

# **The Evidence-Based Practice Movement: Contributions, Controversies, and Recommendations**

**Author:**

Dr. Edward W. Gondolf, Former Director and Research Associate  
Mid-Atlantic Addiction Research and Training Institute  
Indiana University of Pennsylvania  
Indiana, PA 15701

**Contact Information:**

Edward W. Gondolf  
33519 Inverness Dr.  
Evergreen, CO 80439  
Phone: 724-388-0981  
Email: [egondolf@iup.edu](mailto:egondolf@iup.edu)

*Emerging Trends in the Social and Behavioral Sciences*  
Robert A. Scott, Stephen M. Kosslyn, and Nancy Pinkerton, Editors  
John Wiley & Sons Publisher, Hoboken, New Jersey, 2015.

## **The Evidence-Based Practice Movement: Contributions, Controversies, and Recommendations**

### **Abstract**

*The evidence-based practice movement, particularly in the criminal justice field, has meant an increasingly influential role for social science research. Experimental program evaluations, considered to be the “gold standard,” are helping to assess the effectiveness and efficiency of interventions amidst the need to cut costs. However, there continues to be questions about the implementation and conception of experimental designs in the “real-world,” and to be resistance to such program evaluations from many practitioners. Several remedies have emerged including statistical modeling, multiple methods, and consensus panels to promote broader dialogue regarding program effectiveness. The ideal maybe to return evidence-based practice to a more collaborative process rather than a bottom-line verdict.*

*Keywords:* evidence-based practice, program evaluation, experimental field studies, meta-analysis, criminal justice research, offender rehabilitation

### **Introduction**

One of the significant social science trends in the last decade or so has been the increasing prominence of social science in determining intervention programs and policy development. Applied social science, of course, has had a long-standing role in these arenas, but often been diffused by politics, irrelevance, or impracticality. That role today has been heightened in child welfare, substance abuse, homelessness, impoverished family, and domestic violence cases, however, because of the increasing attention to effectiveness and efficiency—and the pragmatism that underlies that. It has been particularly pronounced in the criminal justice field given the soaring cost of intervention amidst contracting state budgets.

This trend has been encapsulated in what has become known as the evidence-based practice movement. At its heart is the call for experimental program evaluations, similar to those used in the medical field to test the effectiveness of medications and medical procedures. The experimental evaluations compare the outcome of subjects randomly assigned to a treatment (or experimental) group and to a non-treatment (or control) group. In accord with basic scientific principles, this sort of design brings us closest to attributing the cause of an outcome to the treatment or intervention, independent of subject characteristics or other mediating factors. The results then give us “evidence” as to whether a program is effective in its aim of ameliorating a certain set of behaviors or not.

This application of social science can help determine which programs warrant implementation and referrals, endorsement and promotion, and funding and other resources. It can also aid in refereeing competing approaches making “success” claims to potential clients. In these ways, experimental evaluations can help bring greater consistency in practices and programs, and offer accountability to clients, the public, and funders. I am most familiar with domestic violence intervention in the criminal justice field and can attest that evidence-based practice is having a major impact on court referrals, funding allocations, program standards, and rehabilitation approaches

for counseling and education programs receiving court-mandated offenders, often referred to as “batterer intervention.”

A recent special issue of a criminology journal devoted to evidence-based practice summarized the extent of the movement this way: “The emergence of the evidence-based movement is arguably one of the most significant developments to occur in criminal and juvenile justice over the past 20 years.” However, the author adds, “It also would be in error to assume that the evidence-based movement has been embraced unconditionally or universally in the research community” (Prybylski, 2012, p.1,7). Many practitioners tend, furthermore, to view evidence-based practice as disruptive and imposing.

In this paper I review the nature and contributions of evidence-base practice in more detail, with reference to its relationship to criminal justice intervention. The methodological and conceptual issues associated with the research underlying evidence-based practice are then discussed, along with the resistance from practitioners to the implementation of such practice. But rather than dismiss evidence-based practice, I conclude with recommendations for conducting, applying, and furthering the social science on which it is based, and for reconciling the concerns of researchers and practitioners who question some apparent misuses and their impacts.

