

## **A Quick Look at In-house Evaluations**

Dr. William R. Geary, Ph.D.

One of the necessary challenges for implementing any program is evaluation. Program evaluation is necessary for two primary reasons. First, many programs are funded by outside agencies that require some record of program effectiveness and cost. In other words evaluations are used as source of legitimating the increasing or continuing of resource expenditures. Second, those involved in program implementation have a genuine concern for providing a solid, effective program that attains its goals. Evaluations are challenging for the simple reason that many people involved in program creation/implementation are not trained in the nuances and complexities of evaluation. Accomplished evaluators are trained in many different areas and are often trained “on the job” through basic knowledge of the scientific method and experiences acquired by conducting evaluations. Evaluation requires not only scientific knowledge related to statistics and research methods. It also requires knowledge related to interpersonal communication and the delicate world of political reality. This essay is an attempt to provide those who work in the field with some basic guidelines for planning their own “in-house” evaluations of mediation (or other) programs.

One of the most important points about evaluation is there must be something concrete to evaluate. While this may seem obvious, it is often the case that program goals have either never been clearly stated in writing, or those involved in programs have not specifically internalized these goals. Goals such as “improve parenting skills” or “improve communication” are abstract and should be defined more succinctly for evaluative purposes. For example, a specific goal for “improve parenting skills” may be “provide

parents with alternative methods of dealing with stress”. For the later, “improve communication” may be specified as a part of “improve listening skills”.

An approach for creating measurable goals from abstract statements is to ask the question, “How?” How will the program improve parenting skills (by expanding stress release options) or how will we improve someone’s ability to communicate (by teaching active listening)? As an evaluator I prefer to have these goals written so I can refer to them as the evaluation process unfolds. Written goals also have the advantage of creating an information source that can be referred to by all those involved in the program (administrators, implementers, evaluators, and sometimes clients). An example of an attainable goal in victim-offender mediation (VOM) is “provide opportunities for victims and offenders to enter into either an indirect, or direct dialogue” (Roberts & Masters:3). From this statement a program evaluation would focus on whether or not this goal was obtained. In other words, did the program provide these opportunities?

What may be more challenging, and more enlightening, is to offer a method for providing this service and a reason why this service should be offered. This leads us to the two primary forms of evaluation. Formative evaluations are assessments of how well the process is going. Therefore a formative evaluation for the above-mentioned VOM program might involve how comfortable mediators are during the program or how comfortable clients are with mediators. On the other hand, summary evaluations involve goal-oriented measures. For example, offender recidivism is often touted as an important reason why we should have VOM programs (Nugent et al., 1999: 1). In this case the goal of VOM programs would be to reduce offender recidivism. In this way, evaluation can be

summed up as either evaluating the process of VOM (how do we do it?) or the goals of VOM (what have we done?).

It is important to understand this concept because the “ends” (what) are often affected by the “means” (how). The methods used to provide VOM will, to a large extent, be judged on the merits of the personnel who deliver the program and vice-versa.

Separating these program elements allows for a more accurate evaluation of program/personnel effectiveness. This type of evaluation allows those who run VOM programs to more easily recognize challenges and isolate areas for improvement and allows for an accurate assessment of what is currently working. With this type of information in mind, predictions can be made regarding the affects of new personnel or different clients on overall program success.

Once you have determined precisely what you are evaluating, the question then arises: how do I do it? In part, the answer depends upon how you want to present the data when you are finished. Basically there are two types of data: quantitative refers to “how much or how many of something a value has” while qualitative refers to “what kind, what group, or what a value is referring to” (Bachman & Pasternoster, 1997: 25). Put another way, qualitative data is measured “in kind, that is, non-numerical values” while quantitative data is “measured in amounts, that is, numerical units” (Voelker & Orton, 1993: 162). The Summer 1999 issue of *VOMA Connections* provides excellent examples of each. As cited earlier in this essay, Nugent, et al. examines the effects of VOM on recidivism rates. The four studies cited examine the “relationship between participation in a VOM program and subsequent re-offense within a one-year period” (1). These studies provide examples of quantitative data. The data collected refer to numbers that have real

numerical value, can be rank ordered, and compared by their value. A re-offending rate of 19% is less than a re-offending rate of 28% and is in fact a difference of 9%.

