7-14-2014

Evidence Based Practice in Juvenile Justice: Nebraska White Paper

Richard L. Wiener
University of Nebraska - Lincoln

Anne M. Hobbs
University of Nebraska at Omaha, ahobbes@unomaha.edu

Ryan E. Spohn
University of Nebraska at Omaha, rspohn@unomaha.edu

Evidence Based Practice in Juvenile Justice:

Nebraska White Paper

Dr. Richard L. Wiener

University of Nebraska/Lincoln

Dr. Anne Hobbs and Dr. Ryan Spohn

University of Nebraska/Omaha

July 14, 2014

Overview of White Paper

This White Paper is the product of the collaborative effort of the University of Nebraska/Lincoln (UNL) Law and Psychology Program, the University of Nebraska/Omaha (UNO) Consortium for Crime and Justice Research and the UNO Juvenile Justice Institute. The purpose of this paper is to provide an overview for understanding, testing, and developing Evidence Based Practice (EBP) interventions that make rehabilitative services available to children in the juvenile justice system. The paper begins with a summary of a proposal for a classification system of EBP programs in the Juvenile Justice System in Nebraska and then goes on to explain the logic of the classification system.[1]

Classification System for Evidence Based Juvenile Justice Programs in Nebraska

I. Model Program/ Fully Evidence Based Practice – The program satisfies the following five criteria:

1) The program demonstrated effectiveness with a randomized experimental study (RCT) or two quasi-experimental studies in which the treatment group showed a significant difference on the target outcome as compared to the control group.
2) The effect lasted for no less than 1 year after the intervention.
3) There is at least one independent replication with a RCT or two more quasi-experimental evaluations.
4) The combination of designs adequately addressed all the threats to internal validity (i.e., the design allowed for a strong inference of causality).
5) The program has produced no compromising negative side effects.

[1] NOTE: To fully understand the classification system that we propose it is necessary to read the full paper, which explains the logic of a temporary assignment and then a final assignment of EBP categories.

II. **Effective** – One RCT or two quasi-experimental designs document the program's effectiveness. Furthermore, an evaluator has replicated the program's effectiveness with an RCT design or two quasi-experimental designs but the researcher was not an independent investigator.
III. **Promising** – There has been one successful RCT or two quasi-experiments that document the effectiveness of the program but there was no replication study available **OR** the program matches the dimensions of a successful meta-analysis practice.
IV. **Inconclusive** – There has been one successful RCT or two quasi-experimental evaluations of the program but there are contradictory findings in these or additional studies OR the program would be promising or effective but the effects are short in duration.
V. **Ineffective** – The RCT or two quasi-experimental evaluations failed to show significant differences between the treatment and control group.
VI. **Harmful** – The RCT or two quasi-experiments showed that the control group scored higher on the targeted outcome than did the treatment group and the difference is statistically significant.
VII. **Insufficient Evidence** – There is no RCT or less than two quasi-experimental evaluations of the program to date.

Logic of the Classification System for Evidence Based Juvenile Justice Programs in Nebraska

The remainder of this paper discusses the logic of the classification system: 1) four methods for verifying the EBP status of programs, 2) current definitions of Evidence Based Practice and a working definition that could be useful in the state of Nebraska, and 3) processes for establishing EBP programs in Nebraska.

Four Evidence Based Practice Evaluation Modes

Experimental or Quasi-experimental Evaluation of Implemented Programs. The most direct way to establish that a program is effective is to conduct a controlled experiment or quasi-experiments in which groups of comparable clients either receive the program treatment or they do not. The strongest form of direct impact evidence employs a randomized experimental design also referred to as randomized control trials (RCT) in which the evaluator administers the treatment to some members of the target client population (treatment or experimental group) and withholds it from others (control group) and then compares the two groups on the targeted outcome measures (e.g., recidivism). In a true experiment (i.e., RCT), the evaluator randomly assigns participants to the treatment and control groups, while in a quasi-experiment participants are already in existing groups (treatment and comparison groups) before the experiment begins. Random assignment produces the greatest experimental control and therefore allows the strongest inference that the treatment caused changes in the outcome variable, while quasi-experiments measure and rule out nuisance factors one at a time to buttress causal inferences

