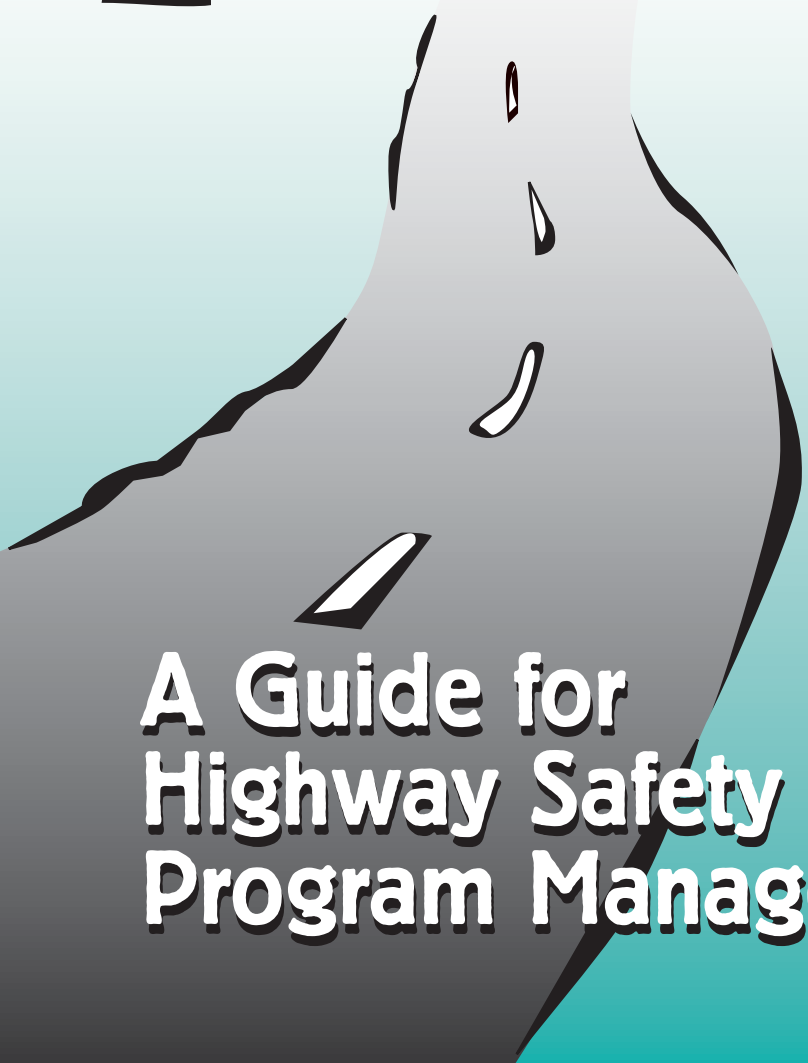
The Art of Appropriate Evaluation



A Guide for Highway Safety Program Managers

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

Once upon a time there was a manager who was responsible for starting up a new pedestrian safety program. Because it was new, her boss asked her to evaluate the program to find out how well it worked. Alarm bells rang in her head; she had never done an evaluation and it seemed way beyond her ability. When she discussed this assignment in her regular staff meeting, one of the staff volunteered to take on the responsibility. Greatly relieved, she gave him free rein.

The staff member immediately busied himself designing data collection forms and survey instruments. He wrote instruction manuals for filling out the forms and distributed them to the folks who were involved in publicizing the program. His research designs called for dividing the city into four regions that would each receive different combinations of the program's components. His weekly project reports were filled with detailed accounts of new forms, focus group protocols, new data collection and analytical procedures, and statistical tests. It seemed that everything was under control.

As the program reached its peak of activity, things took a turn for the worse. Data collectors weren't filling out the forms correctly, and no one could get a handle on the mountains of data the survey produced. The evaluator spent most of his time analyzing the change in public perception of the program. The difference was statistically significant, but so small as to be practically negligible. The progress reports started documenting why it was impossible to conduct a valid evaluation, with terms like changes in data definitions, and confounding variables leading the list of excuses.

The net result was that more than 20 percent of the project's resources were spent on evaluation and no one could answer the simple question "did it work?" The project manager vowed "Never again!"

The term evaluation evokes similar nightmares for anyone working in the public sector. We have all heard stories about expensive evaluation efforts that yield reams of complex data that end up confusing people. None of us wants an evaluation like that. We want to document the good parts of our program and find the things that need to be changed.

Evaluation is a term that refers to the process by which *someone determines the value of something*.

Value doesn't only mean monetary value; so evaluation doesn't necessarily involve converting something into a dollar and cents issue. It is simply examining, appraising, or judging the worth of a particular item or program.

We all conduct evaluations whenever we are contemplating a major purchase. If we are considering a new car purchase, we must decide if a vehicle is worth the price being asked for it. We go through three distinct evaluation processes to make that determination.

1. We first determine what we need in a car and what we would like to have. (Maybe I want a car that makes me "look good" behind the wheel.)

2. We then determine if the car we are looking at will meet these needs and wants. (The sassy red convertible definitely fits the bill.)

3. If it does, we must decide if we are willing to pay the price being asked. (Am I willing to pay $6,000 more than I planned in order to "look good?)

Once we have purchased the car, we probably continue to evaluate, but we sometimes call it "having second thoughts." After the purchase is made, we try to determine if we made a good choice. Did the car deliver on the advertising promises? Did it meet our personal needs and wants? Did it actually cost what we planned or did the car require a lot of expensive maintenance to keep it running. *If I had it to do over, would I buy the same car? Would I recommend it to a friend?*

To conduct an evaluation of a bicycle helmet use campaign, you probably do not have to design a complicated experiment. You really just need to collect helmet use data before and after your program, being careful to follow the exact same procedures both times. Many make the mistake of assuming that unless a program evaluation involves a complex research design, and sophisticated statistical analyses, it can't be a good evaluation. This is not true. Program evaluations do not have to be full-blown experiments in order to be valid. They just have to be carefully

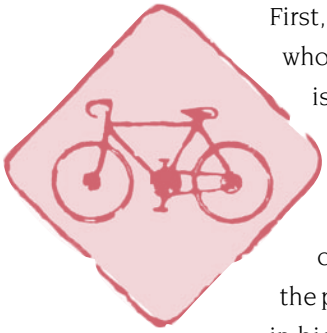
When you are implementing a traffic safety program, you should be making the same types of judgments. You build evaluation into your program so that you can determine:

- The exact nature of the traffic safety problem you are trying to address (*10 percent of the 50 traffic-related deaths last year were child bicyclists. None of the children were wearing bike helmets.*)

- What are reasonable goals and objectives for reducing this problem (*to decrease the number of bicyclist fatalities by increasing bike helmet usage to 80 percent among child bicyclists*)

- How well the program you implemented accomplished your objectives. (*Bike helmet usage increased from 45 percent before the program to 85 percent after the program.*)

What do you notice about these three bullets? They are specific, focused, and practical.



First, the evaluator identified a specific problem (The kids who died were not wearing bicycle helmets.) Next there is one focused program approach to address this problem. (Increase bicycle helmet use.) Note that there is no mention of how you are going to do this: free helmets, school programs, bike safety events or whatever. Finally, there is a practical measure of the progress your program made. (Document the change in bicycle helmet use.)

Why You Want to Read This Guide

A lot has been said over the years about the importance of program evaluation in traffic safety. At various times, program managers have been required to allocate a specified percentage of their program budgets to program evaluation. Training programs have been developed on how to evaluate traffic safety programs using such statistical tools as time series analysis and multiple regression analyses. And despite all of this attention, criticism continues to pour in about the fact that most traffic safety programs are never actually evaluated. And it is no wonder. Some program managers are convinced that program evaluation is too hot to handle, that it causes nothing but trouble, and costs a fortune to boot.

This Guide will convince you otherwise!

It is designed to alleviate your fears about program evaluation and convince you that conducting an appropriate evaluation actually makes your job easier rather than

harder. The focus is on what evaluation can do for you, not the other way around.

The Guide provides an overview of the steps that are involved in program evaluations and gets you thinking about how these steps fit into your implementation plans. It also will provide you with some handy suggestions on how to find and work with an evaluation consultant. And finally it will provide you with a handy glossary of evaluation terms and concepts so that you speak with confidence when the topic turns to "proving results." (When you encounter an underlined term such as Before and After Design, you can refer to the Glossary for its definition.)

The Guide provides an overview of the steps that are involved in program evaluations and gets you thinking about how these steps fit into your implementation plans.

It is equally important that you recognize what this Guide is not. It will not give you detailed, step-by-step instructions on how to evaluate a traffic safety program. Our assumption is that you are already too busy to take on a new career as a evaluation specialist. There are talented individuals in your own community who can help you design and conduct an appropriate evaluation. This Guide will tell you how to find and work with them.

The focus of this Guide is on using limited resources to maximum, practical, advantage. This means conducting an evaluation that is appropriate to the size and scope of the program you are implementing.