### **What is Evidence-Based Practice?**

The concept of evidence-based practice was actually introduced in the early 1990s into the medical field with the explicit mission of bringing greater consistency to medical treatments, medications, and procedures (Gilgun, 2005). Physicians were obviously educated at different times and under different philosophies, and beholden to their own theories and, in some cases, outdated approaches. The introduction of evidence-based practice was a way to help standardize and invigorate practice. The objective was to develop a feedback loop, of sorts, among researchers and practitioners. Practitioners posed questions about effectiveness that researchers investigated; the results were interpreted and applied by the practitioners, and new questions posed.

Admittedly, the intended *process* evolved into a more researcher-directed format in order to ensure greater objectivity, rigor, and focus in the research. Randomized clinical trials (RCT) have, as a result, become more the norm for medical research. (In RCT, subjects are randomly assigned into different treatment models or medication options and a non-treatment or placebo option.) This type of experimental design may, furthermore, include a double-blind condition in which both the patient and the physician are unaware of which medication is being administered. In this way, potential bias or influence of the physician is controlled. Researcher-practitioner partnerships and collaboration, of course, continue to be valued and even necessary, but to a lesser degree than in the process-orientation of the initial conception.

For these reasons, experimental program evaluations have been dubbed, “The Gold Standard,” and are emulated by social scientists, criminologist, and policy makers as the ideal for “evidence-based practice,” as well as by medical practitioners (Dunford, 2000; Sherman, 2009). An early as 1996, the National Academy of Sciences concluded that “randomized, controlled outcome studies are needed to identify the program and community features that account for the effectiveness of legal or social service

interventions with various groups of offenders” (Crowell & Burgess, 1996, p.140). The Division of Experimental Criminology of the American Criminology Association, and its accompanying journal of the same name, have helped to reinforce and further this ideal.

As we discuss below, there are however obstacles, challenges, and limitations that often prevent the experimental ideal from being realized. So the “evidence” often has to include quasi-experimental evaluations and observational studies to formulate evidence-based practice. Some state and federal agencies, and professional and foundation organizations, have therefore calibrated the evidence in terms of evidence-based programs, promising programs, supported programs, model programs, and so on, based on the extent and rigor of the available research. There is also the distinction of “specific evidence,” which is based on research of the particular program in question, and lesser “generic evidence,” which is based on related or similar programs. Agencies and organizations have then cataloged the rated practices in clearinghouse, blueprint, program guide, and “what works” reports for practitioners and policy-makers to consult in determining which programs to support and implement.

### **Meta-Analyses and Effect Sizes**

An increasingly popular tool associated with identifying evidence-based practice is meta-analysis (Durlak & Lipsey, 1991). This statistical computation summarizes the effect of several program evaluations as a whole. Researchers first identify the most rigorous studies based on preset criteria, which generally means relying heavily on experimental program evaluations. They then calculate a standardized measure of effect size for the combined single-site evaluations (Cohen’s *d*, with the value of 0 to 1, is the most commonly used). The results represent a summary of all the evaluations that are included and may also detect variations across program sites and approaches, and research designs and methods.

Interestingly, different meta-analyses in any one field may produce varying results as a result of different inclusion criteria and interpretations of the coefficients. There are, for instance, at least seven meta-analyses that have been conducted on batterer program evaluations. The majority has shown little or no program effect compared to non-treatment controls, reflecting the outcomes of five available experimental evaluations. One more recent meta-analysis from the notable Cochrane Collaboration, however, indicated that the evaluations were too problematic to formulate a conclusion one way or the other (Smedslund et al., 2007). Systematic reviews of the broader research identify program effectiveness with a series of contingencies, such as sufficient court oversight and supplemental services. They tend to consider other research designs, observational studies, and statistical modeling on non-experimental data (discussed further below) that aren’t included in meta-analyses.

Meta-analyses and especially their effect sizes are often presented or used as a convenient bottom-line verdict on programs and policies. There is, however, continued discussion over how best to interpret the effect sizes. The authors of one of the meta-analysis of batterer programs adds a further cautionary note: “One of the greatest concerns when conducting a meta-analysis is the ease at which the ‘bottom-line’ is recalled and the extensive caveats for caution are forgotten or ignored” (Babcock, Green, & Robie, 2004, p. 1046). This is a concern that could be applied to research findings in general, but is particularly acute with regard to evidence-based practice.