(e.g., relying on pretests, measuring potential confounds, and employing statistical control). Quasi-experiments can never produce the same level of confidence in causal inference as do true experiments but replicating findings across quasi-experiments greatly increases confidence in outcomes. When true experiments and quasi-experiments produce significant differences between groups with moderate to large effect sizes, document the nature of the services provided, measure the quality of service implementation and describe the nature of juveniles receiving the treatments, they provide convincing evidence that the program is evidence based (Lipsey, Howell, Kelly, Chapman & Carver, 2010).

The advantages of conducting direct experimental evaluations of implemented programs are that they provide evidence that a program works in the setting in which it was implemented with the population that is in need of services. However, the approach is not without disadvantages, namely, setting up experimental (or quasi-experimental) tests of program effectiveness is not always feasible, i.e., it is not always possible to randomly assign and withhold treatments. Furthermore, when such studies are feasible, they are costly and time consuming to perform and usually require research training that may go beyond the resources of local program administrators.

Model Programs Approach to Evidence Based Practice. There are model programs that researchers have already shown to be effective with replicated experimental or quasi-experimental tests of outcomes. Furthermore, there are outside and independent research agencies such as the University of Colorado Blueprints for Violence Prevention Project (Blueprints) (Mihalic et al., 2001) and the Office of Juvenile Justice and Delinquency Prevention (OJJDP) that verify that the evaluations are rigorous and the effects are reliable. Importing and administering these programs results in an EBP intervention provided that the importation carefully implements the program according to the program manual rules and procedures. Model programs frequently require special training for program personnel so that they can administer the programs following as closely as possible the procedures that produced the original direct evaluation findings. For this reason, model programs often come with instruments to measure and document the fidelity of program implementation (Lipsey, et al., 2010). The University of Colorado at Boulder, Center for the Study and Prevention of Violence has developed a menu driven website "Blueprints for Healthy Youth Development" that allows researchers and administrators to search for established model programs that have shown effective experimental or quasi-experimental impact studies (www.blueprintsprograms.com/contact.php).

Instituting a model program as an EBP takes advantage of the prior research, which assures that the program will produce the desired outcome provided that that the program personnel follow the program guide as closely as possible. The major disadvantages include the high startup costs (including training) to implement the program exactly as the developers intended and the difficulties of interjecting the program in a different subculture with different participants than the evaluators used in the original demonstration project. Two examples of model programs with known and verified impact studies for youth are Functional Family Therapy (FFT) (Alexander & Robbins, 2011) and Multisystemic Therapy (MST) (Henggeler, Melton, & Smith, 1992; Henggeler, Melton, Brondino, Sherer, & Hanley1997).

Comparing Existing Interventions to Program Specific Meta-analyses. In the last 15 years, program evaluators have conducted a number of meta-analytic research projects that summarize the aggregate effects of experimental and quasi-experimental intervention studies to reduce recidivism and increase well-being among youth. A meta-analysis is a quantitative review of a large number of studies that analyzes and summarizes the treatment effects (statistical estimates known as effect sizes) and characteristics of programs in a way that tests the overall effects of a class of interventions across a number of program and sample characteristics. Lipsey (2009) points out that meta-analyses in juvenile justice have examined the effects of specific types of programs including boot camp (MacKenzie, Wilson, & Kider, 2001), cognitive-behavioral therapy (Landenberger & Lipsey, 2005), prison visitation (Petrosino, Turpin-Petrosino, & Buehler, 2003), family therapy (Latimer, 2001), drug court (Wilson, Mitchell, & MacKenzie, 2006), victim-offender mediation (Nugent, Williams, & Umbreit, 2004), and Multisystemic Therapy (Littell, Popa, & Forsythe, 2005). To understand the significance of these meta-analyses it is important to draw a distinction between "programs" and "practices". In the language of the National Institute of Justice (NIJ) a program is "a specific set of activities carried out according to guidelines to achieve a defined purpose", while a practice is "a general category of programs, strategies, or procedures that share similar characteristics with regard to the issues they address and how they address them" (www.crimesolutions.gov/Default.aspx).