Who the Guide is for

Before we go any further, it's time to share the assumptions we have made about who you are. If you are a state

or local traffic safety project director with at least some curiosity about program evaluation, this Guide is for you. Our assumption is that you do not have a background in experimental design or statistics and have no intention of becoming an evaluation expert. (If you really want to become an expert, you should enroll in some college-level statistics courses—this is not one of those subjects you can teach yourself with a book!) You need to understand:

- what type of evaluation is reasonable for the type of program you are implementing;
- what you can do to maximize the success of a program evaluation; and
- where you can get help.

If that is what you are looking for, this Guide is for you!

How the Rest of the Guide is Organized

The remainder of this Guide is organized into six sections, and an appendix. They are:

II. The Evaluation Mentality—This is where we convince you that program evaluation is always a good idea.

III. In Search of the Appropriate Evaluation—A discussion of what you can reasonably expect a state or community program evaluation to accomplish

IV. Evaluation Step-By-Step—A high level overview of the steps involved in program evaluation, from defining your problem to reporting results

V. **Getting Help**—What you should expect from an evaluator, where to find them, and how to work with them.

VI. **Closing Comments**—A wrap-up of the arguments in support of always evaluating your program efforts.

VII. **Glossary of Terms**—Some basic evaluation terms defined to increase your comfort level around evaluators.

CHAPTER TWO

THE EVALUATION MENTALITY

Evaluation is like regular exercise. We all know that exercise is good for us. And we all try it from time to time. But the majority of Americans fall far short of the recommended exercise requirements. Why is that?

In most cases, we don't exercise regularly because we have convinced ourselves that exercise requires too much effort, that it will hurt, and that it probably won't give us a perfect body anyway.

It's the same way for evaluation. Everyone acknowledges that it is always a good idea to evaluate any program that uses taxpayer dollars, but when it comes time to build an evaluation into a program plan, dozens of excuses are offered as to why it just can't or shouldn't be done, in this particular case. So why do many people shy away from conducting program evaluations regularly?

Everyone acknowledges that it is always a good idea to evaluate any program that uses taxpayer dollars, but when it comes time to build an evaluation into a program plan, dozens of excuses are offered as to why it just can't or shouldn't be done. Why do many people shy away from conducting program evaluations regularly?

Most excuses for not doing an evaluation are variations on the following four themes.

- "Evaluation is too complicated. Program evaluations require complex research designs for sophisticated experiments. I don't fully understand what it is involved but it sounds pretty scary."

- "If I conduct an evaluation, I may discover that my great idea was a total flop. I want to avoid that embarrassment."

- "I have a very limited budget. Evaluations are expensive and time-consuming. I just can't spare the resources required."

- "Evaluation is a lot of work, and I don't have the time. Besides, I never understand the gobbledygook that I read in evaluation reports. So why bother?"

Do some of these sound familiar? Have you found yourself thinking these thoughts? Let's go through them one by one to show you why they aren't true and may be standing in the way of your success in traffic safety.

"Evaluation is too complicated."

Many are intimidated by the whole concept of evaluation. A mystique has built up that program evaluation is very complicated with a hundred ways to do it wrong and only one, very difficult, way to do it right. And in some cases this mystique is justified. When you are talking about establishing a direct cause and effect relationship between a specific traffic safety countermeasure and a reduction in traffic deaths, you need a solid research design with random assignment to experimental and control groups and sophisticated statistical analyses. More importantly, you need large numbers of cases in order to detect any real change in traffic deaths.

In reality traffic safety evaluation need not involve "Ivory Tower" laboratory science.

In reality, however, traffic safety evaluation need not involve "Ivory Tower" laboratory science. Traffic safety evaluation is an applied science that works within the constraints of state and local program implementation. Most local communities simply do not have

the volume of traffic deaths and injuries to conduct that kind of countermeasure effectiveness evaluation. Instead, these communities can focus their evaluation efforts on determining if the particular program they implemented achieved its specific objectives.

If you are implementing an occupant protection program, your evaluation dollars would be better spent demonstrating an increase in safety belt use rather than proving, once again, that safety belts save lives. Documenting an increase in safety belt use over baseline levels involves a much simpler evaluation and should not deplete your program resources.

"I may not like the answer so I better not ask the question."

When you implement a program, you put a lot of yourself into the effort. You believe in your heart that it is a good program, and you do everything you can to make it work. But you never know what might happen.

Some people shy away from evaluations because they don't want their good idea to be proven wrong. The mistake they are making is viewing evaluation as the last step in a process, like a final exam. If evaluation is tacked on at the end of the project, you may very well come up with answers that you don't like .

The key to successful evaluation is to build evaluation in from the start, so it can help you frame the questions you are asking, and even clarify the problem that you are trying to solve. A well-planned

The key to successful evaluation is to build evaluation in from the start, so it can help you frame the questions you are asking, and even clarify the problem that you are trying to solve.

evaluation should not yield last minute unpleasant surprises. Instead, it will provide useful information that helps fine-tune the program at every step of the way. It also tells you what's working and what can be made better.

"I have a limited budget; I prefer to spend my dollars on implementation."

When it is time to ask your funding source to extend your program for another year, your proposal will be much more persuasive if it is backed up with solid data demonstrating that you accomplished your objectives.

These days, everyone is expected to do more with less. Project directors struggle to stretch every program dollar to the limit. When asked to choose between delivering more high school presentations or conducting a program evaluation, many choose more presentations because they believe that their impact will be greater if they can reach more kids. There are two errors in this logic.

1) They believe that an evaluation will cut into their program implementation resources, when in fact there are many ways to conduct an evaluation inexpensively.

2) Reaching more kids with a perfect traffic safety message is always a good idea. But how do you know if your message is being well-received unless you do some sort of evaluation? Collecting and monitoring feedback throughout a project provides the opportunity to fine-tune your message as you go in response to audience reaction. At the conclusion of the program you will also have the data you need to improve the entire program before you offer it again.

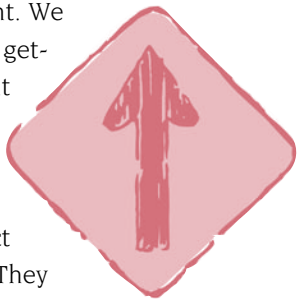
Another point to keep in mind when thinking about conserving program dollars is satisfying your funding source. When it is time to ask your funding source to extend your

program for another year, your proposal will be much more persuasive if it is backed up with solid data demonstrating that you accomplished your objectives.

"Evaluation is too much work."

Evaluation can be labor-intensive (note that we didn't say expensive) and very tedious. This is why evaluation is one of those areas that, as a project director, you should delegate to someone else. This could be an employee from another department in your agency, a faculty member or graduate student at a local university, or a private evaluation consultant. In addition to doing the time-consuming work involved in data collection, you should expect an evaluation specialist to be able to explain the results in language you and everyone else can understand. There is no place for gobbledygook in an evaluation report. (Section V of this Guide talks about what to look for in evaluators and how best to work with them.)

Now that we have eliminated all your old reasons for *not* doing an evaluation, we can concentrate on some new reasons for *doing* one.



Let's go back to our exercise analogy, for a moment. We all know some people who are fully committed to getting regular exercise. They don't have to think about if they will exercise or how they will fit it into their schedule. Exercise is an essential part of their daily existence, just like eating and sleeping. These individuals report that exercise improves every aspect of their life, not just their physical conditioning. They have more energy, they are more productive at work, and they are less prone to depression. What separates them from the rest of us is their exercise mentality.

To get the full range of benefits from evaluation, you need to have an evaluation mentality. This means that you never even consider implementing a program without first thinking like an evaluator. You approach each new problem with the same set of questions:

- What do I know about the problem?
- If I tried to fix this problem, what could I accomplish?
- How could I measure my results?
- How can I collect the data I will need?
- What are my criteria for success?

With the answers to these questions in hand, you are prepared to convince any funding source that you know what needs to be done and that your ideas have a high probability of success. Your evaluation mentality will also ensure that at the end of this project, you can report back to these same funding sources with solid information on what you accomplished.

An evaluation mentality cannot ensure that every project you implement will be a resounding success, but it can ensure that you fully understand what you tried to do and why things turned out as they did.

An evaluation mentality cannot ensure that every project you implement will be a resounding success, but it can ensure that you fully understand what you tried to do and why things turned out as they did.

If you have an evaluation mentality, you design your program and your evaluation at the same time. The benefits of this approach are substantial. An evaluation mentality will enable you to:

- ***Accurately identify the problem you are trying to solve***—All too often people jump into implementing a

program without really understanding the underlying cause of their problem. Is underage DWI a problem because the liquor stores are selling to teenagers or because the police are not targeting the locations where kids are drinking? It is not enough to suspect that safety belt use is low in your community. You need to determine up front what target groups make up your non-users. An evaluation mentality frees you from having to take a "shot-gun" approach to solving every problem.

- ***Uncover some problems you didn't know you had***—You might assume that your pedestrian safety problem involves the very young and very old until you discover that a significant percentage of your pedestrian fatalities are working age adults who had been drinking at the time of their crash. This problem would require a totally different set of countermeasures than a child pedestrian problem.

