Research's Two Modes for Discovering Evidence-Based Programming

By Howard N. Snyder

In recent years, researchers have attempted to summarize various bodies of program evaluation research in an attempt to answer the practitioner's question, "What works?" A paper by Mark Lipsey, co-director of the Center for Evaluation Research and Methodology at the Vanderbilt Institute for Public Policy Studies, presents some insights that all practitioners should know and understand in this new world where everyone is seeking the Holy Grail of evidence-based programs.[1]

For those not familiar with the term (and this probably would have been almost everyone 15 years ago), a prototypic evidence-based program is one that has been rigorously tested and has been shown to have a meaningful and statistically significant effect on the outcome for which the program was designed (e.g., reducing recidivism). From a researcher's perspective, rigorous testing involves the use of experimental methods:

- defining the program in great detail (e.g., staff skills and duties, target clients, screening and assessment instruments and service-delivery protocols/schedules);
- mounting the program independently in more than one location;
- evaluating each site's implementation and outcome using an experimental design that incorporates adequate control groups developed through randomization or matching techniques — and when control groups are not possible, the use of appropriate statistical controls; and
- then, if the evaluations from the different sites are similar and show that the program model is effective, the program may use the label evidence-based program.

The claim is that if such programs are implemented faithfully with similar populations in similar settings to those used in the validation studies, implementers can confidently expect that the programs will be effective, eliminating (or the least reducing) the need for an outcome evaluation. However, if the program is implemented in a somewhat different setting or a somewhat different population, or if the original program changes due to local pressures, a prudent program director should not assume the program will be as effective as the original proven program model.

There are few programs that have achieved this level of endorsement. In the field of juvenile justice, for example, Delbert Elliott and his colleagues' comprehensive review of hundreds of evaluation studies has identified only about 10 programs that they could label evidence based (what they have called the Blueprints Programs).[2] A more general review of the evaluation literature in the criminal justice field, prepared by Lawrence Sherman and his colleagues, found that some correctional efforts pass the scientific test of effectiveness (e.g., drug treatment in prison-based therapeutic communities that operate as 24-hour live-in facilities within the prison) while many others (e.g., correctional boot camps using the old-style military model) do not.[3]

Lipsey points out that evidence-based programs have many advantages. The programs are specific and well defined, and there is someone to contact (generally the developer) to discuss operational details. But there are also drawbacks with evidence-based programs. The supportive research is generally a small number of studies, often conducted by the developer, on demonstration programs implemented primarily for research purposes. In the real world, it is often difficult to mount evidence-based programs faithfully due to budget limitations, skills of available staff, populations that must be served, or other modifications needed to meet local constraints. If the recipe for the evidence-based program is not followed precisely, the assumption that the program will be effective (or as effective as advertised) is challenged.

So, if this were all there was to evidence-based research, the discussion of evidence-based programs would often lead to practitioner frustration and be another example of the research community's inability to address real-world concerns. But Lipsey points out that research is being used to help practitioners when strict adherence to a well-defined evidence-based program is impossible or impractical. This research pathway is based on the assumption that knowledge about effective programs or program components can be gleaned from evaluation studies of disparate programs that share one or a few structural components. This body of research has adopted the daunting label of meta-analysis.

Assume a social service agency is offering to provide a mentoring program to its local juvenile justice system. Before accepting this offer, a practitioner should want to know what existing research says about the effectiveness of mentoring programs and for what type of youths they are likely to be most successful. To do this, the practitioner should check the literature to see if a meta-analysis has been conducted on mentoring programs. While the math behind a meta-analysis may be complex, its logic is straightforward. To do a meta-analysis, researchers collect all of the evaluation studies that looked at mentoring programs (or programs that involved some form of mentoring) as well as other programs that have been used on the populations of interest. The researchers would classify these programs on a range of attributes and the various effects of these programs on the youths involved. Then, through the mathematics of meta-analysis, the researchers could deter-

Through experimental studies and meta-analyses, researchers are trying to help practitioners — and at the same time show the practical value of their research efforts.

mine if generic mentoring programs (with all of their natural site-to-site variations) had a positive effect on the population in the juvenile justice systems studied, or on some identifiable subset of the population.

As Lipsey points out, compared with the clear guidance to program design provided by experimental efforts, the results of a meta-analysis provide practitioners with less distinct blueprints for implementing an effective program, often with no authoritative source to ask about how to conduct the program. However, in contrast to the experimental work, meta-analysis bases its findings on a larger body of evaluation research, covering a range of program variations, participants, settings and researchers. If the meta-analysis finds that a program or a program component is effective across such diverse program implementations, then practitioners can be reasonably assured that such efforts will be effective despite inexact replication of a given model. And if the meta-analysis finds differential effects across implementation sites, then it may be possible to determine which population or program characteristics are likely to yield the most positive results.

Through experimental studies and meta-analyses, researchers are trying to help practitioners — and at the same time show the practical value of their research efforts. As Lipsey concludes, "The simple truth is that we do not yet know much about how best to translate research findings into advice that practitioners can and will actually use effectively."

Developers and implementers of experimentally-supported evidence-based programs have been very energetic in advertising their products to the field — for obvious reasons. And, given the high visibility of these programs, when evidence is developed that questions a program's effectiveness, it tends to find its way into the general debate (remember Drug Abuse Resistance Education — or DARE). One thing that is lacking is a platform from which the findings of meta-analyses can be either broadcast to interested practitioners or centrally housed. Given the "black box" character of meta-analyses, wherever this platform resides, its contents will have to be carefully reviewed by those who can judge the rigor of the meta-analyses to ensure that the conclusions are grounded in quality research. It is important that this be done because, through meta-analysis, knowledge contained in existing evaluation research efforts can be distilled into useful information that can improve the justice system's responses to crime and offenders. More validated research findings could and should be more influential in guiding correctional policy and practice. In addition, it is important for the corrections field to be as committed to developing a knowledge base (perhaps through meta-analyses) of effective implementation principles, with steps tailored to the scale and complexity of the organization/agency undertaking the development or seeding of a particular program.

ENDNOTES

[1] Lipsey, Mark. 2005. The challenges of interpreting research for use by practitioners. *American Journal of Preventive Medicine* (Supplement 1), Volume 28 (2), 1-3.

[2] Mihalic, Sharon, Katherine Irwin, Delbert Elliott, Abigail Fagan, and Diane Hans. 2001. *Blueprints for violence prevention.* Washington, D.C.: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

[3] Sherman, Lawrence, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn Bushway. 1997. *Preventing crime: what works, what doesn't, what's promising.* A report to the United States Congress prepared by the National Institute of Justice. Washington, D.C.: U.S. Government Printing Office.

Howard N. Snyder, Ph.D., is director of Systems Research at the National Center for Juvenile Justice and holds the Center's Maurice B. Cohill Jr. Juvenile Justice Policy Research chair. He received the Peter P. Lejins Research Award in 2004 and is currently the chair of ACA's Research Council.