Most meta-analyses (like the ones named above) aggregate multiple studies of individual programs and the results speak to the effectiveness of those programs in a specific area of practice. When a practice level meta-analysis shows moderate to strong effects across a large number of programs as did one for cognitive-behavioral therapy with offenders (Landenberger & Lipsey, 2005), researchers can use the findings to evaluate the status of programs that were not included in the meta-analysis. To the extent to which a program not included in the meta-analysis implements a process that uses similar materials instruments and procedures to those which were successful in producing effects in the meta-analysis, the new program is an EB program with an expectation of effects that will be similar in strength and direction to those in the meta-analysis. The NIJ uses a checklist (available on its website) to review the quality of meta-analyses for any area of practice to determine whether the practice and the programs are evidence based. This approach to establishing an evidence based practice by comparing the program in question to an existing meta-analysis has the advantage of making use of existing programs already in place without collecting additional evaluation data. Therefore, the approach is less expensive and demanding of professional resources. The disadvantage to the approach is that there may not always be a meta-analysis of a practice area that includes the program in question and when such a meta-analysis does exist, it may be difficult to determine whether the program's attributes are sufficiently developed to belong to the set of programs in the original meta-analytic study.

Comparing Existing Interventions to Generic Meta-analyses. While a number of meta-analyses in the juvenile justice literature look at specific practices and the programs that make them up, researchers at Vanderbilt University have taken a different approach and conducted a series of meta-analyses, the most recent in 2009, that aggregated the effects of all types of practices in juvenile justice aimed at reducing recidivism (Lipsey et al., 1998; 2010;

Lipsey, 1992; 2009). Lipsey (2009) included in the latest meta-analysis a large list of predictors of recidivism effects including attributes of the study's methods, characteristics of the programs, characteristics of the offenders and types of treatment, which together determine differences in effect sizes between the treatment and control groups. Lipsey's (2009) most recent meta-analysis included 548 independent studies from 361 research reports that compared treatment and control conditions on recidivism rates for juveniles aged 12 to 21. The studies included tests of all types of intervention to reduce recidivism conducted between 1958 and 2002. Lipsey and colleagues define recidivism as the proportion of juveniles rearrested during the 12 months after treatment. After statistically controlling for methodological differences between the studies, the meta-analysis showed that age, gender mix, and ethnicity of the sample were not related to the interventions' ability to reduce recidivism but delinquency risk was such that programs that focused on high risk youth were more likely to produce a reduction in recidivism as compared to those that aimed at low risk youth. Furthermore, the effects of type of supervision (no supervision, diversion, probation/parole or incarceration) had no impact on the effectiveness of the programs under study. Most importantly, programs based in surveillance, deterrence or discipline had either increased recidivism or had no effect upon it, while restorative justice, counseling, skill building and multiple service programs all reduced recidivism. In short, therapeutic programs – those that "attempt to engage the youth in a supportive, constructive process of change" were successful in reducing recidivism, while those that "rely more on external control and coercion (e.g., through discipline or surveillance)" were not successful (Lipsey, 2009, p. 128). Among the most successful practices, namely, counseling, skill building, and multiple coordinated services, several subtypes (i.e., mentoring programs, group therapy, behavioral management programs, cognitive behavioral therapy, and case management programs), on average, reduced recidivism by at least 20%. Finally, programs that had high quality of implementation (i.e., fidelity) and included a higher dose level (i.e., hours of service provided) decreased the recidivism rate the most. Lipsey et al (2010) argue persuasively that programs that match those that were effective in reducing recidivism in their meta-analysis of almost 550 studies are EBP's. In fact, the Vanderbilt group developed a Standardized Program Evaluation Protocol (SPEP) for scoring and identifying programs that are evidence based because they match the successful programs in their generic meta-analysis. Indeed, shared dimensions between new programs and those that reduced recidivism in the Lipsey (2009) meta-analysis are strong evidence that the programs are evidence based provided that the programs are implemented with high fidelity, target high risk youth, and have high enough dosage (i.e., enough service contact hours).

