Glossary of Terms

This guide is intended to provide states with information regarding the trustworthiness of different types of evidence, how evidence-based is defined, and how several relevant research terms are defined.

Evidence-based

Programs and practices are evidence-based when their effectiveness at producing a desired outcome has been positively demonstrated by causal evidence. Causal evidence is generally obtained through one or more outcome evaluations and depends on the use of rigorous scientific methods to rule out alternative explanations for the documented change.

A program is a planned, coordinated group of activities, processes, and procedures designed to achieve a specific purpose. A program is typically defined by its specific operating procedures and structured components (e.g., a defined curriculum, an explicit number of treatment or service hours, and an optimal length of treatment). Evidence-based programs that have been subject to rigorous scientific study are rated as effective/model programs or promising programs. Examples of programs include: Thinking for a Change (T4C); Big Brothers Big Sisters; and Law Enforcement Assisted Diversion (LEAD).

A practice is a general category of strategies or procedures that share similar characteristics to achieve a similar purpose. Examples of practices include: risk assessment and substance use/mental health screening.

In criminal justice, the evidence-based label also has been applied to principles that rigorous scientific studies have shown to be effective at producing a desired outcome. Examples of principles include the principles of effective correctional intervention, including the Risk, Need and Responsivity (RNR) principles.

Rigorous scientific methods include well designed and executed: randomized controlled trials (RCTs); quasi-experiments with a comparison condition; regression discontinuity studies; interrupted time series studies; and meta-analyses that are based on studies employing rigorous scientific methods.

Effective/Model Program

A program that has been given an evidence rating of effective/model indicates that the program has strong scientific evidence that it achieves its intended outcomes when implemented with fidelity. Typically, multiple highly rigorous studies (including at least one RCT) show evidence of effectiveness, and no rigorous studies show evidence of negative effects.

Promising Program

A program that has been given an evidence rating of promising indicates that the program has some rigorous scientific evidence that it achieves its intended outcomes; however, more extensive research is needed before a definitive conclusion regarding the program’s effectiveness can be made. Typically, at least one highly rigorous study shows evidence of effectiveness, and no rigorous studies show evidence of negative effects.

Evidence Continuum

Evidence in support of the effectiveness of a program, practice, or policy falls on a continuum ranging from very low to very high levels of confidence. The more rigorous the research design of evaluations and the greater the number of evaluations...
with positive outcomes, the greater confidence consumers can have that the intervention will produce its expected results.

**Types of Evidence**

*Opinion Informed*
Evidence with the lowest level of confidence is “opinion informed.” This includes information such as anecdotes, testimonials, and personal experiences obtained from a few individuals. This type of evidence, while useful in developing a program in the early stages, fails to examine targeted outcomes in a systematic way. It does not provide any real “proof” of effectiveness and ranks “very low” on the confidence continuum.

*Research Informed*
Research-informed studies rely on more than testimonial or professional insight by gathering data on program outcomes from surveys or other sources. Such studies provide some evidence of effectiveness, but the level of confidence is “low” because they do not isolate the impact of the program from other possible influences on observed outcomes.

Correlational studies, one type of research-informed study, can reveal if a relationship exists between a program and a desired outcome. However, demonstrating that a relationship exists does not prove that one variable caused the other. For example, a correlational study might show that being in a community-based treatment program was related to lower recidivism rates than being in a state correctional program, but this finding does not prove that the lower recidivism rate was due to being in the community program. Other factors, such as a judge sentencing more serious offenders to the correctional program, could explain the difference in recidivism. Correlational studies cannot show that it was the program that caused the difference in observed outcome.

Other research-informed studies collect data from program participants at posttest only or pretest and posttest. Because a control or comparison group is lacking, it is not clear that the program caused posttest outcomes or changes from pretest to posttest. Changes may well have occurred among similar subjects not going through the program.

Thus, research-informed studies lack an appropriate comparison group and evidence of a causal effect. These studies provide some preliminary support for a program that can help justify more rigorous experimental evaluation, but they rate low on the confidence continuum.

*Experimental and Experimentally Proven*
At the higher end of the confidence continuum are “experimental” and “experimentally proven” studies. These comprise what is commonly referred to as “evidence-based programs (EBPs).” Virtually all web-based registries of EBPs require experimental evidence for certification as an EBP. All experimental studies use designs that involve comparison or control groups. If participants receiving the program have better outcomes than those in the comparison or control groups, or those not receiving the program, then the program likely is having the intended effect.

