james bell associates

EVALUATION BRIEF
Conducting an Outcome Evaluation
December 2008

Overview

An outcome evaluation measures the impact that an intervention or project has on program participants, the community, or even the service provider itself. This evaluation brief will cover the major activities in a typical outcome evaluation. The brief will focus on client-based outcomes, but the principles below can be applied to other kinds of outcomes as well. For grantees in the Healthy Marriage Initiative, client-based outcome evaluations are likely focused on how men's and women's lives, marriages, or relationships were changed for the better as a result of participating in project activities. Similarly, grantees in the Responsible Fatherhood Initiative are interested in measuring changes in the parenting skills and behaviors of program participants. An outcome evaluation is the only way to demonstrate that a project has accomplished its intended objectives and has had an affect on participants.

Getting Started: Defining Outcome Objectives and Using Logic Models

All grantees want their programs to make a difference within the clients and communities that they serve. The first step in this process is to decide what this difference will look like. To do this, you need to have a "theory of change" that describes how your program's outcomes are created by your intervention. This forms the basis for the development of a well-defined set of outcome objectives, which is the beginning of an evaluation plan. Outcome objectives are essentially research hypotheses that state what should happen if the program works as expected. Of course, no evaluation can capture every change that may result from a project. Your outcome evaluation should focus on key expected outcomes. The outcome objectives you choose should be both *appropriate* (i.e. whether the

> **TIPS: Developing Outcomes**
>
> - Be sure that you can tie outcomes to your project and your target population. Participants in a divorce reduction class can realistically expect to increase their knowledge and skills regarding conflict resolution or communication, which could likely result in increased marital satisfaction. In contrast, reducing the divorce rate in a county or state is an admirable goal, but there are many factors that can influence community statistics, and fluctuations in these rates can not be directly tied to one factor.
>
> - Be sure that your project outcomes are not too ambitious. Be realistic about what your project can accomplish and what your evaluation can measure.

project can truly achieve the outcome) and *measurable* (i.e., whether there is a way to determine if the outcome has been achieved).

For example, consider the following outcome objective:

> *Participants in a relationship seminar will experience an immediate increase in their inner sense of well-being.*

This would be a great workshop if this outcome met the two conditions above, but is it most likely that change would be somewhat less immediate and profound. It's also important to consider if such an increase in inner sense of well-being is even accurately measurable? A more appropriate outcome objective could be worded as follows:

> *Participants in a relationship seminar will incorporate conflict resolution skills learned in the seminar into their behavior over time.*

Consider another example:

> *Fathers will reshape their social networks so that they no longer associate with those who encourage or condone irresponsible parenting behaviors such as avoiding child support.*

Although one could argue that friends and family have an important influence on one's behavior, it would be difficult and labor-intensive to explore this outcome. This outcome is also unlikely to be directly related to project activities. An example of a stronger outcome could be worded as follows:
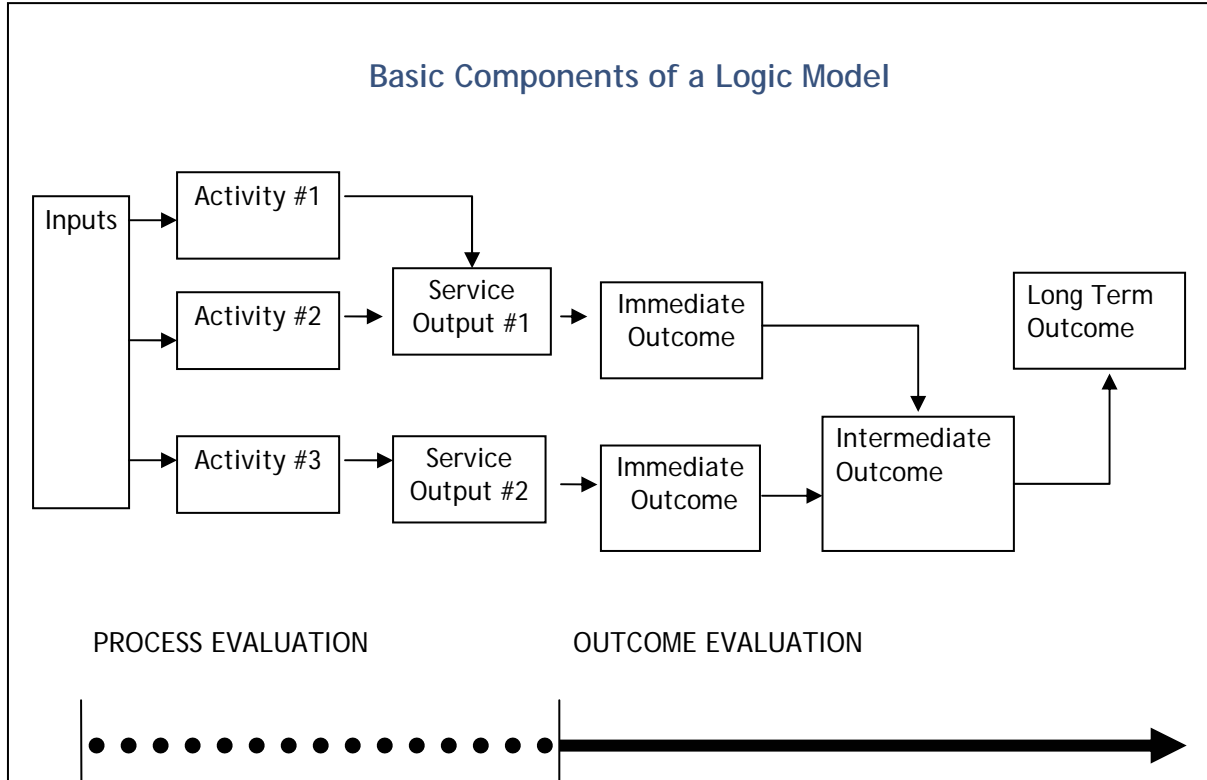
> *There will be an increase in the percentage of participating fathers meeting their child support requirements.*