Using this generic meta-analysis as a guide to rank evidence based programs has the advantage of using the results of over 40 years of research across multiple samples and locations, as well as reducing the expense and time investment associated with conducting independent outcome analyses. According to Lipsey et al (2010), those programs consistent with the successful ones in the meta-analysis of 538 studies are worthy of continued support and those that are inconsistent require modification or discontinuation. The disadvantage of using only this approach is that it may be difficult to determine whether the program's attributes are sufficiently developed to belong to the set of programs in the generic meta-analytic study and that the administrators of the specific program under investigation need not collect any outcome data.

However, with respect to data collection, regardless of the method of EBP analysis used, it is essential for all programs to collect fidelity data to assure that the program staff have implemented the program as intended and it is also essential to collect outcome data to demonstrate that the program as adopted has had its intended effect when implemented to scale in a context different from the original demonstration projects.

Definitions of Evidence Based Practice

Given the different approaches to analyzing Evidence Based Programs in juvenile justice, the problem of developing an accepted definition of an EBP is complicated and fraught with political obstacles. However, underlying all four approaches to EBP analysis is the fundamental requirement that in order for a program to be evidence based, there must be at least one experimental (RCT) evaluation that shows the treatment group outperformed the control group on the target outcome measure. The evaluation could be a direct impact study of the program in implementation, a direct impact study of a model program that was adopted and implemented on site, or it could result from application of a meta-analysis of a specific area of practice or a generic meta-analysis of all interventions related to the problem area (e.g., a meta-analysis of all programs aimed at reducing recidivism in juvenile justice).

In March of 2004 the Department of Justice, Office of Justice Programs brought together researchers and administrators from the U.S. Department of Education's Institute for Education Sciences and the U.S. Department of Health and Human Services' Substance Abuse and Mental Health Services Administration along with some leading private organizations including the National Council for Excellence in Government's Coalition for Evidence Based Policy and the National Center for the Study and Prevention of Violence. The conference established the "Working Group for the Federal Collaboration on What Works" (Working Group) and assigned the task force the job of developing a unified definition of EBP. The result was a quantitative research design approach entitled "A Hierarchical Classification Framework for Program Effectiveness," The Working Group included five criteria for certifying a program as effective: 1) the program demonstrated effectiveness with an experimental (RCT) study in which the treatment group showed a significant difference on the target outcome as compared to the control group. 2) The effect lasted for no less than 1 year after the intervention. 3) There is at least one independent replication with a RCT. 4) The design of the study adequately addressed all the threats to internal validity (i.e., the design allowed for a strong inference of causality). 5) The program has produced no compromising negative side effects.

The working group went on to create a hierarchical classification system with seven levels of program development that programs could attain based upon derivations of the five verifying criteria:

I. **Model Program** – the program meets all five points for certifying a program effective.
II. **Effective** – an evaluator has replicated the program's effectiveness with an RCT design but the researcher was not an independent investigator.

III. **Promising** – there has been one successful RCT of the program but there is no replication study available.
IV. **Inconclusive** – there has been one successful RCT of the program but there are contradictory findings in additional studies OR the programs effects are short in duration.
V. **Ineffective** – the RCT failed to show significant differences between the treatment and control group.
VI. **Harmful** – the RCT showed that the control group scored higher on the targeted outcome than did the treatment group and the difference is statistically significant.
VII. **Insufficient Evidence** – there is no RCT of the program to date.