- ***Establish reasonable, practical objectives for dealing with these problems***—Global objectives are the hardest to accomplish. With good problem identification data, you can focus your objectives on the specific problem you are trying to solve. Instead of trying to "reduce unsafe driving behaviors" you might want to reduce red-light running which is the unsafe behavior that is causing the most concern in your community.

- ***Determine if you have accomplished your program objectives***—A major purpose of any evaluation is to determine if your program accomplished its objectives. Well- thought out objectives are an important

first step, but an evaluation mentality will also help you zero in on what you truly need to measure. Too often, project directors waste time and money collecting data that they can never use, because they can't compare it to any baseline or because it does not relate to their program objectives.

- ***Provide information to funding sources, the media and the public to continue support for program***—Support for a good idea can evaporate if there is no evidence that the idea really works. With an evaluation mentality, you create consensus from the beginning on the criteria for success, and you stay focused on the data you will need to measure your performance against those criteria. You also understand that a significant, but barely discernible, change may not constitute success in everyone's eye.

- ***Determine if and how a program should be revised to increase its effectiveness***—With an evaluation mentality, you don't wait until the end to find out how your idea worked. Very few projects work perfectly. There are always aspects that could be tweaked to make them more effective. With an evaluation mentality, you monitor performance throughout the project so that you can institute mid-course corrections if needed, and so that you are ready to revise your concept for next year.

You do not need to be an evaluation specialist to have an evaluation mentality, just as you don't have to be body builder to have an exercise mentality. You just have to recognize that evaluation, when built in from the beginning, provides benefits throughout the life of your program.

CHAPTER THREE

IN SEARCH OF THE APPROPRIATE EVALUATION

The program evaluation in the introductory scenario is everyone's worst evaluation nightmare because it didn't demonstrate anything other than a lot of good intentions and confused activity. The person running that evaluation clearly did not have an evaluation mentality and did not design an evaluation that was appropriate to the size of the project and to the data that was available.

One of the most critical elements in a successful evaluation (that is, one that actually proves something) is deciding what should be demonstrated. This decision should be based on the type of project you are implementing and the type of data that are collectable (or available). Your goal is to set up an evaluation that is appropriate for your individual circumstances.

What Makes an Evaluation Appropriate?

If the National Highway Traffic Safety Administration (NHTSA) is going to promote a brand new traffic safety countermeasure as an effective tool in reducing traffic deaths and injuries, it is reasonable to assume that NHTSA will have thoroughly evaluated this countermeasure in realistic conditions to make sure it works. This would require conducting several full-scale evaluation research projects that verify the effectiveness of the countermeasure. NHTSA can also call upon large volumes of national and state level crash data with enough records to confirm, with a high degree of confidence, that changes can be attributed to

the countermeasure. A full-scale countermeasure effectiveness evaluation project is the only type of evaluation that would be appropriate in these circumstances.

Two years later, after this new countermeasure has been implemented in several communities, a program manager in a city of 75,000 reads about it in a NHTSA publication and decides that it might be just what is needed to solve a troubling traffic safety problem in his community. This program manager has a solid evaluation mentality so he immediately considers what type of evaluation would be appropriate for his circumstances. He does not need to conduct the same type of evaluation that NHTSA conducted because:

1. He is not trying to prove to the nation that it works; his boss was convinced by NHTSA's evaluation results.

2. His community experiences only a few crashes of the type affected by this countermeasure, (but he still would like to reduce that number even further).

3. Resources are limited.

He needs to determine what an appropriate evaluation would be for these circumstances. There are two types of evaluation questions that are appropriate for most local, and even some State, programs:

- Did you implement the program as planned?
- Did you accomplish your objectives?

Did You Implement The Program As Planned?

At the most fundamental level, you can do an evaluation to determine if you implemented the program as planned. This may sound pretty obvious, but in fact many projects take a wrong turn right off the drawing board. This approach, which is called an administrative evaluation, does not require any elaborate data collection efforts or even a research design. All that it requires is an understanding of what is supposed to happen during a program and a systematic approach to tracking what actually happens.

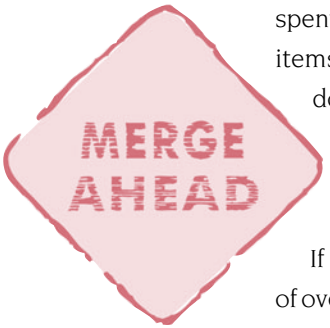
Let's go back to the bicycle helmet program on page 2. Suppose you decide you're going to have two safety fairs over the summer and you're going to give away free helmets, donated by a community sponsor. An administrative evaluation would keep track of the number of helmets you obtained and the number you gave away. It might also document such things as the age, gender and neighborhoods of the children who received the helmets, the number of people who participated in the safety fairs, and the amount of publicity you received about the fairs.

Some managers might dismiss this type of administrative evaluation as simple "bean counting" that doesn't demonstrate anything worthwhile. You will be surprised by what you can learn merely by checking to see whether everything is going as planned.

In one community, a mandatory jail sentencing program for DWI repeat offenders was implemented. The program was evaluated to determine if serving time had any effect on recividism. The evaluators were never able to determine this effect because of an unexpected finding. Although most repeat offenders were sentenced to jail time, the evaluators discovered that very few of them actually served any time. There was no system in place to follow up with individuals when they left the courthouse. Obviously the program manager had to go back to the drawing board to solve the problem of ensuring that the court sentences were actually carried out.

Another community decided to implement an occupant protection traffic enforcement blitz, complete with highly visible public information and media coverage. The evaluator kept track of the number of police officer hours spent and the number and type of citations issued. The program staff were surprised to find that although lots of safety belt citations were issued during the first week, there were no citations issued for child safety seat violations. The police officers did not seem to fully understand the requirements of the State law. This discovery led to a police roll call training session on the child safety seat law and on the importance of enforcing it. During the second week of the blitz, forty-seven citations and warnings were issued for child safety violations.

If you monitor your program from the beginning, you will be able to spot any implementation problems early and determine if the problem can be fixed or if the whole idea should be scratched. There is no sense wasting dollars going through the motions of implementing a program with fatal flaws.



An important element of documenting how the program was implemented is tracking the resources as they are being spent. Every project should have a detailed budget for such items as staffing, supplies, etc. A good evaluation should document whether the project was completed within budget or over budget. The rate at which resources are being spent can sometimes give a good indication if the project is being implemented as planned. If local police are not putting in the budgeted amount of overtime, for example, maybe the sobriety check-points are not being conducted as frequently as planned.

Did You Accomplish Your Objectives?

Everyone knows that you conduct an evaluation to demonstrate that you accomplished your objectives. You don't need an evaluation mentality to realize that. But it does help you understand what objectives you should be measuring.

People usually write goals and objectives to impress a funding source. They are frequently written in grandiose terms that sound impressive but lack a clear focus.

- To reduce traffic deaths (Do you want to promise that in your small town?)

- To increase support for traffic safety (How will you measure this?)
- To improve safe driving behaviors (What behaviors do you care about?)

When challenged, the individuals who wrote these objectives were able to revise them to focus on what an individual project was specifically designed to do, not what sounded good on paper.

Reducing traffic deaths was changed to increase safety belt use—that was what they were really aiming for. Increase support for traffic safety was changed to getting 1,500 signatures on a petition for passage of a bicycle helmet ordinance, and improve safe driving behavior was changed to reduce the incidence of red-light running.

We cannot emphasize enough the importance of carefully defined objectives. They make the difference between a successful and a frustrating evaluation.

We cannot emphasize enough the importance of carefully defined objectives. They make the difference between a successful evaluation and a frustrating one. You should read Section IV for more detailed suggestions on writing SMART Objectives.

What Might Not Be Appropriate to Demonstrate?

It is very difficult to link a countermeasure program to a reduction in deaths and injuries at the local level, (and sometimes, even at the State level). There are several reasons for this.

- Although traffic crashes are a serious national problem, killing more than 40,000 per year, traffic deaths in any

community are relatively rare events. Most communities will experience fewer than a dozen traffic-related fatalities a year resulting from all causes. Furthermore, the number of deaths might fluctuate considerably from year to year, for no apparent reason. Given that the number of deaths might go up or down regardless of what new program you implemented, you might not want to raise expectations that your program will save lives. It would be far better, for example, to demonstrate that your program resulted in an increase safety belt use.

- Traffic deaths are influenced by a variety of factors, all of which can influence whether fatalities climb or drop. These factors, called variables, could include:

✓ The amount of driving in the community (an increase in gasoline costs could reduce the amount of miles traveled, or a new shopping mall on the outskirts of town could increase vehicle travel.)

✓ The weather conditions, (a very bad winter could lead to an increase in fender bender type collisions, but major injuries might go down because people drive less and at slower speeds in bad weather)

✓ A change in the driving age (reducing the minimum age could increase crashes caused by inexperienced drivers)

✓ A change in the population (a downward trend in population growth could reduce the number of drivers on the road.)