## **The Debate over Evidence-Based Practice**

Despite its potential contributions, the evidence-based practice movement is meeting with some objections and controversy. The criticisms of evidence-based practice seem to fall in three main categories: one, the implementation challenges of the experimental “gold standard,” two, the conceptual issues associated with experimental designs that neglect program context, and three, the resistance of practitioners to adopting or fully embracing research results. The overarching concern is that the evidence-based practice movement has become too dependent on experimental program evaluation—an evaluation design that is not always as rigorous as it would appear when implemented. As Robert Sampson, past president of the American Society of Criminology charges: “Criminological [experimental] randomists have overreached in their claims and generated their own folklores, or what I think are more appropriately referred to as myths. Experimental myths are more than just stories or part of a tradition—they have become actively institutionalized in the routine workings of criminology” (Sampson, 2010, p. 490).

## **Methodological Concerns**

There is a fundamental concern about the difficulties in conducting experimental evaluations in “real-world” settings. Randomized assignment of subjects to treatment and non-treatment groups is frequently impractical, yet randomization is the lynchpin of experimental designs. Subjects often will not consent to the assignment for a number of reasons, practitioners (or judges in court settings) will override some assignments based on case needs, and subjects tend to drop out of mandatory treatments or interventions undercutting the “experimental” treatment group. Follow-ups with subjects can, moreover, be logistically difficult to achieve and compounded by resistance to sensitive questions. At its core, there are proverbial ethical concerns, especially over the possibility of putting some subject in a non-treatment control group that deprives them of treatment that may benefit them. There are many more such challenges to negotiate leaving one prominent researchers to re-term experimental program evaluation as the “bronze standard” instead of the “gold” (Berk, 2005), and another to insist that any sense of Olympic medals should be dropped altogether (Sampson, 2009).

Rehabilitation programs with court-referred offenders present a particular challenge in this regard. Dropout rates to an experimental treatment option tend to run between 40-60 percent turning the experimental option into an “intention-to-treat” option rather than “treatment-received.” We don’t know, therefore, the outcome of actually receiving the treatment, or its full “dose,” and whether supplemental treatments or incentives would account for a better outcome of the experimental group. With all these potential implementation issues, experimental designs may be more compromised than their status suggests and warrant qualification and discussion.

One study of 500 evaluations of behavioral treatment programs for adolescents demonstrated that the greater the implementation problems, the lower the effect size of the program compared to a control group (Dulak & Dupree, 2008). Several other studies have revealed, moreover, that the vast majority of the published program evaluations in the criminal justice field, in particular, fail to sufficiently acknowledge the

implementation problems and qualify their results accordingly (Mears, 2003). Making such information explicitly available could in itself help practitioners more appropriately gauge the research implications.

A number of instruments have been developed to systematically examine the extent and nature of the implementation problems and help to offset the issue of underreported shortcomings and limitations—they simply need to be more widely used, according to their proponents. One of the most comprehensive is the Consolidated Standards for Reporting Trials (CONSORT) introduced in 1996 to improve reporting of experimental clinical trials particularly in the medical field. CONSORT tables offer a clear summary of the strengths and weaknesses across 22 implementation issues that include randomization, intention to treat, effect size, conflict of interest, subject withdrawal and dropouts, and adverse or “uncontrollable” events.

### **Conceptual Issues**

A hotly debated conceptual issue goes beyond the methodological concerns raised above: To what degree does the biomedical model (e.g., giving a dose of medication and observing its physical effects) apply to what are more accurately considered to be “social interventions”? That is, many of the program “treatments” are embedded in systems of referral, screening, court oversight, supplemental services, community collaborations, coordinating councils, and so on. A community’s police response, employment level, and cultural norms may be influences as well. All of these components impact the subject pool, level of dropout, treatment quality, and thus outcomes of a program evaluation.

Consequently, a contingent of researchers argues for more complex and sophisticated research designs that account for the program context. As Smyth and Schorr (2009) write in their report on evaluating child welfare programs, “The evaluation tools have to be able to incorporate not only a program’s work, but how that program fits with other interventions. In other words, some of the very factors and situations that the experimental method controls may need, instead, to be explicitly folded into an evaluation” (p. 18). The researchers conclude that, as a result, “the dogma of experimental designs is ultimately detrimental to program development and social intervention” (p.21).