This outcome is measurable and can cover a range of project activities, including those targeting attitudes and expectations toward parenting, employment, and legal assistance.
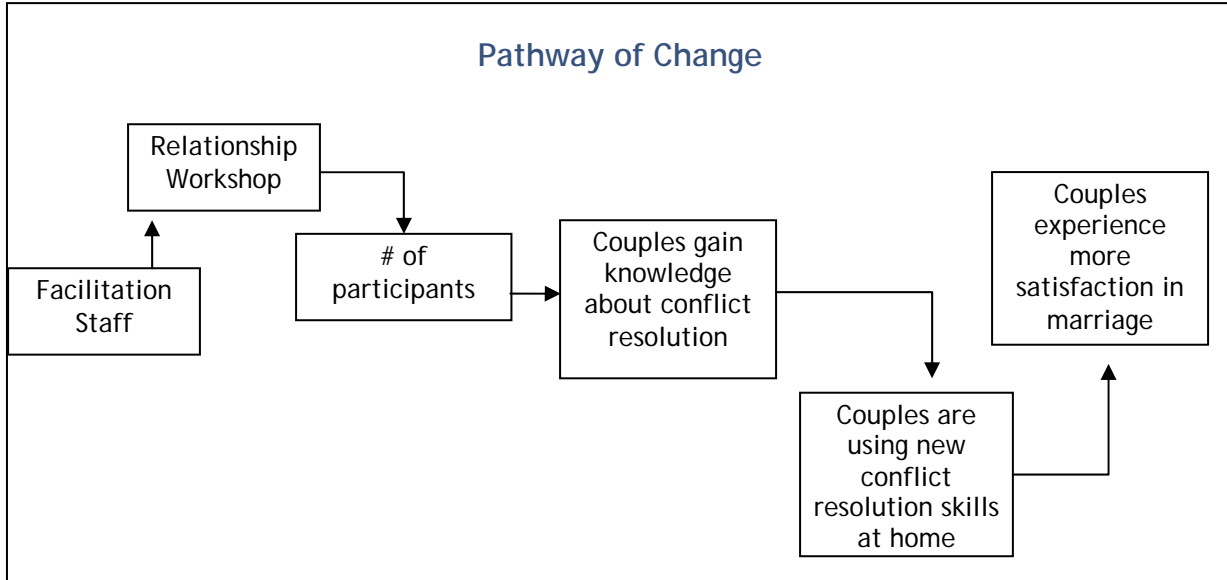
Logic Models

A logic model is a visual representation of the project from inputs to outcomes (see Exhibit 1). The logic model systematically and logically portrays a sequence of events, beginning with project activities and their immediate and tangible effects, and then gradually building towards the intended intermediate and longer term program outcomes. Logic models may reflect one or more pathways of expected change, linking activities to expected outcomes.