✓ Previous extremes (a shift back to "normal" levels after reaching an extreme value, either high or low)

If you are trying to establish a connection between a particular countermeasure and a reduction in deaths and injuries, you have to ensure that none of these variables, or any others you might think of, contributed to that change.

- Since the number of fatalities that occurs in most communities is so small, if you were committed to demonstrating a reduction in fatalities, you would need to aggregate your data over several years in order to have enough deaths to show a real decrease.

This approach creates an entirely different problem related to existing data: it is very difficult to compare data that were collected in widely separated time periods whether you are looking for fatalities or some other measure such as citations issued. Over time, changes in data collection procedures, data definitions, and enforcement thresholds can change significantly. For example, a community may change its policy concerning the collection of blood alcohol content data on traffic fatalities, making it difficult to compare the number of alcohol-related deaths over a five-year period. Or a Traffic Records Department may change its definition of a "reportable" crash from $250 or more in damages to $2,000 or more in damages. This would spuriously decrease the number of reported crashes.

it is very difficult to compare data that were collected in widely separated time periods whether you are looking for fatalities or some other measure such as citations issued.

These problems with linking countermeasures directly to bottom line changes in fatality levels are not insurmountable. However, they do require a significant increase in the complexity and cost of an evaluation. You should undertake this extra effort only when it really is necessary, like when you are trying a countermeasure that has never been tried anyplace else.

If your countermeasure has been around for a while, why do you want to spend precious resources to prove what has already been proven? The traffic safety community has demonstrated to most everyone's satisfaction that safety belts and strong DUI laws save lives. If you are implementing an occupant protection program, therefore, you don't need to link your program to a reduction in deaths and injuries. Instead, you can limit your evaluation to demonstrating that you accomplished your objective to increase the rate of safety belt use by a specific percentage.

Similarly, since the effectiveness of sobriety checkpoints has been thoroughly evaluated, you can focus your evaluation dollars on demonstrating that the number of sobriety checkpoints you planned were conducted and that citations for DWI offenses increased. It is not necessary to attempt to link this accomplishment to a reduction in alcohol-related fatalities.

WHAT WORKS

If your program involves one of the following strategies, you can concentrate your evaluation dollars on documenting that you implemented the countermeasure, not that the countermeasure saved lives.

- Safety belts
- Child safety seats *(always in the back seat!)*
- Bicycle helmets
- Motorcycle helmets
- DWI enforcement
- Sobriety checkpoints
- Tougher impaired driving laws
- Crossing Guards
- Traffic Calming Devices *(e.g., speed bumps)*
- Educating judges and prosecutors

In order to prove that you accomplished your objective of increasing safety belt use or DWI enforcement you will still have to collect data and document your accomplishments. You will probably need to observe safety belt use before and after you implement you strategy, or collect enforcement data for a comparable period before you instituted your "blitz." If safety belt usage or DWI enforcement increased, your program was a success.

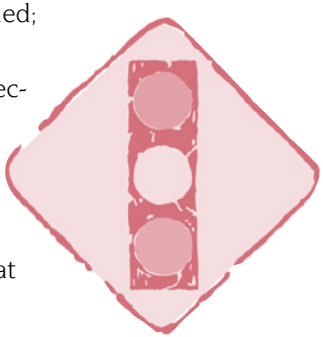
If it did not increase, than you should look at the strategies you used. Perhaps these techniques were not as effective as other options, (e.g., a public information campaign, by itself, will not be as effective at changing behavior as an enforcement campaign coupled with continuing media coverage.) Learning that something did not work does not make your evaluation a failure. It simply provides you an opportunity to learn more about your problem and to revise your approach in the future.

Learning that something did not work does not make your evaluation a failure.

Summary

A program evaluation can provide you the following information about your program:

- That you implemented the program as planned;
- What resources were spent, and
- Whether your program accomplished its objectives.



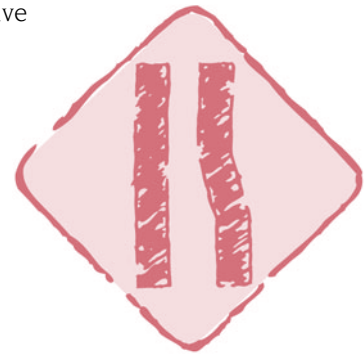
That level of detail is appropriate for most local and state level evaluations. In the next section we provide you with a high level overview of what will be involved when you take the plunge.

CHAPTER FOUR

EVALUATION STEP-BY-STEP

By now you should have gotten the message that evaluation is an integral part of program implementation and needs to be built in right from the start. Therefore, the primary steps involved in evaluation mirror the steps you follow when you implement a program.

1. Identify the problem you are trying to solve
2. Develop reasonable objectives
3. Develop a plan for measuring results
4. Gather baseline data
5. Implement your program
6. Gather data and analyze results
7. Report results



The remainder of this section provides a brief overview of what you should keep in mind in each of these steps.

Step 1—Identify Your Problem.

It may sound obvious but you need to understand the problem you are facing before you can expect to solve it. All too often, decisions are made to implement a program based on a reaction to a single, tragic fatal crash. It is always wise to take the time to understand your problem before you try to solve it. Problem identification serves two important functions.

- It provides you the information you need to select an appropriate countermeasure and target audience for your program. You will be looking for information on the magnitude of the problem, the underlying causes, and the target groups most affected. This information should enable you to select the most effective countermeasure.

- It may provide you with some of the baseline data you will need to determine if your program meets its objectives. You may start your problem identification with crash data, but you will also need to collect other types of data in order to understand the problem you have and to select the most effective strategy for dealing with it. This might include baseline observations of safety belt use, measures of enforcement levels, public opinion and awareness surveys, or speed counts. At this stage, it is also helpful to gather any trend data that may have been collected over the prior few years so that you will be able to show a trend before and after your program.

A Safety Team was formed to identify, develop, and implement countermeasures to reduce crashes on the Capital Beltway in suburban Washington, D.C. Everyone knew there was a problem, but there were a lot of opinions about what was causing the problem and how it should be solved. An evaluation was conducted at the beginning of the project, rather than at the end, to identify how, why, and where crashes occurred on the Beltway. This research led to a number of specific actions, including engineering changes, increased enforcement, and speedier incident response times.

During the problem identification step, you also lay the foundation for your data collection efforts throughout the program evaluation. As you collect your baseline data, it is critical that you carefully document the procedures you follow, so that data collected later in the project can be compared with your baseline. In order for the data to be compared, it has to be collected at the same locations and times of day, using the same collection forms, and

ideally the same observers. Failure to follow the same data collection procedures can make it difficult to document your accomplishments.

Step 2. Develop Valid Objectives for Your Traffic Safety Program.

Once you have identified your problem and selected your strategy for addressing it, you need to define what you expect to accomplish. Many would argue that this is the most critical step in the evaluation process because it determines what success will be and how it will be measured.

Program objectives should be SMART (Specific, Measurable, Action-oriented, Reasonable, and Time-specific).

Volumes have been written on how to write program objectives, each with its own set of do's and don'ts. These rules are all similar and it is not important which set you follow. The one advantage to the list shown below is that it is easy to remember.

Program objectives should be SMART (Specific, Measurable, Action-oriented, Reasonable, and Time-specific). Let us elaborate.

Objectives should be SPECIFIC: Avoid using generalities like "improving traffic safety" or "increasing awareness." If you identify exactly what you want to happen, then you can document your success. Sometimes you can be specific about the amount of change you anticipate, expressed either in absolute (increase safety belt use to 75 percent) or relative (increase citations by 15 percent over the baseline) terms. At other times, you can simply observe and record the change in behavior.

Objectives should be MEASURABLE: For an objective to be measurable, there must be something you can quantify, like DWI citations, and you must be able to detect a change over time. To the extent possible you should also be able to isolate the targets of the countermeasure. For example, you want to increase by 10 percent the number of DWI citations issued to young drivers.

Objectives should be ACTION-ORIENTED: Action is good. You usually can see an action and count the number of times it happens. It is much easier to document that safety belt laws were enforced, by counting the number of traffic stops and citations, than it is to document if public support for belt law enforcement increased.

Objectives should be REASONABLE: A small community implemented a public information campaign on the value of traffic safety enforcement. The published objective of this public service campaign was to reduce traffic deaths community-wide. While this would be a desirable outcome, it is not reasonable to expect that an advertising campaign alone would change behaviors and ultimately reduce traffic crashes, at least not within the time- frame of the study. This community should take another look at the problem they are trying to solve, select a countermeasure that will address that problem, and then establish a reasonable target for success.

Objectives should be TIME-SPECIFIC: Projects don't last forever and objectives should have deadlines. Deadlines make it clear to everyone when results can be expected. They also keep people focused on what needs to be accomplished by when.

SMART objectives don't leave you a lot of wiggle room. It will be very obvious if you meet them or not. They challenge you to accomplish what you set out to do and serve as a constant reminder of your criteria for success. This is all the more reason to be honest and practical when you write them.