There are several alternatives being used to remedy these concerns. An analytical approach to address the conceptual issues, and also many of the implementation issues, is statistical modeling—specifically the use of instrumental variable analysis and propensity score analysis. The goal of both analytic approaches is to simulate experimental conditions by controlling for potential differences in subject characteristics across the comparison groups, such as program completers versus dropouts. That is, the analyses attempt to balance two nonequivalent groups on measured subject characteristics in order to produce a more accurate estimate of the effects of a treatment. Addressing non-experimental data in this way avoids the difficulties and disruptions of randomization and thus allows for more “real-world” or naturalistic circumstances. Instrumental variable analysis additionally controls for contextual program factors, and propensity score analysis produces outcomes for sub-groups or types of subjects as well as the sample as a whole.

These two methods have been used extensively in education, agriculture, public health, and economics, especially when experimental research is impractical or too difficult to implement. Criminal justice researchers have also begun to use them in order to approximate equivalent comparison groups of criminal offenders (Angrist, 2006). The main shortcoming is that both analytic approaches require extensive subject characteristics and a large sample size, which are not essential in well-implemented experiments. Some critiques argue, as well, that statistical modeling, even under the best circumstances fails to establish truly equivalent comparison groups, and to create reliable measures for program context. The proponents of statistical modeling claim, on the other hand, that the modeling effort is worthwhile considering the inherent “naiveté” of experimental designs and “the promising developments in the theory and practice of non-experimental evaluations” (Heckman and Smith, 1995, pp. 108–109).

Another way to address the context of “social interventions” is through multi-site program evaluations. In this approach, program evaluations are conducted concurrently in different community settings to see if the results hold up across variations in settings. Multi-site studies of alcohol treatments, depression therapies, and criminal rehabilitation, in fact, have overturned some of the conclusions drawn from single-site evaluations, as a result. Participants assigned to a particular treatment have better outcomes in one city, but poorer outcomes in another. Multi-site studies of this sort are, however, very costly to implement, complicated to supervise, and sometimes difficult to interpret.

Multiple methods are also increasingly recommended to help represent the broader intervention and its context (Government Accountability Office, 2009). This might include direct observations of rehabilitation programs, court transactions, and probation procedures, as well as open-ended interviews with staff and community leaders. While determining what effect is attributable to the batterer program remains problematic, descriptive information regarding the context can help qualify and interpret a program’s outcomes. It also can bring a deeper understanding of the intervention in question—how it works or why it doesn’t work.

There are, additionally, increasing recommendations for system-analysis in the evaluation field (Kelly, 2007). Systems analysis is, of course, a broad term but is commonly used in business management to represent the operations of an entire corporation and its component parts. It is also a perspective increasingly brought to public health projects considered to be an “open system” interacting with a variety of other service agencies, informal networks, and the community at large. A 2007 special issue of the *American Journal of Community Psychology* was devoted entirely to the topic. Textbooks, such as *Fourth Generation Evaluation* (Guba & Lincoln, 1990), are also available to help design a systems approach to program evaluations.

The criminal justice field has been applying a system perspective in its notion of “community coordinated response” to sex offenders, domestic violence cases, prisoner re-entry, and substance abuse cases. The assumption is that a variety of criminal justice and community service components interact together for “successful” outcomes in these cases. In the domestic violence field, this approach is reflected in the “system audits” to monitor the actions and coordination of the system components, such as the response to 911 calls, police arrests, court actions, case monitoring, offender rehabilitation, victim services, and case management. There is also increasing evidence

that the components of a community coordinate response improve batterer program outcomes (Gondolf, 2012)

## **Practitioner Resistance**

A recent published roundtable on court innovations highlights a third area of concern about evidence-based practice. It noted a “cultural suspicion of anything academic” among practitioners despite the need for decision-making based on data and the self-reflection that promotes (Berman, 2008, p. 99). The pressure for practitioners to “live in the moment” adds to the tension in a practical way. Crises order the day and tend to preclude long-term planning, according to the roundtable panel. As a result, “There is almost a complete disconnect between practice and the parallel university of research” (Berman, 2008, p. 103).

