Assessing Project Performance: Building Blocks of Evaluation and Performance Measurement

Thursday, January 19, 2012
2:00 p.m. – 3:30 p.m.

Presenters: Mary Poulin and Carrie Williamson, JRSA; Moderator: Stan Orchowsky, JRSA

Webinar Chat Questions:

1. Could you speak to the utilization of evaluation information – how and when should adjustments be made to services if outcomes aren’t being met?

   Performance measurement and evaluation are useful tools to determine the reasons behind a failure to achieve intended outcomes. Determine what adjustments need to be made by examining these elements:
   
   • Implementation—Did program activities follow the program plan? Did the youths receive the intended dosage?
   
   • Data collection—Did you measure program activities and outcomes accurately and consistently?
   
   • Program logic—Why did you think program activities would cause the outcome? Is there research evidence to support this assumption?
   
   • External factors—Did external factors beyond the scope of your program prevent program activities from having the intended effect? How can the program adjust to avoid the effects of these external factors?

2. Please clarify – are there three types of evaluations: process, outcome, and performance?

   Process and outcome are two different types of evaluation.

   Process and outcome are also two different types of performance measures.

   Both process and outcome evaluations assess program performance, but process evaluation is focused on the implementation of the program and outcome evaluation is focused on the impact of the program or policy.
Performance measurement is a component of evaluation, but it is not a form of evaluation. Performance measure data depict the activities and objectives being accomplished by the program (process) and the characteristics of participants the program hopes to impact (outcome). A true evaluation will demonstrate that program activities caused any change seen in program participants.

3. **Defining “program” evaluation is tough. If we are providing probation supervision services to a population, do we evaluate each individual service or intervention or “probation” as a whole?**

The focus of an evaluation depends on the questions you want to answer about your program and the data available for the evaluation. For example, if you want to know if intensive supervision has a greater effect on recidivism than traditional probation, you would want to examine supervision as a whole, but examine variation in outcomes by type of service received.

4. **At my facility, evaluation of programs is done is using quasi-experimental designs. When results are questioned, executive administrators become openly demeaning and hostile to the line of questions. How does one handle this type of environment?**

This is a difficult situation. It is important to be clear about what data were collected, how they were collected and the answers they were intended to provide. A couple of suggestions: 1) Be transparent about how you arrived at certain conclusions. For example, rather than saying “the program was not successful” say, “we defined program success as….” According to this definition, X% of clients successfully completed the program.” 2) Consider working with the executive administrator prior to the release of the results to help make decisions about how to interpret data.

5. **There has been a shift in our state with agencies choosing to conduct their own evaluations internally. Can you speak to the value of using external evaluations to conduct an evaluation of a program?**

It is certainly possible to conduct a high quality internal evaluation of one’s own program, but there are several advantages to hiring an external evaluator. An outside evaluator brings a fresh perspective to the program and potential ways to improve both operations and data collection. In addition, if you intend to use the evaluation to demonstrate effectiveness to funders or other stakeholders outside of your agency or organization, an external evaluation is likely to be viewed as more objective and credible. However, the primary benefit of having an internal evaluator is having extensive knowledge of the program before the evaluation begins.

6. **How can drug and alcohol treatment programs be evaluated in situations where the participants do not have access to drugs or alcohol?**

In this type of situation, consider what can reasonably be achieved and measured by the program, and think outside the box about the types of performance measures you can collect to assess program achievement. One objective of a drug and alcohol treatment program might be to change attitudes and perspectives about drugs and alcohol, and to
increase participants’ knowledge of ways to avoid consumption of drugs and alcohol. These outcomes can be measured in these circumstances. Also, consider measuring the use of alcohol and drugs after the participants are no longer participating in the program. This measure of follow-up provides evidence about the duration of program effects.

7. **How does a quality assurance process fit in with performance measurement and evaluation?**

It is tough to answer this question without knowing more specifics about your quality assurance process, but it is likely that this process complements performance measurement and evaluation. Quality assurance and performance measurement have a common goal: tracking the activities of a program, ensuring that these activities are following a plan, and improving the execution of these activities. One key difference might be the comprehensiveness and consistency of the data collected, in that performance measurement should cover program activities from beginning to end, while quality assurance may involve random checks or site visits. Further, performance measurement and evaluation ask questions about the outcomes of the program. Quality assurance usually only focuses on implementation.