Other classification systems recognize value of quasi-experimental studies that do not reach the level of confidence that RCT evaluations produce. For example, the National Institute of Justice (Crimesolutions.gov) recognizes that quasi-experimental designs, although not as strong as RCT studies can, when replicated, show convincing evidence that programs are effective at bringing about improvements on the targeted outcome. Evidence based practice in Nebraska should adopt a definition that weighs RCT evaluations more highly than quasi-experiments but should still give credibility to the latter. For example, the NIJ uses a rating scale that assigns different amounts of points depending upon the rigor of the evaluation design:

3 points —Experimental (well-designed randomized field trial)

2 points – Quasi-experimental Level 1 (there is a credible control group with extensive information provided about the equivalency of the groups before the treatment began)

1 point – Quasi-experimental Level 2 (there was a control group but it lacks comparability on pre-existing variables or lacks information concerning equivalency of the groups before the treatment began)

0 points – Non-experimental design without a control group

The Working Group's definition requires a direct impact evaluation of each program before it is willing to certify the program as EB. The Working Group does not take into consideration the value of program- specific or generic meta-analyses. The Nebraska definition should make use of the more current approach and partially rely on meta-analyses to classify new programs as evidence based when they are similar enough to other existing programs, which have produced positive outcomes in meta-analyses. The following Hierarchical Classification System for Nebraska Juvenile Justice Programs (NJJHCS) combines the Working Group Definition with the meta-analysis approach and adds some reduced value for quasi-experimental studies:

I. **Model Program/ Fully Evidence Based Practice** – The program satisfies all five Working Group points for certifying a program effective. However, in our system one RCT evaluation is weighted the same as two quasi-experimental evaluations (e.g., NIJ Quasi-experimental evaluation Level 1)

II. **Effective** – An evaluator has replicated the program's effectiveness with an RCT design but the researcher was not an independent investigator. Again, one RCT evaluation is weighted the same as two NIJ Level 1 quasi-experimental evaluations.

III. **Promising** – There has been one successful RCT (or two Level 1 quasi-experiments) of the program but there is no replication study available ***OR*** the program matches the dimensions of a successful meta-analysis practice. Thus, programs that match the dimensions of other programs in successful meta-analyses are promising but not effective or fully evidence based programs. To be promising without direct outcome studies, programs must demonstrate that the treatments are similar enough to the successful programs in the meta-analysis ***and*** that enactment of the procedures that define the interventions show high levels of fidelity. Programs that are effective or fully evidence based (EBP) in Nebraska must show RCT or quasi-experimental evidence of their successful outcomes with studies conducted of those specific programs in their own settings. (See Model Programs and Effective Programs above.)

IV. **Inconclusive** – there has been one successful RCT (or two Level 1 quasi-experimental evaluations) of the program but there are contradictory findings in additional studies OR the program would be promising or effective but the effects are short in duration.

V. **Ineffective** – The RCT (or two quasi-experimental evaluations) failed to show significant differences between the treatment and control group.

VI. **Harmful** – the RCT (or two Level 1 quasi-experiments) showed that the control group scored higher on the targeted outcome than did the treatment group and the difference is statistically significant.

VII. **Insufficient Evidence** – there is no RCT or quasi-experimental evaluations of the program to date.

UNL/UNO Processes for Measuring Evidence Based Practice in Nebraska

The UNL/UNO collaborative evaluation team proposes the following four stage process to determine whether or not Nebraska Juvenile Justice programs are evidence based, to assist programs that show insufficient information or programs that are inconclusive, promising, or effective move up the EBP scale, and to help effective and model programs collect evaluation data (process, implementation, and outcome data) to maintain their current ranking. First, the team will establish a screening survey on a website that includes a series of questions for program administrators to answer. The questions will come directly from practice specific meta-analyses that match the program's intervention target (e.g., cognitive behavioral, counseling, family therapy and the like) and the Lipsey (2009) generic meta-analysis. These screening data will allow us to assign each program into 1 of 3 temporary classifications: 1) Temporary Level 1 for promising, effective, or model programs. 2) Temporary Level 2 for programs with inconclusive or insufficient evidence. 3) Temporary Level 3 for ineffective or potentially harmful programs. Second, the team will visit program sites, observe the programs in operation, interview program staff and clients, and complete a UNL/UNO Nebraska scoring card to classify the program into one of the 7 NJJHCS levels. This scoring card will be based upon the research