Exhibit 1

Basic Components of a Logic Model

Inputs	
Activity #1	
Activity #2	Service Output #1 → Immediate Outcome
Activity #3	Service Output #2 → Immediate Outcome → Intermediate Outcome
	Long Term Outcome

PROCESS EVALUATION   OUTCOME EVALUATION

The pathway of change presented in Exhibit 2 shows that one immediate (short-term) outcome of a relationship seminar might be participant knowledge gain about how to resolve conflicts with their partner in a healthy way. Knowledge is something that participants could gain immediately, but the application of this knowledge might occur in the future. For example, following their participation in the workshop, a couple could be expected to use their newly learned skills to constructively resolve disagreements. This would be an intermediate outcome. The long-term outcome is the realistic result of the behavior change that ultimately might be achieved over time. In this example, the long-term outcome regarding marital satisfaction is expected to follow this initial change in behavior.

Exhibit 2

Pathway of Change

```
Relationship
Workshop ──────┐
               ↓
Facilitation   # of          Couples gain                        Couples
Staff ───────→ participants → knowledge    ─────┐                experience
                              about conflict     │                more
                              resolution         │                satisfaction in
                                                 ↓                marriage
                                           Couples are ──────────→
                                           using new
                                           conflict
                                           resolution skills
                                           at home
```

This process of developing immediate, intermediate, and long-term outcomes should be completed for each key activity implemented through a project. In addition, remember that each project activity will likely have more than one set of associated outcomes.

The Next Step: Outcomes to Indicators
Developing outcomes is a main component of laying the groundwork for your evaluation plan. The final step in this process is assuring that your program outcomes are connected to indicators. The difference between these two terms is explained in the definitions below.

Outcome: The changes in the knowledge, skills, attitudes, behavior, and/or functioning of individuals, families, organizations, or the community as a result of the program.

Indicator: The specific, measurable information collected to track whether an outcome has been achieved (Compassion Capital Fund National Resource Center, 2003).

OUTCOME	INDICATOR
Couples experience greater satisfaction in their marriage	Statistically significant increase between pre- and post-test average scores on the Marriage Satisfaction Scale (Norton, 1983)
Fathers will improve their parenting knowledge and skills	Statistically significant increase between pre- and post-test average scores on the Adult-Adolescent Parenting Inventory (Bavolek, 1979)

Keep in mind that indicators do not necessarily have to be reflected by scores on questionnaires or surveys alone. While standardized measures can provide powerful and straightforward evidence that outcomes were achieved, there are other types of data that can offer more detail and context. For example, change could be measured through participant focus groups or open-ended interviews. See the discussion below for more information on choosing data collection methods and the differences between these methodologies.

Designing an Outcome Evaluation Research Plan

When designing an outcome evaluation plan, you may find yourself walking a line between rigor and practicality. We want to measure the success of our efforts in the most thorough way possible, but staff and clients alike can feel quickly overwhelmed by an overly ambitious evaluation plan. It is best to choose instruments that collect the information needed but do not overly burden clients or staff. By the same token, it is best not to collect more data than you can analyze. An outcome evaluation plan should consist of the simplest plan that will achieve the task at hand.

Using a comparison group
Comparison is a necessary, but possibly complicating, piece of an outcome evaluation that will measure the changes that occur as a result of the program intervention. There are several different ways that comparisons can be made. Some examples of comparative evaluation designs are provided below.

- *Pre-Post Design.* This design involves identifying an "event" that marks the beginning of an individual's participation in the program intervention. Data are then collected before that "event" or intervention begins, which is referred to as the pre-test or baseline assessment. After completion of the intervention, data are collected a second time from the same participants, which is referred to as the post-test or follow-up assessment. The follow-up data are then compared to the baseline data to identify whether participants changed or improved on the outcome measure.

- *Comparison Group.* This design involves the identification of a group of individuals that are "comparable" to the individuals in a participant group, but who have not been exposed to the services or interventions offered to program participants. A comparison group can be identified within the program's agency (i.e., similar individuals who could have benefited from the program but did not) or from another agency or community that does not have the service intervention available. Typically, demographic characteristics and other key variables are examined, such as presenting conditions, to establish the comparability of the intervention and comparison groups. A comparison group may be identified before, during, or after the start of an intervention, and can be created at either the client level (i.e., individuals in the participant group are directly matched and compared with comparison individuals) or the aggregate level (i.e., outcomes for the participant group as a whole are compared with outcomes for the comparison group as a whole).

- *Historical or existing data as a comparison.* When it is not possible to locate a group of individuals that is comparable to the group of program participants, historical data can sometimes serve as a benchmark for comparison. For example, a program

implementing agency-wide practice change could potentially see that all clients served by the agency are exposed to the intervention in one form or another. In this case, a program might rely on data regarding services and outcomes maintained by the agency prior to the changes in practice and compare these to the outcomes observed over time following the project implementation.

Deciding on data collection methods
Deciding upon data collection methods is the next step in the process of designing an outcome evaluation. There are a number of issues to consider when creating a comprehensive data collection plan.