At the moderate range of confidence are a set of designs that are commonly called quasi-experimental designs (QEDs). These designs all lack the element of random assignment that characterize randomized control trials (RCTs) and the certainty that the intervention and control groups are identical at the start of the study. A higher level of evidence comes from RCTs, where participants are randomly assigned to treatment and control groups. Randomization assures that the treatment and control groups are essentially identical on both measured and unmeasured characteristics at the start of the study. The highest level of confidence comes from multiple RCTs that show program benefits in different samples of randomly assigned subjects.
Direction of Effects
The direction of effects must also be considered. Studies can demonstrate positive effects, no effects, and harmful effects. Evidence that a program is ineffective or harmful requires the same level of evidence and confidence as the claim that a program is effective.

Research/Analysis Terms
Interrupted Time Series (ITS)
An interrupted time series is a quasi-experimental research/evaluation design in which data on an outcome of interest (i.e., crime) are collected at multiple time points before and after an intervention. These studies attempt to detect whether the intervention had a significantly greater effect than any underlying trend over time.

Meta-analysis
Meta-analysis is a research method for synthesizing the results of multiple studies in order to reach a conclusion about a program's effectiveness. It provides a means for combining the results of many evaluations of a program's effectiveness into one large study with many subjects (the total number of subjects from the individual studies). This is important, because single studies based on a small number of subjects can produce distorted findings about a program's effectiveness. By pooling the original studies, meta-analysis counteracts a common methodological problem in evaluation research – small sample size – thereby helping the analyst draw more accurate and generalizable conclusions about an intervention's effects. Moreover, a definitive conclusion about whether a program works needs to be based on an entire body of evidence about a program's effectiveness, taking into account both the trustworthiness of the evidence and the consistency of findings across studies.

Meta-analysis is especially useful for determining the effectiveness of generic interventions, as generic interventions tend to be the focus of a far larger number of evaluations than brand name programs. Additionally, meta-analysis reports its findings in terms of an average effect size, that is, the average effect of the program across multiple studies. This helps the evaluator gauge both the strength and consistency of a program's effect. Rather than using a vote-counting approach to determine the proportion of studies that found a statistically significant program effect, the evaluator can rely on a quantitative measure that more accurately captures and summarizes program results.

Regression Discontinuity Design
A quasi-experimental research/evaluation design for determining the effect of an intervention in which the assignment of study subjects to treatment and control conditions is determined on the basis of a cut-off score on a relevant pre-program measure. The design is advantageous because it produces highly trustworthy findings without potentially assigning study subjects most in need of treatment to a no-treatment control condition.

Other Relevant Terms
Brand-name Program
Brand-name programs are defined by their specific content and procedures, which are well documented (typically manualized) and uniformly followed in every application. Examples of brand-name programs include T4C and LEAD. All T4C programs have the same content and follow the same delivery procedures regardless of the implementation site.

To be effective, a brand name program typically must be implemented with a high degree of fidelity — that is, those responsible for implementing the program must closely follow or adhere to the content and protocols that are defined as part of the intervention. Deviating from that content or the specified delivery procedures can degrade the program’s effectiveness or result in a situation where an otherwise effective program does more harm than good.
Core Program Elements

Core program elements (or core components) are the parts of a program that rigorous scientific research has identified as being vital for program success. They are essential features that should not be eliminated or altered prior to or during program implementation. If they are altered, or left out in whole or in part, the nature of the program itself changes, and the expectation for a positive result that accompanies an evidence-based program is no longer valid.

Generic Intervention

Generic interventions are defined by the general nature and focus of their services, rather than by uniform content and a specific set of procedures that are followed in every application. Examples of generic interventions include drug courts, mental health counseling, restorative justice programs, and treatment for sexual offenders. All drug courts, for example, have a similar nature and focus, yet a drug court program’s specific content and procedures can vary from one jurisdiction to the next. For some generic interventions, such as drug courts, core program elements have been identified through scientific study, and ensuring that these elements are incorporated when the program is implemented helps maximize program success.

Implementation Fidelity

Implementation fidelity refers to the degree to which a program’s implementation in any real-world setting matches the content and procedures specified in the original program model.

No matter where an evidence-based program is adopted and implemented, the goal should be to achieve the highest degree of implementation fidelity possible. In other words, an organization should always strive to deliver the program’s content and adhere to its procedures precisely as prescribed in the program model. Adhering to the model as closely as possible substantially increases the odds of attaining positive outcomes. Conversely, deviating from the model can not only degrade program effectiveness, it can create a situation where the program does more harm than good.