Practitioner resistance to evidence-based practice may also come from the frustrations with limited resources and staffing, inconsistencies in court referral and oversight, and administrative shortcomings and ineptness. Practitioners tend to think in global terms—that is, they consider broader, multifaceted, and entangled relationships, and are sensitive to a variety of idiosyncrasies, exceptions, and contingencies among their program participants. We frequently hear as well that the evaluated programs don’t apply to the circumstances of their particular program, or their program has evolved or changed substantially since the program evaluations of the evidence-based practice. As a result, so-called judgment-based or clinical-based practice may be more the de facto rule (Pollio, 2006).

An additional practitioner concern is the tendency of evidence-based practice to put forth a bottom-line or authoritative verdict. That is, research findings are too often reduced to a seemingly categorical statement about what works and consequently betrays the complexity, nuance, and qualifications of research. The more severe critics fear, moreover, an autocratic hierarchy of experimental researchers end up dictating policy (or at least influencing it heavily) to marginalized practitioners (Pollio, 2006). Practitioners counter that their “evidence” derived through clinical observation, practitioner experience, and case studies are generally excluded from the consideration of evidence-based practice.

Well-aided in the mental health literature is also the challenge of translating evidence-based recommendations directly into practice (Westen, Stirman, & DeRubeis, 2006). Most experimental evaluations rely on manualized treatment to ensure the integrity of what is being tested, while most clinicians favor flexibility with diverse clients and circumstances. Evidence-based practice has done poorly when applied to people from non-dominant cultures and ethnic groups. As a number of critiques also point out, program evaluations with community-based services serving minorities are few in part because those services tend to be under resourced and not “research ready”.

Research results can be downright confusing to practitioners, as a few examples from the domestic violence field illustrate. The noted Minneapolis police study of the 1980s and its replications seemed to disagree on the impacts of arrest in domestic violence cases (Garner & Maxwell, 2000); A more recent multisite evaluation of judicial oversight of domestic violence cases, produced mixed results that run counter to the experience of the practitioners involved in the study (Visher, Harrell, & Yahner, 2008).

Our own multi-site evaluation of batterer programs, using statistical modeling, counters the experimental program evaluations that suggest no effect (Gondolf, 2012).

### **Practitioner Input**

The academics writing about practitioner resistance propose a democratized interaction between researchers and practitioners, and reinstate the process-orientation of its initial conception (Holmes, Murray, Perron, & Rali, 2006). “Research readiness” among practitioners, or “critical consumers” of research, is needed to give feedback and respond critically. Under the heavy workloads and crisis-driven schedules of most practitioners, this sort of “research readiness” is difficult to achieve and maintain. There are efforts in many fields to compensate for a lack of training in research basics through professional conferences, technical assistance, and research briefs, but the gulf continues to be a substantial one to bridge.

In turn, researchers also would benefit from greater “practice wisdom” in order to appreciate the outlook and experience of those affected by their research. One federal agency, coupled with a national non-profit organization, convened a series of seminars joining leading researchers and practitioners to frankly debate the evidence-based practice research and its application to batterer intervention. The summary reports have then been disseminated to inform and engage others in the cross-training experience.

Finally, there are grant solicitations for practitioner-initiated research that enable unique and distinguished programs to develop their own documentation and evaluations. Federal agencies have, as well, issued solicitations for long-term research-practitioner collaborations addressing criminal justice interventions, beyond the more superficial cooperative agreements that accompany program evaluations and research.

In medical settings, consensus panels are also established for new innovations and treatments. A variety of researchers and practitioners, along with administrators and advocates, convene to discuss reviews of the research, practitioner experience, and administrative issues. There is some wrangling to sort out what might be the “best practice” based on a number of criteria that may include patient satisfaction and program feasibility, as well as research evidence. It generally suggests a systematic sorting of researcher and practitioner recommendations, and an emerging consensus around certain practices. This approach may extend to establishing program standards or guidelines, or “standard of care,” for the field. Some argue, however, that standards have relied too much on practitioners and advocates rather than on the evidence-based research, as is the case with regard to domestic violence batterer programs.