SAMPLE OBJECTIVES

NOT SO SMART	S.M.A.R.T
• To encourage increased safety belt enforcement	• To increase safety belt citations by 15 percent in 6 months
• To reduce underage drinking	• To reduce the number of liquor establishments that serve minors by 40% in 12 months
• To work with the legislature to advocate tougher impaired driving laws	• To get a .08 law introduced and passed through committee in the next legislative session

Once you have drafted your objectives for your program, you need to circulate them to those decision-makers who hold the fate of your program in their hands. You need to get buy-in at the outset as to what you are trying to accomplish. If they are expecting dramatic bottom line results (i.e. a reduction in fatalities), now is the time to explain to them why that would be difficult, if not impossible to demonstrate in the short term. If you wait until the program is over, they will likely come to the conclusion that the program failed because it did not meet their objective even if the program accomplished your objective!

This does not mean that your community may not experience a reduction in deaths and injuries over time. If you continue to implement effective countermeasures targeting specific traffic safety problems, you should begin to observe a downward trend in crashes, deaths and injuries. Your decision-makers need to understand, however, that this improvement will not occur over night.

Step 3. Develop a Plan for Measuring Results

Before you can begin implementing your program, you have to plan how you will conduct your evaluation. This plan will address the questions:

- What will you measure?
- How will you measure it?
- How will you analyze your results?

While all of these questions are important, the first, what you will measure, is critical to the success of your evaluation.

What will you measure?

What you will measure must be tied directly to the objectives you have established for your program. If your objective is to reduce speeding on a given roadway, the most logical thing to measure would be average speeds on that given road. Since your objective is tied to speeding, you don't need to spend time or money trying to measure a reduction in crashes.

One of the reasons that you don't want to be forced into counting lives is that fatal crashes don't occur very frequently in most communities. It is very difficult to even observe a change in fatalities, let alone connect that change to a specific countermeasure.

The problem that you will face again and again, is that everyone else will be urging you to tie program success to saving lives. Rather than getting backed into that corner, you should point out that the traffic safety literature indicates that excessive speed contributes to serious crashes. Since you have documented that there is a problem with excessive speeds on specific highways in your community, you are going to implement

a countermeasure whose objective is to reduce speeding. You will measure program success by monitoring speeds on the selected road segments, before during and after your program is in effect. Wherever possible you should try to measure observable phenomena - things you can see and quantify, and that occur with a high degree of frequency. The phenomena can include behaviors, knowledge, opinions, and attitudes, and institutional responses. Here are some examples of each of these:

Behaviors

- Using safety belts and child safety seats
- Wearing bicycle and motorcycle helmets
- Speeding
- Red-light running
- Jaywalking

Public opinion, awareness, and knowledge

- Awareness of Public Information and Education campaigns
- Support for legislative initiatives
- Knowledge of a safety belt law
- Teen attitudes about drinking and driving
- Perceived risk of getting a traffic ticket

Institutional responses

- Citations issued by the police
- Special police patrols and check-points
- Presentations
- Training programs
- Media coverage
- Policies and legislation

Changes in these observable phenomena can be caused by your program or by some other confounding factors such as engineering improvements along a roadway. It will be important to understand what these confounding variables might be and how you can control them. This is a an area in which an evaluation specialist can be extremely valuable.

How will you measure it (and when)?

Once you have decided what you will measure to determine if your program achieved its objectives, you will need to decide how you will gather the information needed to make the measurement. There are four basic ways that you can measure program effects:

- Field Observations,
- Surveys,
- Forms, and
- Archival Data.

Field observations are used to measure changes in safety behaviors. They can detect the presence or absence of a behavior, (wearing, or not wearing a helmet), or record some measurement of a condition, such as a vehicle's speed, or the size of a traffic gap that a person accepts before pulling into traffic. To conduct a valid field observation you, or your evaluation specialist, will need to determine where and when to make the observations, how many observations will be needed, and what procedures will be followed to record the data.

Surveys are used to collect attitude, knowledge, and opinion information about individuals. They can be administered in person, over the telephone, or by mail, (and with growing frequency, via e-mail.) Each of these approaches has its own strengths and weaknesses which your evaluation specialist can describe for you. Surveys can provide a wealth of information but the survey instrument you use must be designed very carefully and tested thoroughly, and the procedures you use for including individuals in the survey (the sampling plan) must be well thought out. The classic problem that can plague a survey effort is bias introduced into the sampling plan. For example, if you conduct a telephone survey with your sample drawn from the telephone directory, you are limiting your respondents to households that have a telephone. Your target population might be college students who are not adequately represented in the telephone directory.

The classic problem that can plague a survey effort is bias introduced into the sampling plan.

Forms should be used to collect "process" data such as the number of presentations made (and where and when), the number of requests received for a brochure, the number of visits made to liquor establishments and the outcome, etc. These forms should be tailored to the specific data you need to capture and should be designed in coordination with the people who will be using them. There is a fundamental conflict between the people who would like to know the information and the people who actually have to collect it. Consideration has to be given to the conditions under which the forms will be completed (at a busy PTA meeting with people milling around or back at the office with access to a computer) and the amount of

Consideration has to be given to the conditions under which the forms will be completed (at a busy PTA meeting with people milling around or back at the office with access to a computer) and the amount of time it is reasonable to expect someone to spend on the task.

time it is reasonable to expect someone to spend on the task. The forms should be tested with real users prior to giving them out to be sure that there is no confusion.

Archival Data can be used to document a variety of issues. They are powerful because their use allows you to consider trends such as how a behavior, such as speeding, has changed over time. Archival sources would include:

- Police crash records
- Department of Motor Vehicle driver records
- Traffic citations logs
- EMS transport records
- Emergency room records
- Traffic court files
- Hospital disposition records, etc.

It could also include newspaper archives, city council or State legislature records, and any other files that document program activity or responses to program activity.

The biggest challenge you will face with archival data is getting access to it. Any organization that maintains databases with any personal information will have very strict guidelines for who can access the information and what can be done with it. Make sure that your evaluation specialist understands these data sources and has experience accessing them. Since your evaluation is not concerned with the identity of individuals, you can usually obtain summary, data with the personal identifying information deleted.

Keep in mind that archival data may change over time as improvements are made in data collection. You will need to check each data item that you are interested in to see if it is consistent.

If you are conducting a State-level evaluation that focuses on fatalities and injuries, you can also access the Fatality Analysis Reporting System (FARS), the General Estimates System (GES), and the National Automotive Sampling System (NASS), all of which are maintained by NHTSA. Information about these archival data sources can be obtained from:

> The National Center for Statistics and Analysis, NHTSA
> 400 Seventh Street, SW
> Washington, D.C. 20590
> Phone: 1-800-934-8517
> World Wide Web: www.nhtsa.dot.gov/people/ncsa

Once you have determined the type of data you will be collecting, and its source, you will need to develop systematic procedures for data collection. You cannot leave this important step to chance. You will likely have multiple people collecting data and you want to minimize any variations in how they interpret what they are seeing. You accomplish this by designing data collection forms that can be used by everyone, and by providing training on how to make observations, read police forms, etc. You want each individual to collect data in exactly the same way. If you collected observational data as part of your problem identification activity, use the same procedures you used then

You want each individual to collect data in exactly the same way. If you collected observational data as part of your problem identification activity, use the same procedures you used then so that you can make valid comparisons.

so that you can make valid comparison. Your evaluation specialist will be responsible for ensuring systematic data collection procedures.

There is one last consideration under the topic "How Will You Measure." That deals with the timing of your data collection efforts. We have discussed how your pre-and post data should be collected under similar conditions, which could include time of year, time of day, etc. You must also consider when the post data should be collected in relation to the implementation schedule. For example, you will want to collect safety belt use data immediately after a major enforcement blitz to determine if belt use changed. Traditionally, each increase after an enforcement blitz will level off over time. It won't go all the way back down to the "pre" level, but it will go down. So, you also need to know what the long-term effects of that enforcement blitz may be. You will therefore need to plan for follow-up data collection at scheduled intervals after implementation is complete.

Your schedule for data collection should be determined before implementation begins, so that it will not be influenced by the implementation itself.

Step 4. Gather Baseline Data.

Now that you have refined you program objectives and developed a plan for measuring results, you may collect some additional data about other aspects of your program.

During problem identification, you gathered preliminary data on such factors as safety belt use, and documented how you collected this information so that you can repeat these procedures after implementation. Now that you have refined you program

objectives and developed a plan for measuring results, you may collect some additional data about other aspects of your program. You may need to conduct an opinion poll to document what your citizens think about DWI enforcement, before you implement a campaign to conduct sobriety checkpoints on weekends. This information should all be gathered before you actually start implementing anything, so that you can easily isolate any effect your program may have.

Step 5. Implement Your Program.