8. **How do you link program activities to outcomes?**

Conducting an evaluation is the only way to demonstrate with certainty that program activities caused program outcomes.

However, one can make reasonable assumptions about program effectiveness if the program follows an evidence-based model, adheres strictly to the program plan, and high-quality performance measurement data are collected.

9. **Good, true evaluation is expensive. Can you help ensure that grant solicitations do not ask for evaluation (as opposed to performance measurement) without enough dollars available in the award to cover evaluation?**

We are not responsible for writing grant solicitations, but recommend that you contact the granting agency when a solicitation requests an evaluation to ensure what their expectations are regarding evaluation activities. Sometimes agencies may use a template for some of the information that goes into a solicitation, and depending on the solicitation the request for evaluation may be more or less salient. Also, it is not uncommon for solicitations to use the term evaluation when they are referring to performance measurement. For this reason it is always important to seek clarification from the agency who posted the solicitation.

10. **How does evaluation work with prevention programs?**

Evaluating a prevention program is a complex endeavor, because it is difficult to measure something that doesn’t happen. However, there are still some good options for assessing prevention program performance.
One way to show program achievement is to follow an evidence-based program model and demonstrate that the program is meeting its objectives. Remember that objectives are the measureable accomplishments of the program that act in furtherance of the ultimate goal or mission of the program. For example, if research shows that increasing youth awareness about cyber predators reduces the likelihood that youth will be victimized by someone on the internet, you should focus on collecting performance measurement data that assesses the degree of knowledge about this topic before and after participating in your program.

Another way to evaluate or assess prevention program performance is to compare rates of a certain behavior or condition among program participants with similar individuals who have not received the treatment or intervention. For example, if your program seeks to prevent a behavior that is fairly common (such as underage drinking), you might compare rates of alcohol consumption among program youth with rates among a group of similar youth who did not participate in the program.

11. Can you please give an example of a well-written performance measure? In my experience providers use “the count of service provided” as an appropriate measure; this does not show how well the program is executed or if they are following the plan (fidelity).

You touch on an excellent point, in that a simple count of program participants doesn’t tell us much about who the individuals are who are being served, if they meet the specifications of our target population, what services are being provided, and what the time frame is for consideration. It is important to be very specific when developing performance measures for your program so that the information you collect provides a clear picture about WHAT you are doing and WHO you are serving.

Here are two ways to improve on your example:

1) Number of 9th through 12th grade youth enrolled in 2012 spring semester mentoring program

2) Number and percent of 9th through 12th grade youth matched with community service opportunity June-August of 2012

For more information, attend the Introduction to Program Logic webinar February 22nd. We will focus on developing good process and outcome performance measures and improving on existing measures.

12. Does OJP take first generation immigrant populations into consideration for evaluation questions that are potentially incriminating? For example, we know that some schools are definitely populated with juvenile delinquents, etc., but that is not showing in the scores because students are not indicating so on their surveys, perhaps because they fear very real consequences (deportation) if they’re found out. We suspect students are artificially depressing their scores.

It is always a concern that individuals are not responding truthfully to questions, particularly those that may be seen as incriminating. There are a couple techniques to
try to address this: if identifying information (i.e., youth’s name) is not needed, don’t ask for it and provide privacy to youths when completing surveys; have the surveys administered by a third party who don’t know the youths and will not share the responses of individual youths (i.e., they will only share group results); and tell the youths the purpose of the survey and how the responses will be used.

13. How critical is measure of ROI for federally-funded program evaluation?

This is a difficult question for us to answer. Agencies within the Office of Justice Programs have carried out and funded a variety of projects that have incorporated cost-benefit analyses. Information about this work is available at the National Criminal Justice Reference Service.

14. How do we find the OJJDP performance measures – DC TAT, I believe they are called?

The OJJDP performance measures can be found at https://www.ojjdp-dctat.org/.

15. How can I perform an evaluation without hiring an outside evaluator?

We would suggest that you not attempt to conduct an internal evaluation if you do not have program staff with the knowledge and skills to conduct an evaluation. Rather, consider doing performance measurement.