### **Recommendations**

This overview is not meant to dismiss or undercut the evidence-based practice movement. Rather the intent is to broaden and refine it. The call for evidence-based practice arises out of a need for more substantiation, accountability, efficiency as well as effectiveness in intervention and treatment. It contributes a logical, rational, and systemic thinking to important questions that are sometimes skewed by personal philosophy, limited observation, and political intents. This sort of thinking, ideally, brings more objectivity to policy and program development.

As discussed above, there are inevitably challenges and misuses, and even distortions of evidence-based practice. Critics for instance object to the exclusive

reliance on experimental evaluation, the bottom-line verdicts regarding effectiveness, and the disruptive impositions of research on practice. The extent and impact of these issues are admittedly debatable, but the wide-range of researchers raising them at least warrants pause and caution. In response, a host of remedies attempt to address the concerns, but need to be more vigorously introduced, especially to practitioners caught unwittingly by bottom-line assertions.

Critics recommend that researchers be more forthright in acknowledging the limitations of their work and alternative interpretations of it. Practitioners have, in particular, called for more attention to the nuance and complexity of outcomes, the mediating effects of context, and more familiarity with the “real world” experience of intervention. These concerns might entail more extensive data collection and sophisticated computer modeling. On the other hand, practitioners are in need of more “research readiness” both in terms of their understanding of research demands and their program’s ability to accommodate them. If they are to truly collaborate or be more involved in the research, they need to be conversant in basic research principles, as well as their more global and idiosyncratic appreciation of their clients. All this begs for cross-training and shared conferences that federal and regional agencies have convened.

Federal and state agencies have posed some additional alternatives to develop more “grounded” research and evidence. They have stipulated documentation of collaborations with practitioners (beyond “drive-by” practitioner sign-offs), practitioner-initiated-research projects, technical assistance to establish program research-readiness, and research review and dissemination procedures that ensure practitioner response and input. One might argue that these efforts do not preclude experimental evaluations, rather they supplement them. They usually entail different research designs and approaches (e.g., case studies, action research, longitudinal follow-ups, and community ethnography) that follow the recommendation of policy commissions calling for diversifying the sources of evidence.

Additionally, there are structural inducements for integrating research knowledge and clinical experience more broadly—as well as a findings from different research methodologies and approaches. The medical field, in particular, has long-standing consensus panels that bring together practitioners and researchers to review research and its applications to clinical settings. Similar committees and commissions have convened to develop “best practices” or “what works” that represent agreement of research and practitioners over what appears to be most effective intervention or treatment. Standards of care or program standards have been negotiated in most fields, often with stakeholders as well as practitioners, researchers, and insurance companies or state funders. These ventures certainly help to impose an exchange and collaboration, but accounts of some of these efforts expose the difficulties in establishing an ideal partnership.

Ultimately, the question is how to realize the ideal of “evidence-based” practice as a process—one that is a collaborative feedback loop of researchers and practitioners. One in which properly qualified research findings are part of a discourse rather than policy pronouncement. Such process-based partnerships do exist in the domestic violence field, as well as others; and have been documented and forwarded as models to emulate. All of this takes us back—and also forward—to the founding

principles of evidence-based practice in the early 1990s. The lingering question is whether these principles can reconcile the increasingly entrenched factions, specifically in the domestic violence field—and beyond.