Many people would be surprised to see implementation as a step in the evaluation process. But remember, you should be monitoring how your project is going right from the start, rather than waiting until everything is over. You should be keeping track of project costs and other process data that could indicate if program activity is at expected levels. You might do periodic opinion polls to see if the public is paying attention to your public information campaigns. You should also gather feedback at any training programs or public presentations. You may discover that there is a serious problem that should be fixed before any other contacts are made. If you include any media events in your program, you definitely want to pay attention to the amount of media coverage you receive. This information is much easier to capture in real time rather than to recreate the records weeks or months later.

Whatever you learn during program implementation, it is critical that you document it.

Whatever you learn during program implementation, it is critical that you document it. You may have planned for weekly sobriety checkpoints at five locations in the county,

with support from the State Patrol. Due to unexpected budget cuts, the State Patrol can only support one location per week. This may be a problem that you cannot fix, but you need to factor it in when you analyze your citation data. Based on this development you may want to adjust your program objective or extend the duration of your implementation phase. You will definitely want to document how actual implementation differed from your plan, and what impact you believe this change could have.

Step 6. Gather and Analyze Data.

While the work involved in planning an evaluation is critical to success, it is in this step that your evaluation specialist will earn his or her fee. Gathering the data is the most labor intensive aspect of the program evaluation, and analyzing it may be the most complex. As a manager, your biggest concern during the data collection phases is that the effort is adequately staffed and that everyone has been trained on the correct procedures to follow. Your evaluation specialist should also keep you informed about any changes that have to be made because of some external event that could influence the outcome. For example you may have collected baseline data on child safety seat use outside of a child care center. One year later, when you are looking to see if your campaign had an effect, you discover that the center has closed. Your evaluation specialist will need to find a suitable alternate site so that you don't miss any data.

CAUSAL OR CORRELATED?

A final word of caution about statistical analyses and how they are reported: Your evaluator will very carefully choose the right words to describe the outcomes observed and their relationship to the countermeasure implemented. Distinctions will be made between a ~~causal~~ relationship (Implementing A caused outcome B.) and a ~~correlation~~ (A was implemented and B happened, and they appear to be connected.). The distinction is an important one, and should not be lost in the excitement of success. If your evaluator does not use the term causal relationship, it is because she does not believe that a causal relationship can be proven with the data available. Even though correlation is harder to explain than cause, don't undermine the validity of your effort by slipping into sloppy terminology.

During the analysis phase, your main focus should be becoming comfortable with the statistics. Your evaluation specialist will determine what statistical tests, if any, are appropriate. There is no point into going into any detail here on the various tests that could be used and the circumstances under which they are most appropriate. Your evaluation specialist should be able to explain them all to you in terms that you understand.

When you start to get results from your evaluator, there is one thing that you should keep in mind. Just because something is "statistically" significant doesn't mean that it is also "programmatically significant" or meaningful. Meaningful in this context means that both you and your funding sources will be satisfied that the program really made a difference.

Meaningful in this context means that both you and your funding sources will be satisfied that the program really made a difference.

Your evaluator may tell you that there is a statistically significant decrease in the number of repeat DWI offenders following implementation of your mandatory sentencing program. She can report with a high degree of confidence that this change is not due to chance. However, when you look at the actual numbers, you discover that the total number of repeat offenders only dropped by ten. While your evaluator is tickled that she was able to prove that your program was a success, you are worried that your funding source may view this result with less enthusiasm. Trust your instincts. You don't want to be in a position of claiming victory based on statistically significant results that no one else can really see.

Step 7. Report Results.

A successful evaluation is worthless if no one knows about it.

The results are in, and your program was a big success. Before you celebrate, however, you need to pay attention to a very important step in the process. A successful evaluation is worthless if no one knows about it or can understand what is being said.

Your purpose in reporting evaluation results are two-fold:

1) You want to convince your funding source that they should continue funding your traffic safety program, and maybe even increase their support.

2) You want to generate support for your program among the media, the general public and among the other organizations you would like to take a more active role in traffic safety.

As program manager you will need to report your results to your funding source, and to the media, at a minimum. If other organizations were involved in implementation, you should share the results with them, along with appropriate thanks for their participation.

The presentation of your results will vary depending on your audience. You should create a detailed report for your funding source, to convince them that you take evaluation seriously. It must include a short, punchy, Executive Summary which hits the high points and emphasizes the conclusions. The detailed report should include an accounting of how your program funds were spent. The detailed report should follow a standard research format, with the following sections.

- *Table of Contents*

- *Executive Summary*—No more than three pages in length, ideally shorter.

- *Background*—Why the study was conducted and the questions it attempts to answer. It should include the objectives for the program being evaluated and the criteria for success.

- *Methods*—Complete descriptions of the design, procedures, techniques etc. that were used to collect and analyze the data. Questionnaires and data collection forms should be included in an appendix.

- *Findings*—The outcomes of the research presented in tables and graphs.

- *Discussions and Conclusion*—Interpretation of the findings, how they relate to the purpose of the evaluation and the objective of the program being implemented.

- *Recommendations for Action*—Discussion of changes that should be made to the program to increase effectiveness. This section could also include proposals for continued, or even increased funding, based on the results provided.

Your evaluation specialist should be principally responsible for the Methods and Findings sections and should have major input to all other sections.

Your report to the media, and through them to the general public, should be very different. It can be issued as

a press release which specifies what was done, and why, and what the results were. This information should focus on the impact the program will have on the general public. Will they be seeing more enforcement on the street? Will their children be safer walking to school? A clear table or graph of the most significant findings should be included if possible. Your audience will understand percentages use them whenever possible. People also understand the concept of risk when applied to traffic safety. Try to include a discussion of the average person's risk of being involved in a crash, and how that risk may have changed as a result of your program.

Once you have communicated your results to everyone, you need to turn your attention to what changes should be made before you implement the program the next time. You should review all the documentation on what went right, and what obstacles were encountered, so that you can do some contingency planning the next time. You should also review your performance against your budget and milestone schedule to determine if you need to request more money or allow more time in the future. Did you have enough data collectors? Did the media understand what you were doing? Did you get enough cooperation from the local police or school system? All of the factors should be reviewed and built into your planning for future implementation of this same project or any others. Each experience should provide important lessons learned that can save you time money, and frustration in the future.

Each experience should provide important lessons learned that can save you time money, and frustration in the future.

CHAPTER FIVE

GETTING HELP

You're probably thinking that this Guide is glossing over the really hard parts of evaluation just to trick you into taking the plunge. Well you are right—but not because we are trying to trick you. There are hard parts to evaluation and they can create real nightmares if they are not done right. But there is no point in trying to teach you how to do them right in a little guide that you can fit in your pocket.

Evaluation is one area where it makes sense to bring in outside expertise. It will save you considerable time and effort, and will increase your probability of success significantly. Ah, but what about the opening scenario when the manager delegated responsibility for the evaluation to a volunteer from another office and it was a disaster? Obviously, you can't be so eager to delegate the evaluation tasks that you hire the first person that comes along.

What Can (and Should) an Evaluation Specialist Do for You?

You want your evaluation specialist to:

- Design the evaluation
- Recruit and train the data collectors
- Collect the data
- Provide interim feedback to you during the conduct of the program
- Analyze the data and present the findings
- Provide input to you as you draw conclusions

Since the evaluation should be designed right along with implementation, you want your evaluation specialist working with you at the beginning, when you establish your program objectives. Your evaluation specialist can help you focus on what can be measured and what evaluation questions you will be able to answer.

The evaluation specialist will also be able to counsel you about the problems you might encounter gathering or analyzing particular types of data. For example, the evaluator will recognize the potential effect of seasonal differences in driving levels, or the impact political changes might have on enforcement levels. Beyond recognizing the potential problems, however, the evaluator will know how to deal with them.

With an experienced evaluator on board, you will not have to worry about the design of surveys, sampling plans or data collection forms, or the training of your collectors , or the appropriate analytical procedures that should be followed. You of course, need to stay in touch with the evaluation to make sure that the processes and products are in line with your expectations. You will also need to listen to your evaluator when she points out the problems with your carefully thought out plans. An evaluator is trained to be objective and you should take her concerns seriously.

What to Look for in an Evaluator

If you are like most managers, you get a little nervous delegating responsibility for a critical activity to someone you don't know very well. The anxiety increases significantly if the activity involved is highly technical and is outside your area of expertise, because you know you can't just step in and take over data analysis if there is a problem. The way to overcome this anxiety is to have a very clear understanding of what to look for in an evaluation specialist. The following is a list of criteria you can use for selecting an evaluation specialist.

When you are hiring an evaluation specialist, look for

1. Someone who explains things in terms you can understand

   ✓ If you can't understand him or her in the interview, you won't understand the final report. Even if someone gets high marks on all the other criteria, pay attention to this one.

2. Someone who understands characteristics and limitations of traffic safety data

   ✓ Some evaluators waste time and resources trying to make traffic safety data do the impossible. You also don't want your evaluator to become educated on your nickel. You should give your evaluator a copy of the Compendium of Traffic Safety Research Projects 1987-1997 (NHTSA, Document DOT HS 808599) for more detailed information about what has been learned about traffic safety evaluations.