Who will participate in the data collection? Will all program participants be included in the evaluation, or will only a sample of participants be included? Sampling is a good choice if the service population is large, as is the case with several Healthy Marriage projects. Trying to capture data from all participants may be prohibitively difficult and expensive.

How will data be collected? What data collection methods will be used? All data does not necessarily have to be in the form of numbers (*quantitative data*). Focus groups and open-ended interviews provide *qualitative data*. Qualitative data can be beneficial, as it provides richer detail, but it can also make analysis more difficult and may be less convincing to some audiences. In an overall research design, a *mixed quantitative/qualitative approach* is often the best method of data collection. It can capture both sides of program impacts: the quantifiable change, and the less tangible, but richer, aspects of participant experience.

As ethical concerns may arise when conducting a research study, it is important to keep a few issues in mind. Participation in research is *voluntary,* and this should be expressly noted to participants at the time of data collection.

TIP: Simple Sampling

Sample Size: Determining an appropriate sample size does not have to be complicated. There are multiple sample size calculators on the internet that can do the work for you. One useful resource is Practical Sampling by Gary Henry (Sage: 1990).

Simple Random Sample: If the research population is homogenous (i.e., participants are alike in terms of key factors such as race, ethnicity, or class), you may only need a simple random sample. When using this sampling method, the key is to choose a sample that is truly random. One option is to use the software program EXCEL, which can assign random numbers to a list of clients. Once each client has a random number, sort by number (i.e., high to low or low to high), and then choose the desired number of clients that you would like to have in your sample.

Stratified Random Sample: Also called proportional or quota random sampling, this is the best choice for a relatively simple sample when the target population is diversified. This strategy may be effective for programs that serve diverse populations and want to ensure that each population is represented in their sample. In this method, the client population is divided into homogeneous subgroups, and a simple random sample is then taken from each subgroup.

Participants also have the right to be *informed* about how the data will be used and any risks associated with participation. Finally, participants have the right to *confidentiality*. A breach in confidentiality could result in inappropriate use of client data. For example, consider an interview participant from a fatherhood program who admits to the interviewer that he has used illegal drugs in the past. Without confidentiality, it is possible that such information could be released to an employer or an ex-spouse, resulting in negative consequences for the interview participant. In order to ensure that this does not happen, we

must protect all client confidentiality to the greatest extent possible. To do this, data should not be identified by name. Rather, identify data using a number, code, or other system so that respondents cannot be identified, and be sure to store data in a secure location.

When will the data be collected? What will be the schedule for data collection? Will baseline data be collected for comparison? Will there be pre, post, and follow-up data collection? When will follow-up data collection occur? The driving force behind these decisions is the specific nature of the expected project outcomes and the theory of change. There is no "one size fits all" formula, so you must decide when your program can logically expect a change or outcome to occur. For example, will fathers start using newly learned parenting skills with their children in three months, six months, or twelve months?

Putting the Evaluation Plan into Action

Prior to implementation, it may be useful to complete a pilot test of instruments and data collection procedures. By resolving evaluation design issues prior to implementation, you may avoid discovering problems midstream, which could result in compromised data.

Once you reach the implementation stage, remember that there will be inevitable bumps in the road, and that you may run into issues in the future. Thus, it is important to check in periodically to see how things are working, rather than letting the operation run on autopilot. Be reflective and flexible, and if procedures have gone astray, consider the reasons as to why this has occurred. Are response rates low? Do participants seem confused by the instrument? Are questionnaires being administered correctly?

TIP: Response Effects
The manner in which are collected can influence the quality of data gained. Consider how you might respond to questions about the quality of services received if you are asked directly by the person that provided the service. What if the person interviewing you seemed to want a specific answer? Are you more likely to give it? Understandably, participants can be reluctant to tell the whole story in certain social situations. Consider the possible effects created by the race, ethnicity, or gender of the person collecting the data. By the same token, emotionally charged or culturally taboo subjects are not appropriate topics for focus groups. All of these situations can cause *response effects*, or changes in the data caused by the process of data collection itself.

By the same token, do not forget about data storage. If the data that you worked so hard to collect is not stored where it is safe and accessible, all your effort and work may be wasted. The database chosen for your evaluation is dependent upon the evaluation goals and program capacity. SPSS or SAS are good choices when collecting data with the intention of statistical analysis, while MS Word or other processing software would be appropriate for storing qualitative data. Remember to store data without any identifying characteristics, and take care to ensure that that someone stumbling upon the database will not have access to confidential information. Access to the database should be limited and password protected.