16. Will you elaborate on the benefits of the logic model?

A logic model connects the goals, objectives, activities, and performance measures of a program, and shows HOW the program should operate and WHY we think it will have the intended effect. Logic models are an excellent way to demonstrate what you’re doing to outside stakeholders, particularly funders. Logic models also provide the opportunity for those involved in program operations to assess what your program is actually accomplishing, particularly if the program involves multiple sites, agencies, or organizations.

17. Can a program have program measures, but can’t be “evaluated”? For example, I have short-term secure detention facility. A measure is the reduced number of assaults, but the outcome cannot necessarily be “no re-arrests”. Comment?

Whether or not participants are re-arrested could be a potential outcome for your facility, depending on the program objectives. Though there are some factors that may make a program easier to evaluate, any program can be evaluated. If you need more program-specific assistance, fill out the Training and Technical Assistance form on the NJJEC website.

18. How can you tell if the change in crimes pre-program and post-program is attributable to the program?

A true evaluation must be conducted to establish with reasonable confidence that changes are attributable to the program. Evaluation extends the information provided by performance measurement by ruling out other explanations for the change.
19. If data related to conditions was not tracked before the intervention, how can they be ruled out during a causal analysis? (e.g. say for some reason race, SES, or family make-up were not tracked for the target population before the intervention).

You may be unable to rule out the effects of these conditions or factors if you do not have relevant data collected before an intervention or program was implemented. Be sure to check for outside sources of data that might be usable, such as arrest records for a neighborhood or attendance records for youth.

20. How do you use baseline data when working with youth who may be more engaged in risky behavior as they age, i.e. middle school students are more likely to use drugs/alcohol and be sexually active in 8th grade than in 6th grade regardless of program intervention.

Evaluation methods can account for the effect of age on risky behaviors, but as we know, conducting a true evaluation is not always feasible.

Collecting baseline data from a control or comparison group is another way to show that age is linked to the prevalence of risky behaviors. For your example, you should collect baseline data for both your program participants (the treatment group) and a second control/comparison group of youth who are similar to the program participants but are not actually involved in the program activities. It is very important that the comparison/control group youth are similar to the treatment youth, particularly with regard to the prevalence of risky behaviors in 6th grade.

Collect information on these risky behaviors in both 6th grade and 8th grade for both groups. The comparison/control group youth who are not involved in the program will give you a good idea of the natural progression of these behaviors when the youth do not receive the intervention, and allow you to show that rates of risky behavior of your program participants in 8th grade differ from the comparison/control group youth in 8th grade.

21. Does OJJDP accept studies that use single subject research design?

We’re not quite sure what is meant by “accept studies,” but it is our understanding that OJJDP has funded evaluations that use quasi-experimental designs incorporating pre- and post-tests.

22. Can you talk about “fidelity” as it relates to implementing an evidence-based program in a new location? Local situations may make it difficult to implement the program the way it was originally designed. How do you get at that type of fidelity?

This is a very important issue in implementing evidence-based programs and practices (EBP). It is important to maintain the key components of the EBP that are expected to affect outcomes for the target population, but there are many situations in which changes to an established evidence-based program or practice may be necessary and advantageous for the target population or context. In this situation, it is important to
clearly explain the reasons for making adjustments to the original program plan, and to collect good performance measure data.

For more information on this topic, attend the Advanced Program Logic webinar February 23rd, 2012.

23. Please briefly define/distinguish between evaluation and performance measures.

Both evaluation and performance measurement can be used for monitoring and program improvement, but the key difference is that performance measurement generates evaluative information. Performance measurement tells us about program activities and the state of our target population before and after program execution, while evaluation establishes a cause-and-effect relationship between these activities and any changes that occur in the target population.

24. What tools do you recommend or have available on the website that can be used as guides or examples of effective program evaluations?

There are a number of useful program evaluation resources under the EBP and Evaluation Resources tab on the NJJEC website, including links to national organizations that conduct program evaluation or compile high-quality research. The Evaluation and EBP in the States section provides a number of state juvenile justice agencies conducting program evaluation and performance measurement. A tutorial created by the Juvenile Justice Evaluation Center is available on the NJJEC Trainings and Presentations page, and an updated version of this tutorial will be available soon.

View and subscribe to the NJJEC Bulletin for information about research of interest, new resources available, and conferences and trainings focused on juvenile justice program evaluation. There are also several publications available under the NJJEC Publications tab that might be useful for you, including Juvenile Justice Program Evaluation: An Overview, Hiring and Working With An Evaluator, Strategies for Evaluating Small Juvenile Justice Programs, and Evaluability Assessment: Examining the Readiness of a Program for Evaluation.