In contrast, Reichwald (1999) concludes her article by stating, “I think participant satisfaction is the final determinant” (10). Participant satisfaction is subjective because the data does not refer to numbers and it is difficult to affix any particular value or measurable point of comparison between answers. We may be able to rank order each answer, for example someone who answers “satisfied” is obviously more satisfied than someone who answers “not satisfied”, but there is no way to know how far apart these two answers are from each other in the minds of the respondents.

For our purposes then, it is safe to say that when a respondent provides answers in numerical form, the data is quantitative. If a respondent’s answer is in words the data is qualitative. Both types of answers are attempts to interpret a respondent’s reality in some way that can be transferred to a mass audience and both processes require a researcher’s interpretation before conclusions are drawn. Finally, both types of data are best interpreted with the aid of some type of control group.

The control group acts as a reference point for comparison. In order to make assumptions about the effect of a program, it is helpful to compare and contrast those who were exposed to the program (experimental group) to those who were not (control group). It is important to keep in mind that control groups should be as similar as possible to the experimental group in order to narrow the effects of outside influences on behavior. The best control groups are identical to the experimental groups in every way imaginable except exposure to the tested program. Researchers often use complex statistical tools to counter the effects of uncontrollable influences on control groups.

For basic, in-house VOM evaluations, the simplest design is a before/after design or a pre/post test. That is to say, clients are compared to themselves at a later time. Of course, no design is perfect, but pre/post tests give evaluators a relatively strong, convenient design. In terms of qualitative data an evaluator would measure attitudes, norms, or beliefs before a client is exposed to the program and again after exposure. For quantitative data, an evaluator could count the number of offenses or problems a client had before a program and compare that to the same behaviors afterwards.

Another possibility is to infer a control group from a larger sample not necessarily related to your program. For both types of data, the control group comes from a different random sample. For example, if I am interested in evaluating a VOM program in Anytown, USA, I could compare my VOM clients to similar clients from that community. More specifically, if I am interested in recidivism rates of VOM clients I could compare them to overall recidivism rates for Anytown, USA, or the country. It is important here to match my clients to the control sample as much as possible. For example, Caucasian males between the ages of 18-25 who have committed minor offenses should only be compared to each other. Sometimes it is difficult to find perfectly matched control samples, but remember, the closer the match the surer you can be of the results. At the very least, it is important to know the differences between your clients and the sample group to whom you are comparing them. This will allow you to place appropriate limits on your data set and conclusions.

Still another possibility is to follow the experimental design carefully and create a matched sample. Matching each client to an identical control subject does this. The match should be identical to your client in as many ways as possible. You would then compare

your client's subsequent behaviors to the match. This method is particularly difficult in the best of circumstances and very rare when resources are limited. Keep in mind that with any control group design quantity does not mean quality. Some characteristics are more important to match than others. The basics in social research seem to be race/ethnicity, gender, income, age, education, offense type, offense history, and perhaps marital status.

Now that you know what you're looking for and whom you are going to study, the next step is how to get the data. Data may come from the client or from those who have information about the client. The obvious method of collecting data from the client is to ask them either orally (interview schedule) or in writing (survey/questionnaire). Both methods require a standard set of questions and some method of accurately recording information. Face-to-face interviews typically require more resources (especially in terms of time) than questionnaires. However, interviews can provide opportunities for clarity that are limited in a written survey. Also, it may be easier for someone to explain ideas verbally than in writing simply because it takes less time. On the other hand, the wide range of variables that affect all face-to-face interactions affects interviews. For example, body language, tone, and personal space all become issues in interviews.

Survey questions are either open-ended or close-ended. Close-ended questions force respondents to answer in a particular way. For example:

Do you feel you were treated fairly by the mediator?