Analysis and Interpretation

You will reach the mid-point in your outcome evaluation when you are ready to analyze the resulting data and draw conclusions. Whether you have collected qualitative or quantitative data, analysis must be logical, thorough, and systematic. As is the case in research design, analysis should also be as simple as possible. In fact, be wary of standard statistical tests for significance (f & t tests) when the sample size is relatively small (less than 100). Many projects may only need to report descriptive statistics to demonstrate change (i.e., percentages or frequencies). For larger samples, quantitative data can be analyzed using paired t-tests to determine statistically significant differences between pre- and post-tests. Qualitative data analysis can be accomplished through *content analysis* of interview and focus group data to discover themes and patterns. Those collecting a large amount of qualitative data might consider software such as Atlas ti. and NVivo. These programs can help organize qualitative data, which can be overwhelming in amount and detail.

The final step in an outcome evaluation is interpretation, which holds equal importance as actual analysis. This is an important area of collaboration between the program and evaluation staff. If your outcome comparison revealed a difference, consider the reasoning as to why this occurred. If no difference was found between the pre- and post-test or between the intervention and the comparison group, consider the reasoning behind this as well. Could it be that some segment of the target population experienced a change while others did not, thereby confounding the results of statistical analyses? Could there be something in the process or formative evaluation that can help explain change or lack of change? Analysis and interpretation are both critical issues to discuss with an experienced evaluator.

Reporting and Dissemination

Although there is a tendency to approach reporting as a task done only to meet funding requirements, it can be of great value to programs and evaluators. Reports can inform the direction of a project or the creation of future projects; tell the "story" of the project from implementation to outcomes; demonstrate a need for future funding; and advance any number of fields including sociology, social work, human services, relationship and/or parenting education, and family studies (ACYF, 2003). When writing a report, think about an audience that extends beyond current funders. Consider policy makers, program staff, potential funders, agency directors, other community agencies, other healthy marriage and responsible fatherhood grantees, and family and marriage scholars.

Remember that it is important to report negative or non-conclusive findings in addition to positive findings. For example, the discovery that a particular curriculum does not perform as well as expected with a Hispanic, Korean, or Caucasian middle class population is key information that can help in planning future projects in your agency and would be helpful for other organizations to know. Try and think about the information needs of several potential audiences.

In addition, consider multiple methods of disseminating your findings. Explore the feasibility of dissemination through press releases, newsletters, presentations, or online resources. Professional organizations and journals may or may not be the right venues for your evaluation findings. Perhaps making presentations to local community organizations and service agencies might give you the local impact that you desire. Remember that the evaluation findings of projects in these initiatives are valuable and unique. Many grantees in these initiatives are exploring new frontiers in family support and services, and the knowledge produced through your project's evaluation may benefit many families far beyond the scope of your individual project.

Sources

Administration on Children, Youth, and Families. (2003). *The program manager's guide to evaluation; An evaluation series from the Administration on Children, Youth, and Families.* Retrieved on July 16, 2008 from http://www.acf.hhs.gov/programs/opre/other_resrch/pm_guide_eval/reports/pmguide/pmguide_toc.html

Compassion Capital Fund National Resource Center. (2003). *Intermediary development series: Measuring outcomes.* Retrieved on July 16, 2008 from http://acf.ccfgrantees.org/resources/gbk_pdf/om_gbk.pdf

The National Science Foundation. (2002). *The 2002 user-friendly handbook for project evaluation.* Retrieved on July 16, 2008 from http://www.nsf.gov/pubs/2002/nsf02057/nsf02057.pdf

W. K. Kellogg Foundation. (2004). *Using logic models to bring together planning, evaluation and action: Logic model development guide.* Retrieved on July 16, 2008 from http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf

Weiss, C. (1997). *Evaluation* (2nd Ed.). Upper Saddle River, NJ: Prentice Hall.

For more specific guidance on conducting an outcome evaluation, contact your Federal Program Officer or James Bell Associates for assistance.

About James Bell Associates

James Bell Associates (JBA) is a leading and respected evaluation and research services firm based in the Washington D.C. area. Over the last three decades, JBA has established a national reputation for excellence in evaluation and research focusing on innovative health and human services programs run by Federal, state and local governments, as well as universities, foundations and private firms. JBA is widely known for its highly customized approaches to planning and conducting independent program evaluation and applied research and for offering practical technical assistance (TA) and training to enhance clients' performance measurement capability.

<div align="center">

James Bell Associates
1001 19th Street, North, Suite 1500
Arlington, Virginia 22209
703-528-3230 or 800-546-3230
www.jbassoc.com

</div>