be taught in one of the official languages and all instructors must speak, read and write one of the official languages. Programs delivered for specific ethnic language groups may be exempted from this provision but only same language participants will receive certification.
17. **Dignity of risk:** the program must be taught in such a way as to encourage individuals "to try" even if the outcome of their efforts is not successful since there is dignity in being given the opportunity to learn from our mistakes.
18. **Autonomy:** the program must allow for the decision of any individual to be respected permitting all participants the opportunity to come to their own conclusions based on the facts and on their personal construct of life.

## Validating Content

This will prove difficult in many instances due to a paucity of research in the road safety education field. Given that this situation exists, it is essential to be as cognizant as possible of the research available. In many instances it may be efficacious to utilize research from other areas of education and/or safety, (injury prevention). Often, research that relates to social change, education, behaviour modification, culture and/or engineering can be supplanted into road safety education.

The two ways to validate information and content to be used in any program are to cite the research supporting the material (scientific validation) or to argue logically for the concepts being taught (philosophical validation). One must be careful to argue from true premises in order to arrive at true conclusions, i.e. valid and true information. Generally this is accomplished by finding out the facts that scientific enquiry has validated, and then extending the scope of the argument.

All information and content to be used in courses must meet one of three criterion to be accepted as valid:

- a) it must be scientifically sound or
- b) logically argued for, or
- c) have reached “critical mass” for acceptance within the field as the most appropriate methodology to date (it should be noted that this criteria will always be subject to change if more relevant information is found). When a more analytical method is suggested it will receive higher standing as being more efficacious since it is also less restrictive.

## Evaluation

There are two major areas that need to be addressed regarding course evaluation:

- a) Is the material appropriate and being utilized by the recipient drivers or instructor/trainers?

This type of question is usually addressed by using a questionnaire format in assessing the efficacy of the delivery system for entry level and/or upgrading programs.

- b) What effect is the upgrading and/or updating of instructors having on the issue of road safety?

This issue is of paramount importance but extremely hard to assess. Usual measures of effectiveness are often inappropriate since, in areas of low probability and high consequence risk, there are too many intervening variables to delineate specific causal factors for reductions or increases in risk at the actuarial level.

Therefore, a more appropriate measure may be to design studies that assess intermediate outcome measures to evaluate the effectiveness of such educational programs. Again, the questionnaire format may be an appropriate tool to measure intermediate outcomes such as socially responsible actions like decreases in impaired driving and increases in occupant restraint use and therefore may be better indicators of the effectiveness of our educational efforts. Further to this, Driver Competency Assessment Protocols Inc. will track all drivers having taken the Driver Competency Assessment (DCA) and use this for evaluation purposes as well.

## Where RSEA Fits In

To maintain registration “in good standing” as a professional road safety educator within the association one will be required to keep abreast of current issues. To do this the association will review and endorse updating and upgrading courses to be delivered by professional educators/trainers to professional instructors/educators/trainers since it is our belief that the expertise lies in our own membership. The members that have developed and prepared to deliver these

programs will be granted permission to do so under the RSEA umbrella. In return the successful candidates of these programs will be credited with completing the “in good standing” annual requirements for registration within the professional association. Candidates will be required to offer these to the registrar to have their registration updated before applying for continuance as a registered member as a professional road safety educator. In this way we will be able to track our progress as a professional association and as a profession.

Individuals will have on their RSEA transcript a list of courses they have participated in and the marks that were granted although only those courses successfully completed will be listed in the RSEA REGISTER and marks may be withheld at the request of the individual. Marks at this level will be letter grades corresponding to the marks on the course.