Yes                      No

The respondent must answer in a manner you have designated in advance. More typical are questions using Likert Scale responses:

How often were you treated fairly by the mediator?

All the time

Most of the time

Rarely

Never

Again, the respondent must reply from an already existing choice of responses. Unlike closed-ended questions, open-ended questions provide little to no guidance for respondents:

Please explain how you feel the mediator treated you during mediation.

Open-ended responses provide fewer limitations of responses but are harder to analyze. It is important for you to determine what type of information fits your needs and how your clients will respond to these differing formats. It is also important to keep in mind the time necessary for analysis and presentation. Typically, open-ended questions are much more time consuming in both areas.

If resources are limited, the questionnaire may be best. However, even with limited resources, interviews can still be useful. I have been involved in evaluations where every client is assigned a closed-ended questionnaire and then a random sample of these clients are chosen for open-ended interviews. My opinion is that this system provides a nice mix for gathering data. It has all the comforts of a mass questionnaire with some of the benefits of interviews.

As mentioned, an evaluator does not have to go to the client for information. Official records or other sources can provide corroboration for client data as well as data that stands on its own. It may be important to know how others perceive of the client in order to gain a more complete understanding of the program. Asking co-workers, friends, teachers, etc., can provide insights otherwise not available. Be warned however, to always keep confidentiality in mind. For example, for youth, asking teachers is common. Other sources are police/court records, school records related to attendance/performance,

guidance counselors, etc. Again, be careful with confidentiality here. Typically within this design clients must sign a waiver of release of information unless the source is considered within the public domain.

It is usually best to not rely solely on clients memories for numerical information whenever possible. Clients may simply not remember the exact number of times they committed a crime in the past 12 months or they may give out misinformation. Past field research indicates that client responses are fairly reliable as long as clients are given appropriate time to consider the question and are assured of confidentiality.

Given all of this, I have found that the most common in-house evaluation involves questionnaires. Quite often this is done in a non-experimental method of simply handing out a survey after the program is over. Grant providers generally accept this data collection method as proof of an evaluation. Within this context one is left with non-comparative, descriptive data- how many people answered “yes” and how many answered “no”. Obviously this type of evaluation is extremely limited, but its acceptance is completely measured by your goal. If your goal is merely to maintain funding, this may be adequate in the short term. However, even with minimal evaluation standards, a minimal evaluation may be damaging to the program’s long-term future. Without a more complete understanding of the program’s strengths and weaknesses it becomes more and more difficult to meet client needs. As in any other service organization, the longer you do not meet clients needs the greater the likelihood the program will eventually crumble. On the other hand, short-term successes often add up to long-term successes.

So far, I have gone over the evaluation process in terms of understanding the goal, types of evaluations, categories of data, types of data collection methods, the roles and



types of control groups, and how resources and providers expectations shape evaluation methods and outcomes. I will now briefly discuss some basic information on how to construct an appropriate survey. In researcher's lingo, this is referred to as an instrument.

When evaluating a program it is quite common to use a survey/questionnaire form. The first decision to be made is whether you want to use a pre-existing instrument or one created specifically for this evaluation. In either case, you should collect as many pre-existing instruments as your resources allow. If you decide to use a pre-existing instrument, a variety of instruments to choose from obviously helps find an instrument that best suites your needs. If you are creating your own, then you can use ideas from existing instruments to guide your creation process. In either case there are some basic guidelines to follow for writing a good survey.

The underlying principle for survey instruments is to write it with the audience in mind. To this end, it is important to test the survey on respondents who are similar to the prospective sample in order to receive feedback and make improvements before the actual sampling process occurs. Below is a list of basic guidelines to help you design an appropriate questionnaire:

1. Keep the questionnaire as brief as possible- respondents have a tendency to "zone out" when surveys are too long. At best respondents will not give attentive answers. At worst they will simply stop answering. Brevity requires focusing questions in areas that you are interested in. It is important when creating or deciding on an instrument to constantly refer to your program goals in order to ask the right questions and avoid questions that don't really address your needs.