25. What about external factors – how might they impact participant outcomes?

Accounting for external factors is the key difference between performance measurement and evaluation: Evaluation rules out the impact of these external factors (factors outside of the program that may affect program outcomes) on outcomes, and demonstrates that program activities caused the expected outcome. Evaluation may demonstrate that external factors reduced the strength of the program’s effect or prevented it from having the expected effect at all, or may show that even though the program appears successful, some outside influence actually caused the change in participants or the community.

26. Is there online support for those that will be putting together project performance measures?
NJJEC has several resources available to aid in the development and collection of performance measures on our website, including publications, briefings, and links to additional resources. You may find our presentation *Evaluation and Performance Measurement 101: An Introduction to Key Concepts* useful. The presentation includes guidance on developing good performance measures and improving existing measures. If you need more program-specific assistance, fill out the *Training and Technical Assistance form* on the NJJEC website.

Also, be sure to attend the OJJDP/NJJEC March 22nd webinar on data collection and analysis!

**27. How does the size of the sample affect evaluation? What formula is typically used to determine a statistically significant sample size?**

The sample size needed to conduct an evaluation depends on the population size, because an evaluator needs enough individuals in the sample so that it is able to be representative of the entire population. For example, a sample of 100 people is likely to be a sufficient representative of a population of 150, but it is not enough people to represent a population of 1,000. See JRSA’s new BJA-funded publication entitled, “Is This a Good Quality Evaluation Report? A Guide for Practitioners,” available on the NJJEC website or the JRSA website (www.jrsa.org), for more discussion about sample size.

**28. How is a logic model different from a strategic plan?**

A logic model may be one component of a strategic plan, but a strategic plan is a much more extensive endeavor. A logic model is a visual depiction of a specific program, practice, or policy you intend to implement or are currently operating. Its purpose is to explain how the components of a program fit together—how activities affect objectives and goals, and the ways you will measure the program’s accomplishments. Typically, a strategic plan relates to an entire organization which may comprise several programs or projects. It helps the organization assess where it is now and envision where it wants to be in the future. It includes additional elements such a vision statement, broad funding and policy decisions, explanations for why these decisions are being made, and the expected impact of these decisions. While a logic model gives a quick glance at the operations of a program, a strategic plan provides a more comprehensive and future-oriented look at an agency or organization’s goals and the methods to achieve those goals.

**29. Is there a template available that has the basic components [of a logic model]?**

There are many ways to structure a logic model, and the components depend on the complexity of your program and the audience for your logic model. At a minimum, we recommend including the goal, objectives, activities, and performance measures (process and outcome) in your logic model. This sample logic model is available on the OJJDP website, and we will include several other styles in our discussion of logic models during the Introduction to Program Logic webinar February 22nd and Advanced Program Logic webinar February 23rd.
30. If you are delivering a program to a populations that can’t be denied services, what other techniques do you suggest so that you can set up a scenario where one group receives services and one group does not?

If there are more individuals than can be provided the service, treatment, or intervention, those awaiting enrollment or participation can serve as a comparison group for your study. If you are able to have a comparison group for this reason, there are statistical techniques that an evaluator can use to account for differences in group size.

31. How can one evaluate a program by using qualitative methods efficiently, particularly participant observations?

For information on qualitative research, check out the following resources:

- [http://www.socialresearchmethods.net/kb/qual.php](http://www.socialresearchmethods.net/kb/qual.php)
- [http://learningstore.uwex.edu/assets/pdfs/G3658-12.PDF](http://learningstore.uwex.edu/assets/pdfs/G3658-12.PDF)

32. Is there a distinction between a “comparison” group and a “control” group?

There is a slight distinction. In both cases, the comparison/control group is used in contrast to the treatment group, which is the group of individuals receiving program services or some type of intervention.

Usually, we think of control groups in the case of experimental designs involving random assignment, and comparison groups in the case of quasi-experimental designs. The key difference is that a comparison group is considered similar to the experimental group, but because of the lack of random assignment, we cannot be completely sure that the group receiving the treatment is similar to the group not receiving the treatment (i.e., the comparison group).