# Driving School Association of the Americas in cooperation with Road Safety Educators' Association

## Application for Curriculum Review

*Please fill in this form and forward to the Curriculum Review Chair at the address provided.*

*The application fee is \$50.00 and should accompany submission of the Application.*

*Typically a full curriculum review costs between one and three thousand dollars depending on the sophistication and length of the program. If other work is contracted in order to meet the criteria additional costs are borne by the applicant.*

*You will be contacted once the reviewer has prepared a cost estimate.*

Developers Name

Address

Contact Information

Date of Submission

Title of the Program

Developer of the Program

Target Population of this Program

List the Performance Outcome Measures for Successful Completion of this Program

Do you have any clearly defined behavioral goals? (eg increasing seat belt use by 15%)

How will you measure the success of the program? (This relates to the Goal of the program)

How does this program propose to meet this goal?

Are there any restrictions or guidelines you must meet in your jurisdiction that we need to be aware of? If yes please include a copy of these.

Please include a description and any other comments you feel are relative to this curriculum review.



## **APPENDIX J: Hiring an Evaluator**

One important aspect of an effective program evaluation is determining who will participate in the evaluation and whether outside resources are needed. This appendix provides specific guidance to facilitate this process. The information provided here has been adapted from several sources, listed on page 335.

The focus of this appendix is to provide guidance on hiring and working with an outside evaluator. Many driver education program personnel may not have experience hiring and working with external evaluators. This information can help determine how and when to consider doing so. In some cases, an external evaluator may be brought in to conduct the entire evaluation for an organization. The most likely scenario, however, is that a combination of internal and external resources will be used when an outside evaluator is involved.

To begin, consider whether your organization can manage the entire evaluation without outside assistance. In addition to the resource and logistical considerations identified in the *Guidelines*, here are some important questions to answer before deciding to undertake an evaluation on your own.

*Which evaluation skills do your staff and evaluation team have?*

Staff members may have most of the organizational and administration skills required to carry out the evaluation. If this is the case, identify who has the necessary skills and whether they are available to participate in the evaluation. If the staff does not have these skills, consider in-house training, workshops, or conferences to provide or enhance them. The investment may be worthwhile in establishing an in-house evaluation capability that will be available on an ongoing basis.

*What is your staff's interest in evaluation?*

Teaching your staff new skills can be very rewarding for staff development and organization building. It can also enhance job performance and commitment to the program—staff members can see the connection between evaluation and their work, and the evaluation does not place an unreasonable burden on their existing workload. Since there has to be buy-in, consider an organization-wide discussion about the philosophy and objectives regarding program evaluation.

*Will staff involvement in the evaluation compromise the objectivity of the results?*

Because staff members may have a significant stake in the evaluation (e.g., their performance may be judged by the findings), they should not be involved in evaluation tasks that will bias

the results. Staff members, for example, can usually maintain their objectivity when administering pilot or pre-tests, and post-tests; however, a questionnaire or a focus group on customer satisfaction might provide more honest feedback if the customers can remain anonymous. Program managers and supervisors should also supervise their staff adequately to ensure the evaluation's integrity.

*Will using program staff to perform essential evaluation activities benefit the evaluation?*

Program staff can play a crucial role in the evaluation. In fact, using internal staff to conduct the evaluation is one way to improve its usefulness. An evaluation plan can be broken into a series of activities that various people can take on without overburdening their workload. It may be that some of the evaluation tasks are already being performed by staff members in their work. Instructors, for example, may already be collecting information about customer satisfaction or student preferences related to instructional materials or delivery methods, but no coordinated process to compile this information is in place. A simple but systematic examination of the data, when coupled with the experience of program staff, can yield sensible recommendations for program improvement (adapted from First 5 LA 2003).

Undertaking a solely in-house evaluation may be feasible; however, it would not be unusual for the evaluation team members to feel they need some help. The gaps that exist between what will be evaluated and the availability of internal resources and expertise are good indicators of whether outside expertise and assistance are needed.

## **Working with an External Evaluator**

An external evaluator can be a tremendous asset to an organization. Choosing someone from outside the organization can increase the program's learning process by offering new perspectives on program development and implementation. The right evaluator can offer a fresh perspective and also has the time and expertise to conduct the evaluation. External evaluators will also have specialized resources available to them, such as computer equipment, support staff, libraries, and research databases.