2. Keep the vocabulary as universal as possible. There are two basic guidelines that apply here: avoid jargon and target your audience. Don't assume that your clients' experiences are similar to each other's or to yours. It is also a good idea to avoid abbreviations.

Finally, try to avoid biased language. The most common mistake involves adjectives used to describe certain groups or individuals.

3. Keep the language as specific as possible. It is always a good idea to receive feedback for test questions in advance in order to avoid questions that respondents think are vague or confusing. It is also important to practice assessing the answers in advance to make sure you can accurately interpret the answers. Specific language provides greater reliability of responses. In other words you want respondents to answer the question you are asking. Within this context it is also a good idea to avoid "double barreled" questions. These questions are confusing because they ask two questions at the same time and you can't be sure which question is being answered. For example: "Do you think the mediator and the mediation was fair?" The answer could refer to either the person or the process.

4. Also avoid double negatives as they can lead to confusion. For example: "Is it true that you 'can't get no satisfaction' " may be an effective musical double negative, but can be confusing when trying to answer accurately and even more confusing in trying to interpret what the answer means.

5. Finally, offer respondents an even number of possible responses to circle. Offering three negative responses and one positive response may result in more negative responses simply because simple chance leads to a negative response 3 out of 4 times. It also provides more negative than positive ideas for a respondent to consider before answering.

This list is by no means complete, but it does provide basic considerations before implementing a survey. Further information on this is available online (for example Frary, 1996) or from various texts on the subject (for example Newman, 2000). In addition, a sample survey on group counseling is available from Umbreit and Fercello (1997) and the results from various evaluations are discussed in Umbreit (1994).

This essay provides a cursory glance into the world of VOM evaluation. The final two steps to this process, data analysis and reporting, are beyond the scope of the essay. Suffice it to say here that analysis involves counting responses and comparing them to a control group. Data presentation is completely dependent upon your organizational goals and your target audience. It is important to keep in mind that no evaluation instrument is perfect for all situations and political/social contexts. A successful evaluation addresses the needs of an individual program and provides answers to questions that allow for program growth. Having said that, there are two primary questions that should run through the mind of an evaluator throughout the entire evaluation process: what are the program goals? And how are client needs unique or similar to clients in other situations?

### **Works Cited:**

Bachman, Ronet & Paternoster, Raymond. (1997). *Statistical Methods for Criminology and Criminal Justice*. New York: McGraw-Hill.

Frary, Robert B. (1996). Hints for Designing Effective Questionnaires. *Practical Assessment, Research & Evaluation*, 5(3). [<http://ericae.net/pare/getvn.asp?v=5&n=3>].

Newman, W. Lawrence. (2000) *Social Research Methods: Qualitative and Quantitative Approaches (4<sup>th</sup> ed.)* Boston: Allyn & Bacon.

Nugent, William R., Mark S. Umbreit, Lizabeth Wiinamaki, & Jeff Paddock (1999). "Participation in Victim-Offender Mediation Reduces Recidivism." *VOMA Connections* Summer 1999:1+.

Reichwald, Sarah (1999). "Defining Success in Mediation." *VOMA Connections* Summer 1999:7+.

Roberts, Ann Warner & Guy Masters (1999). "Introduction To Group Conferencing." In Roberts, Ann Warner & Guy Masters (ed.). (1999) *Group Conferencing: Restorative Justice In Practice Draft*. Minneapolis, MN: Center for Restorative Justice and Mediation.

Umbreit, Mark S. & Claudia Fercello. (1997). "Program Evaluation Kit: Family Group Conferencing." In Roberts, Ann Warner & Guy Masters (ed.). (1999) *Group Conferencing: Restorative Justice In Practice Draft*. Minneapolis, MN: Center for Restorative Justice and Mediation.

Voelker, David H. & Peter Z. Orton. "Statistics." (1993). Statistics. In *Cliffs Quick Review*, Jerry Bobrow (series editor). Lincoln, NE: Cliffs Notes.