3. Someone with previous evaluation experience, particularly in use of behavioral observations, public opinion questionnaires, and analysis of archival data

   ✓ Read reports your evaluator has written to assess writing skills.

   ✓ Talk to program managers for these projects to assess the evaluator's performance.

   ✓ Look for examples that resemble the type of work you will be doing.

4. Someone who fully understands research design and statistical techniques and when they should and should not be applied

   ✓ Good research design is always needed, even on simple evaluations.

   ✓ Statistical analyses may or not be appropriate, depending on what you are trying to measure. Describe your situation and ask the evaluator what he or she would recommend. Ask for as description of situations which would not require statistical analyses. If the answer is "Statistical analyses are always required." that may indicate a problem.

5. Someone who can present results clearly, both verbally and in writing

   ✓ Ask for a variety of writing samples

   ✓ Ask the evaluator to briefly explain a recent project to you. Watch for the use of technical jargon. Ask a lot of questions to measure the evaluator's patience. This individual may need to represent you someday in a meeting. How will he or she do?

6. Someone who is skilled in presenting quantitative information graphically so that it highlights key issues

   ✓ Ask for a report with a variety of graphs and charts. Are they easy to understand? Do they make things

clearer or more confusing? Ask the evaluator to explain them to you to see how the he or she can communicate complex information in simple terms.

7. Someone who can get access to data

✓ You want your evaluator to make your job easier for you, not harder. A good evaluator will already know how to get access to a variety of data sources. He should not be relying on you to locate the most appropriate sources and to negotiate access to the best sources. If a candidate has never worked with any of the data sources listed below, does he or she at least seem capable of figuring things out independently?

— Archival data (FARS, NASS, State data files)
— Police Reports
— Court Records
— Medical Records
    EMS
    Emergency Room
    Hospital
    Rehabilitation

8. Someone who can get access to data collectors

✓ You want your evaluator to have experience dealing with data collectors and all the challenges they can provide. Potential data collectors could include students, volunteers from the community, and temporary workers.

Links to Sources

Transportation Institutes. Many universities have institutes which specialize in transportation, including traffic safety. A list of many transportation institutes is on the web at http://utop.pti.psu.edu. When talking to them, be sure to find a person who has evaluated safety programs. If you do not have access to the Internet, the University Transportation Centers Program Clearinghouse at (814)863–3614 can give you the information you need.

Governors' Highway Safety Representatives. Each state has an office that reports to the governor on highway safety issues. They may be able to help you find an evaluator close to you. They are listed on the Internet at: http://www.nhtsa.dot.gov:80/people/outreach/safesobr/13qp/resource/resgov.html. (It may be easier for you to connect with http://www.nhtsa.dot.gov and then use NHTSA's search engine to look for "governor." Also try this if the longer link has disappeared.)

NHTSA's Regional Offices. NHTSA has ten regional offices that may know evaluators in your area. They are listed on the web at: http://www.nhtsa.dot.gov:80/people/outreach/safesobr/13qp/resource/resregions.html. (or connect with http://www.nhtsa.dot.gov and search for "region.")

If you do not have access to the Internet, call the NHTSA Auto Safety Hotline at 1-800-424-9393 or 1-800-327-4236. When the recording starts, press 0 to speak with an operator. Ask for the telephone numbers of your NHTSA Regional Office and your governor's highway safety representative. Call early in the morning (east coast time) for the fastest service.

Other Researchers. The Office of Research and Traffic Records at NHTSA keeps a list of researchers who have evaluated safety programs for them. Call them at 1-202-366-4892 to see if one of them is near you.

PLEASE NOTE: Any names received through any of these sources *do not constitute an endorsement of that person or company by NHTSA.* The United States Government does not endorse products (including research reports) or manufacturers (including statisticians and researchers).

✓ When a manager is intimidated by the whole concept of evaluation, he or she tends not to ask the nagging questions for fear of looking uninformed. This can be disastrous! While you are delegating the tasks associated with evaluation, you cannot delegate your own responsibility for managing all aspects of your program. Therefore, it is critical that you hire an evaluation specialist with whom you are completely comfortable. You should feel free to ask any questions that occur to you, no matter how fundamental, and you should understand every answer that is given. If you don't have that relationship with a potential evaluator, keep looking!

Where Should You Look for Evaluation Assistance?

With your criteria for selecting an evaluator in hand, you can begin your search for an evaluator right in your own backyard. You should explore the resources in other departments of your own agency and in other agencies within your jurisdiction. However, don't make the mistake the manager in the opening scenario did. Examine an in-house evaluator with the same rigor that you would an outside consultant. A few college statistics courses do not qualify someone as an evaluator.

There are a variety of other sources of evaluation expertise that you can explore as well. They include:

- Local colleges and universities

  ✓ Check with the Education, Psychology, Business, Public Health, and Epidemiology departments at

nearby universities. The whole discipline of program evaluation originated in the field of education. You should be cautious recruiting in a Math department because of their focus on theoretical statistics rather than applied statistics.

✓ You should expect to pay for the evaluation services you obtain from local universities, but they may be an excellent source of low-cost data collectors and graduate students. You will need to review the procurement procedure in your community to determine how long it might take to contract with the university. Be sure to clearly specify milestones and due dates so that your project doesn't get lost in the shuffle.

- Private consultants

✓ There are numerous consultants in the private sector with evaluation expertise. The firm you hire should have traffic safety evaluation experience, and should be familiar with the archival data sources that you will be using.

✓ Depending on your procurement requirements, you may need to conduct a competition in order to hire a private consultant. This may take several months to complete. Be sure to allow enough time for this process. You don't want your evaluator coming on board several months into implementation.

✓ Private consultants do not have the schedule limitations caused by the semester breaks and summer vacations at universities.

Working with an Evaluator

Before you even start looking for an evaluator, you should prepare a clear statement of work with the specific tasks to be accomplished, a description of all deliverables, and a schedule for their completion. This document will serve as the foundation for your relationship with your evaluator. It should be as specific as possible so that there are no misunderstandings down the road. You will also need to prepare a budget for your evaluation, but this should be negotiated with your evaluator.

It is reasonable for you to expect your evaluator to complete all tasks on time, or to notify you in advance if there will be any delays. Likewise your evaluator should be able to expect you to complete any reviews of deliverables on time. (The schedule for your project should include feasible review periods.) If you have changes that you want, you need to be very specific.

Meet with your consultant regularly, either in person or over the phone. This demonstrates that you are interested in what she is doing and that you want to be involved in any major decisions that need to be made. At the same time, avoid the temptation to micro-manage. You should be focusing your attention on the overall implementation, not the details of the training for data collectors. If there is a problem, your evaluator should tell you about it. Think about how many reports you want your consultant to provide you. Frequent progress reports are essential if they are the only form of communication you have with your consultant, but if you can meet face to face every week, progress reports become just an extra burden that takes time away from your evaluation effort.

CHAPTER SIX

CLOSING COMMENTS

We want to leave you with four final thoughts on program evaluation.

1) It doesn't have to be hard!

If you start out with the intention of keeping your evaluation as simple and straight-forward as possible, you are much more likely to have useable results. Resist anyone who tries to expand the focus or complicate the design. Keep the level of evaluation consistent with the size of the program and the objectives you are trying to meet.

2) It doesn't have to be expensive!

First, re-read # 1, and keep your design as simple as you can. Second, take advantage of the resources that exist in your community. You might be able to convince a university professor to take your evaluation on as masters thesis project. Maybe you can hire an evaluator and recruit volunteer data collectors from local citizens organizations. Work with your evaluator to identify activities on which you can economize, and which areas are worth spending a little extra.

3) Investing in evaluation can save you time and dollars over the long haul!

With the information you learn from a worthwhile evaluation you can focus your resources on the most critical problems and the most effective countermeasures. You will also be able to adjust programs mid-stream to improve effectiveness. And most importantly, you will be much more likely to convince your funding sources that their dollars have been well-spent, which means that you are a good investment for the future.

4) It's never too late to start!

We have spent a lot of time stressing that evaluation should be built into a project right from the start, and not left until the final act of your program performance. However, if you are in the middle of a project right now and are eager to try out your new evaluation mentality, go right ahead. You certainly can check to see if implementation is going as planned and how resources are being spent. An evaluator should be able to help you review what baseline data exists and develop some simple performance measures that you can use to assess how the program did in meeting its objectives. It's even not too late to write some SMART objectives to clarify for everyone what you expect the outcomes to be.

The purpose of this Guide was to convince you that evaluation does not have to be a scary thing. You will only truly be convinced when you apply the information you have read here to evaluate a program of your very own. What are you waiting for?

Once upon a time there was a project manager who was faced with a problem. The head of her department informed her that there were two new projects being planned as part of a national effort to reduce nighttime collisions. Two county supervisors each had their own favorite solution. However, the funding source informed the department that the money they were providing could only go toward one new initiative. The department head refused to choose one project over another without empirical proof to justify her decision. So, the responsibility of pilot testing each approach and recommending one project over another was placed on the shoulders of the beleaguered manager. What a dilemma!