## References

- Angrist, J. (2006). Instrumental variables methods in experimental criminological research: What, why, and how? *Journal of Experimental Criminology*, 1, 23-44.
- Babcock, J., Green, C., & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violence treatment outcome research. *Clinical Psychology Review*, 23, 1023-1053.
- Berk, R. (2005). Randomized experiments as the bronze standard. *Journal of Experimental Criminology*, 1, 416-433.
- Berman, G. (2008). Learning from failure: A roundtable on criminal justice innovation. *Journal of Court innovation*, 1, 97-122.
- Crowell, N., & Burgess, A. (1996). *Understanding violence against women*. Washington, DC: National Academy Press.
- Dunford, F. (2000). Determining program success: The importance of employing experimental research designs. *Crime and Delinquency*, 46, 425-434.
- Durlak, J., & DuPre, E. (2008). Implementation Matters: A review of research on the influence of implementation on program outcomes and the factors affective implementation. *American Journal of Community Psychology*, 41, 327-350.
- Durlak, J., & Lipsey, M. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology*, 19, 291-332.
- Garner, J., & Maxwell, C. (2000). What are the lessons of the police arrest studies? *Journal of Aggression, Maltreatment & Trauma*, 4, 83-114.
- Gilgun, J. (2005). The four cornerstones of evidence-based practice in social work. *Research on Social Work Practice*, 15, 52-61.
- Gondolf, E. (2002). *Batterer intervention systems: Issues, outcomes, and recommendations*. Thousand Oaks: Sage Publications.
- Gondolf, E. (2012). *The Future of Batterer Programs: Reassessing Evidence-Based Practice*. Boston: Northeastern University Press.
- Government Accounting Office (2009). *Program evaluation: A variety of rigorous methods can help identify effective interventions*, Report to Congressional Requestors, No. 424. Washington, DC: U. S. Government Accounting Office ([www.gao.gov/new.items/d1030.pdf](http://www.gao.gov/new.items/d1030.pdf)).
- Guba, E., & Lincoln, Y. (1990). *Fourth generation evaluation*. Thousand Oaks, CA; Sage Publications.
- Heckman, J., & Smith, J. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, 9, 85-110.
- Holmes, D., Murray, S., Perron, A., & Rali, G. (2006). Deconstructing the evidence-based discourse in health sciences. *International Journal of Evidence Based Healthcare*, 4, 180-186.
- Kelly, J. (2007). The system concept and systemic change: Implications for community psychology. *American Journal of Community Psychology*, 39, 415-418.

- Mears, D. (2003). Research and interventions to reduce domestic violence revictimization. *Trauma, Violence and Abuse*, 4, 127-147.
- Pollio, D. (2006). The art of evidence-based practice. *Research on Social Work Practice*, 16, 224-232.
- Prybylski, R. (2012). Editor's Introduction. *Justice Research and Policy*, 4, 1-15.
- Sampson, R. (2010). Gold Standard Myths: Observations on the Experimental Turn in Quantitative Criminology. *Journal of Quantitative Criminology*, 26, 489–500.
- Scott, R., & Shore, A. (1979). *Why sociology does not apply: A study of the use of sociology in public policy*. New York: Elsevier.
- Sherman, L. (1992). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.
- Sherman, L. (2009). Evidence and liberty: The Promise of experimental criminology. *Criminology and Criminal Justice*, 9, 5–28.
- Smedslund, G., Dalsbo, T., Steiro, A., Winsvold, A., & Clench-Aas, J. (2007). Cognitive behavioural therapy for men who physically abuse their female partner. *The Cochrane Database of Systematic Reviews*, Issue 4, Article No. CD006048 ([www.cochranelibrary.com](http://www.cochranelibrary.com)).
- Smyth, K., & Schorr, L. (2009). *A lot to lose: A call to rethink what constitutes "evidence" in finding social interventions that work*. Harvard Kennedy School of Government Working Paper Series, Harvard University, Cambridge, MA ([www.hks.harvard.edu/socpol/publications\\_main.html](http://www.hks.harvard.edu/socpol/publications_main.html)).
- Visher, C., Harrell, A., & Yahner, J. (2008). Reducing intimate partner violence: An evaluation of a comprehensive justice system–community collaboration. *Criminology and Public Policy*, 7, 495–523.
- Westen, D., Stirman, S., & DeRubeis, R. (2006). Are research patients and clinical trials representative of clinical practice? In J. Norcross, L. Beutler, & R. Levant (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions* (pp. 161–189). Washington, DC: American Psychological Association.

### **About the Author**

Edward W. Gondolf, EdD, MPH, is currently a research associate and former research director for the Mid-Atlantic Addiction Research and Training Institute (MARTI), based at Indiana University of Pennsylvania (USA). His most noted book *Batterer Intervention Systems (2001)* summarizes a seven-year evaluation of batterer intervention systems in four cities funded by the US Centers for Disease Control, and a related NIJ study using the longitudinal data to identify risk factors for re-assault. Under grants from the National Institute of Justice (NIJ), he more recently evaluated the effectiveness of specialized counseling for African-American men, a study of case management for domestic violence offenders, and a 4-year evaluation of supplemental mental health treatment for batterer program participants. Dr. Gondolf's current book, *The Future of Batterer Programs: Reassessing Evidence-Based Practice (2012)*, addresses the debate over the effectiveness of batterer programs and the means to improving it.