In addition, external evaluators may have broader evaluation expertise than internal evaluators, particularly if they specialize in program evaluation or have conducted extensive research on the target population. External evaluators can bring a different perspective to the evaluation because they are not directly affiliated with the program. This lack of affiliation, however, can sometimes be a drawback. External evaluators are not staff members, and they will have limited knowledge of the program's needs and goals.

If you decide to hire an evaluator, staff may still need to be involved in key aspects of the evaluation design and implementation. A partnership can be created between the evaluation team and the evaluator to determine evaluation questions, design the evaluation, interpret the results, and apply the findings. The evaluation team must also decide how the evaluator will be

used. Will the evaluator be a hired hand—doing tasks that the team does not know how or have time to do? Will the evaluator be selected for his or her expertise in a particular area to assist with a specific task? Or, will the evaluator be asked to work as a partner with the organization, providing guidance and support? Depending on organizational and evaluation needs, it is quite possible that the evaluator will take on a combination of these roles.

Once the role of the evaluator has been decided, it is crucial to determine what the expectations are of this person. How often should there be contact with the evaluator? What will be the final product? Will the evaluator be required to recommend program changes or get involved in implementing suggestions? Make a list of the required tasks and the desired working relationship. Who will be the contact person(s)? Who will supervise the evaluator? Answering these questions first will help decide whether the right evaluator has been found.

## **Finding An Evaluator**

The first and usually the best place to start your search is with other organizations that have experience working with external evaluators and do similar work. Referrals are a good sign that the evaluator has previous experience working in the field. Other places to search include professional associations, local colleges or universities, large corporations (pro bono or low-cost consultants), and on-line and print directories. Also, graduate students who are doing research in the driver education or young driver areas may be willing to help with little or no monetary compensation, especially if funding can be acquired through their university.

Try to identify and interview at least two prospective evaluators, and invite them to meet with the evaluation team. During the initial meeting, be sure to discuss: 1) the program's background and evaluation needs; 2) the expectations of the evaluator's role and possible tasks; and 3) the evaluator's background, expertise, and experience with similar programs, young novice drivers, and evaluation projects. Suggested questions for the evaluator include:

1. What strengths do you possess that will prove particularly helpful in connection with this evaluation?
2. Have you worked with similar evaluations? What did you learn from the experience? What would you do differently if you could repeat the experience?
3. How would you propose to divide up the tasks among team members?
4. Talk about the responsibilities the program must assume in order to make our work together successful.
5. Are you available to complete this work during the time we've specified?

---

First 5 LA 2003, adapted from Consultants ONTAP, [www.ontap.org/advice.html](http://www.ontap.org/advice.html).

Pay close attention to professional style, demeanor, listening skills, philosophy, and overall fit with your needs. If a good fit seems to exist, ask the evaluator to submit a proposal (including cost, approach, timeline, and deliverables) and a list of current clients and references. Review each proposal according to a predetermined set of assessment criteria, such as:

- Understanding of program and evaluation needs
- Required experience and expertise
- Excellent written and verbal communication skills
- Affordable budget
- Track record in field
- Able to meet schedule
- References

## **The Role of An External Evaluator**

While there is no “best” time to hire an evaluator, experience has shown that successful project managers hire evaluators sooner rather than later. Once an evaluator has been hired, it is important to establish a working relationship with program staff. The evaluation should not be isolated from the program’s day-to-day activities. Generally, an evaluator should work collaboratively with the evaluation team to:

- Learn about program goals, objectives, and activities
- Understand the perspectives of everyone involved in the program
- Set the boundaries of the evaluation
- Select the evaluation methods
- Collect and analyze the data
- Report the findings to appropriate audiences
- Recommend strategies for program improvement
- Always abide by specified ethical standards

The figure on page 333 presents guidelines for working with evaluators.