Remembering her training in evaluation management, the manager decided to approach this problem with an evaluation mentality. She was determined to save herself as much wasted time and effort as possible, so she decided to build evaluation procedures into each of the projects right from the start. With the assistance of a carefully selected professional evaluator, she asked five essential questions to put herself in the right mind-frame: "What do I know about the safety problems involved in night driving? What is the objective of each of these projects? How would I measure results? How can I collect the data I need? What are my criteria for success?"

Feeling like they had a firm grasp on each project, the manager and evaluator settled on reasonable objectives for each pilot test according to the SMART guidelines and created a plan for measuring results. They hired assistants to collect appropriate baseline data according to each project's focus. Next, the pilot programs were implemented according to the carefully outlined schedule. In the following weeks, the collected data was analyzed and the report was carefully drawn up. "Hey," the manager said to the evaluator, "with your help, this wasn't as hard as I thought."

The big day arrived. In the conference room gathered the department head, the two supervisors, and the funding representative, all anxious to hear the results. Calmly and confidently, the manager presented her findings. While one approach indicated modest success, she explained, the other program clearly surpassed it, raising safe night driving behaviors by 50%. Impressed by the convincing results, the funding representative heartily agreed to fund the successful project for three years. The department head recommended the manager for a long-overdue promotion. The victorious supervisor patted himself on the back for having though of such a brilliant idea. And even the not-so-triumphant supervisor took the news well, reassured that the outcomes had resulted from an impartial and professional study. Breathing a sigh of relief, the manager thanked her lucky stars that she had used her evaluation training.

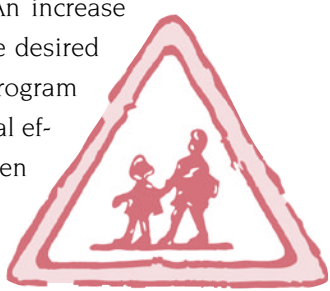
And they all lived happily ever after…

CHAPTER SEVEN

GLOSSARY OF TERMS THAT EVALUATORS USE

Administrative Evaluation (Process Evaluation)—An assessment of the extent to which a program was implemented or conducted according to plan. Administrative evaluations are useful to establish that a program actually reached its intended target audience with the appropriate messages the desired number of times through the selected media. Process evaluations are most useful in troubleshooting unsuccessful programs delivering proven countermeasures.

Before and After Design—An evaluation design that assesses the change in an outcome measure as the difference between pre-program levels and post-program levels. An evaluation of a school- age pedestrian safety program, for example, might observe street crossing behaviors before and after the educational program had been implemented. An increase in the proportion of children observed using the desired search patterns would provide evidence of program effectiveness. This design is sensitive to historical effects, however. If something else happened between the two assessment periods that might affect the observed behavior, then the outcome can not be unequivocally attributed to the program. In this example, the outcome would be confounded if the local news media gave extensive coverage to a child killed or injured by a hit-and-run driver. This design is stronger if a comparison group is also assessed at the same time periods as the treatment group.



Bias—A potential characteristic of non-random samples that affects the program's outcome. For example, an evaluation of a driver improvement

program that is provided to volunteers cannot determine how well the program conveys information because volunteers have different motivations than "average" drivers. Researchers prefer to use random samples whenever possible to avoid bias.

Confounding Factors (or Variables)—Events other than those being investigated that may also have an effect on the outcomes of the program. For example, the results of an evaluation of a speed enforcement program could be confounded by the highway department making engineering changes in the same areas as the enforcement efforts.

Comparison Group and Treatment Group—In order to demonstrate a program's effects, evaluators may compare a group that receives a countermeasure with an equivalent group that does not. The group getting the countermeasure is the "treatment" or "experimental" group and the other is the "comparison" or "control" group.

Correlation—A mathematical technique that assesses the extent to which one variable increases (or decreases) in value as another variable changes in value. Temperature in Fahrenheit and temperature in Celsius is perfectly correlated — as one goes up, so does the other. If one event causes another, they are necessarily correlated, but two variables that are highly correlated are not necessarily causally connected — they might both be caused by a third, unmeasured, variable.

Cost-Benefit Analysis—A process comparing the cost of a program with the savings resulting from the outcomes of

the program. While it is often difficult to identify and enumerate all the costs and benefits, the process can be meaningfully applied to a single program. For example, a law requiring motorcycle riders to wear protective helmets has limited enforcement costs compared with fairly large benefits in health care expenses and welfare benefits avoided.

Cost-Effectiveness Analysis—A process for determining the relative benefit of alternative programs by comparing the amount each program costs with the extent to which each affects a common measure of effectiveness. In this analysis, the outcomes of the program need not be converted to actual dollars saved. In comparing two approaches to increasing safety belt use rates, for example, one could calculate the cost of increasing belt use by, say, 5 percentage points for each program.

Evaluation Design—The plan for conducting an evaluation in a way that permits the evaluator to rule out the possibility that other factors (other than the program) caused the observed outcomes. This plan should include a clear statement of the objectives of the program, how success will be measured, what populations will be exposed to the treatment, how treatment and comparison groups will be constituted, and how the data will be collected, analyzed, and reported.

Field Test—A study of a limited-scale implementation of a new program in a setting similar to where it is likely to be used. Field test sites are generally recruited from candidates showing a high level of interest in participation; a quality that sometimes provides an "ideal" environment rather than a "representative" one. This is not all bad, as

it shows the potential benefit of a countermeasure unfettered by implementation problems.

Impact (or Outcome) Evaluation—An evaluation that determines the extent to which a program achieved its stated outcome objectives. For example, an impact evaluation of a program designed to reduce pedestrian crossings against red lights could compare the observed post-program change in the number of pedestrians crossing on the red and green cycles at selected intersections with an appropriate comparison group.

Outcome Objectives—A specification of the events that would mark the successful achievement of the program's goals. These should be easily and unambiguously measured and closely related to the issues addressed by the program. While all traffic safety programs hope to reduce the number of traffic fatalities, reduction of fatalities is not often closely related to the program's activities. Rather, appropriate objectives should be related to increasing use of safety belts, reducing the number of drinking drivers, improving street-crossing behavior, increasing helmet use, etc. Objectives may specify the populations of interest (e.g., decrease driving after drinking among Native Americans living in Nevada); and, in an ideal world, objectives should state a quantifiable level of change (e.g., increase belt use by pickup truck drivers on 2-lane rural roads in Iowa by 10 percentage points).

Quasi-Experimental Design—A system of procedures for ruling out alternative explanations for study results when study groups could not be constituted by random assignment. While random assignment to groups is the

preferred method for ruling out bias in samples, many real-world situations do not permit random assignment. Consequently, evaluators must turn to other techniques (e.g., additional comparison groups, multiple levels of treatment, comparisons over long time periods) to dismiss threats to the validity of the study.

Random Sample—A subset of a population chosen in such a way that each member of the population has equal probability of selection. Random samples permit the use of certain statistical procedures that provide measures of the potential error in estimates of means (averages) and differences between means of two groups. A simple system for making random selections is to create an alphabetical listing of population members and selecting every nth name. If the population list contained 1000 names and the evaluator needed a sample of 100, she would select every 10th name.

Reliability—An assessment of the extent to which a measurement system will give the same results if used to assess the same events on repeated occasions. A measure can be reliable, however, without being valid. For example, a weekly count of citations for driving while intoxicated may be highly repeatable. However, it is not a valid measure for evaluating a program designed to reduce the incidence of impaired driving because it is so dependent on other factors, including police motivation, program funding, and department priorities.

Representative Sample—A group of individuals deliberately chosen from a particular population to try to emulate the characteristics of the target population as a whole. When random sampling is not possible, use of a

representative sample, with careful attention to defining the relevant population characteristics may be an acceptable option. Focus groups are usually constituted using representative samples. For example, participants may be selected to match the following characteristics: 60% male, 40% female; ages 21 through 30; primary vehicle is pickup truck; drives more than 10,000 miles per year; graduated from high school and attended college for 2 or fewer years.

Statistical Significance—An estimate of the probability that the differences observed between. treatment and comparison groups occurred by chance alone (i.e., that the treatment had no effect). The probability level below which results are said to be significant is somewhat arbitrary, but is usually .05 (5 chances in 100) or .10 (1 chance in 10). Statistical significance can be obtained with extremely small differences if the size of the groups is sufficiently large. While statistical significance can tell you if the results are not likely due to chance events, it cannot tell you if the size of the difference is programmatically meaningful (that is, worth the effort).

Validity—An assessment of the extent to which a measurement system actually measures what it is supposed to measure. For example, observed belt use is a much more valid measure of compliance with belt-use laws than is self-report on a survey. However, there are some circumstances (e.g., nighttime, fogged windows, high-speed locations) under which observations are not very reliable.

DOT HS 808 894
May 1999

U.S. Department of Transportation
National Highway Traffic Safety Administration

www.nhtsa.dot.gov
nhtsa
people saving people