tools found on the Blueprints and OJJDP websites along with materials that other juvenile EB intervention rating systems use (e.g., Chorpita, Bernstein, & Daleiden, 2011). (Note: the Nebraska Commission on Law Enforcement and Criminal Justice has indicated they will award Juvenile Accountability funds to the research team for one year to develop the screening instrument and to visit 14 program sites to complete the second stage for those programs.) Third, for those sites that are effective or model programs on the NJJHCS, the UNL/UNO research team proposes to visit with the site administrators and assist them in designing program evaluation projects to collect implementation and limited outcome data to maintain the programs at their current levels on the scale, provided that funding is available. Fourth and finally, for inconclusive and insufficient evidence programs on the NJJHCS, the UNL/UNO research team proposes to assist those programs in conducting experimental or quasi-experimental independent evaluations that comply with the Working Group's recommendations for EBP, provided that funding is available. The UNL/UNO team plans to also offer these services to effective and promising programs to help increase their rankings on the NJJHCS scale.

References

Alexander, J. F., & Robbins, T. L. (2011). Functional Family Therapy. In R. C. Murrihy, A. D., Kidman, T. H. Ollendick (Eds.), *Clinical Handbook of Assessing and Treating Conduct Problems in Youth* (pp. 245-271). New York: Springer.

Chorpita, B. F., Bernstein, A., & Daleiden, E. L. (2011). Empirically guided coordination of multiple evidence-based treatments: An illustration of relevance mapping in children's mental health services. *Journal of Consulting and Clinical Psychology, 79*, 470-480. doi: 10.1037/a0023982

Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60*, 953-961. doi: 10.1037/0022-006X.60.6.953

Henggeler, S. W., Melton, S. W., Brondino, M. J., Sherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833. doi: 10.1037/0022-006X.65.5.821

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1*, 451-476. doi: 10.1007/s11292-005-3541-7

Latimer, J. (2001). A meta-analytic examination of youth delinquency, family treatment, and recidivism. *Canadian Journal of Criminology, 43*, 237-253. Retrieved from http://heinonline.org/HOL/LandingPage?handle=hein.journals/cjccj43&div=23&id=&page=

Lipsey, M. W. (1992). The effect of treatment on juvenile delinquents: Results from meta-analysis. In F. Loesel, D. Bender, & T. Bliesener (Eds.). *Psychology and law: International Perspectives* (pp. 131-143). Berlin; NY: Walter de Gruyter.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders: An International Journal of Evidence-based Research, Policy, and Practice, 4*, 124-147. doi: 10.1080/15564880802612573

Lipsey, M. W., Howell, J. C., Kelly, M. R., Chapman, G., & Carver, D. (2010). *Improving the*

effectiveness of juvenile justice programs: A new perspective on evidence-based practice. Washington D. C.: Georgetown University, Center for Juvenile Justice.

Lipsey, M. W., & Wilson, D. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.). *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions* (pp. 313-345). Thousand Oaks, CA: Sage.

Littell, J., Popa, M. & Forsythe, B. (2005). Therapy for social, emotional, and behavioral problems in youth aged 10-17. *Cochrane Database of Systemic Review, 4*, 1-42. doi: 10.1002/14651858.CD004797.pub4

MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of correctional boot camps on offending. *The Annals of the American Academy of Political and Social Science, 578*, 126-143. doi: 10.1177/000271620157800108

Mihalic, S., Irwin, K., Elliott, D., Fagan, A., & Hansen, D. (2001, July). *Blueprints for violence prevention* (OJJDP Juvenile Justice Bulletin). Washington D.C.: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Nugent, W. R., Williams, M., & Umbreit, M. S. (2004). Participation in victim-offender mediation and the prevalence of subsequent delinquent behavior: A meta-analysis. *Research on Social Work, 14*, 408-416. doi: 10.1177/1049731504265831

Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *The Annals of the American Academy of Political and Social Science, 589*, 41-62. doi: 10.1177/0002716203254693

Wilson, D. B., Mitchell, O., & MacKenzie, D. L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology, 2*, 459-487. doi: 10.1007/s11292-006-9019-4