## The Ten Rules of Working with Evaluators

### **Evaluator's Responsibilities**

- I. Include the input of staff in designing the evaluation plan and selecting evaluation tools.
- II. Get to know the program through observation, interviews, participation in meetings etc.
- III. Use various methods to collect information about the program.
- IV. Collect data in the least intrusive and cost-effective way possible.
- V. Be sensitive to the needs and characteristics of program participants.
- VI. Maintain the privacy of participants and confidentiality of the data collected at all times.
- VII. Prepare reports about progress throughout the course of the contract.
- VIII. Provide feedback to staff and management about the program, and recommend how to use this information to improve the program.
- IX. Make a presentation or prepare materials of the final results.
- X. Hand over all data and documents to the program at the end of the contract.

### **Team's Responsibilities**

- I. Be an active participant in the evaluation process.
- II. Be clear about what the evaluation will accomplish and which resources are available.
- III. Communicate regularly with the evaluator and keep the lines of communication open.
- IV. Be honest with the evaluator about any problems or challenges the program is having.
- V. Make accessible any documents or people that the evaluator will need.
- VI. Inform the evaluator of any changes that will affect program implementation.
- VII. Inform others of their role and that of the evaluator.
- VIII. Be patient with the evaluation process.
- IX. Be willing to accept and implement the recommendations of the evaluator.
- X. End a bad relationship with an evaluator.

Adapted from First 5 LA 2003.

The table below presents some advantages and disadvantages of carrying out an evaluation on your own versus hiring an external evaluator to help.

### Trade-Offs Between Internal and External Evaluators

Internal Evaluator		External Evaluator	
Advantages	Disadvantages	Advantages	Disadvantages
Knows the organization, the program, and operations	May lack objectivity and thus reduce credibility of findings	May be more objective and find formulating recommendations easier	May not know the organization, its policies, procedures, and personalities
Understands and can interpret behavior and attitudes of program members	Tends to accept the position of the organization	May be free from organizational bias	May be ignorant of constraints affecting feasibility of recommendations
May possess important informal information	Is usually too busy to participate fully	May offer new perspective and additional insights	May be unfamiliar with the local political, cultural, and economic environment
Is known to staff, so may pose less threat of anxiety or disruption	Is part of the authority structure and may be constrained by organizational role conflict	May have greater evaluation skills and expertise in conducting an evaluation	May produce overly theoretical evaluation results (if an academic institution is contracted)
Can more easily accept and promote use of evaluation results	May not be sufficiently knowledgeable or experienced to design and implement an evaluation	May provide greater technical expertise	May be perceived as an adversary arousing unnecessary anxiety
Is often less costly	May not have special subject matter expertise	Able to dedicate him- or herself full time to the evaluation	May be costly
Doesn't require time-consuming recruitment negotiations		Can serve as an arbitrator or facilitator between parties	Requires more time for contract negotiations, orientation, and monitoring
Contributes to strengthening internal evaluation capability		Can bring the organization into contact with additional technical resources	

Source: *UNFPA Evaluation Toolkit*, Adapted from UNICEF Guide for Monitoring and Evaluation 1991.

## Resources

### *Program Evaluation Kit*

First 5 LA

Los Angeles County Children and Families First Proposition 10 Commission, Research and Evaluation Department

[http://www.first5.org/docs/Community/CommRsrc\\_EvalKit\\_0603.pdf](http://www.first5.org/docs/Community/CommRsrc_EvalKit_0603.pdf)

### *Programme Manager's Planning, Monitoring and Evaluation Toolkit*

The United Nations Population Fund (UNFPA)

<http://www.unfpa.org/monitoring/toolkit.htm>

### *W.K. Kellogg Foundation Evaluation Handbook*

W.K. Kellogg Foundation

<http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf>

The Evaluation Center, University of Western Michigan Checklists:

- *Checklist for Negotiating an Agreement to Evaluate an Educational Program*  
<http://www.wmich.edu/evalctr/checklists/negotiating.htm>
- *Budget Development Checklist*  
<http://www.wmich.edu/evalctr/checklists/evaluationbudgets.htm>
- *Evaluation Contracts Checklist*  
<http://www.wmich.edu/evalctr/checklists/contracts.htm>