

EFFECT: The effectiveness of a program is based on the program's ability to impact the intended outcomes (i.e., program effect).

- **Harmful:** A program is harmful if it has been shown to lead to harmful outcomes (i.e., rigorous designs have been used – true experimental designs, quasi-experimental designs). A program can be harmful due to its inherent nature (i.e., some aspects of the program are harmful) or certain settings/contexts in which it is implemented.
- **Ineffective:** Ineffective programs have been evaluated in at least two well-done studies (true or quasi-experimental design), and no impact on outcomes has been demonstrated.
- **Effect is Undetermined:** Programs that have not been evaluated or have been evaluated poorly (neither a true nor quasi-experimental design was used, not based on sound theory) are considered to have undetermined effectiveness.
- **Potentially Positive Effect:** Some programs may have demonstrated some positive effects, are based on sound theory, but have been evaluated with a less rigorous design (e.g., exploratory study). Although no causal inference can be made as no rigorous designs were used (i.e., true experimental design, quasi-experimental design), there are indications that such programs may have potentially positive effects.
- **Some Evidence of Effectiveness:** Some programs may not have been rigorously evaluated more than once but are based on sound theory and results indicate positive effects. Thus, these programs show preliminary evidence of effectiveness; however, such effects cannot be considered as convincing as those that have been through a least two rigorous evaluations, and shown to positively modify intended outcomes.
- **Found to be Effective:** Effective programs are based on sound theory, have been rigorously evaluated in at least two studies (true or quasi-experimental design), and have exhibited significant positive effects.

TYPE OF EVIDENCE/RESEARCH EVIDENCE: Elements of a research study, such as participant selection and group assignment, determine the research design. The type of design determines how well effectiveness is measured. The more rigorous the design, the higher its internal validity, which means a greater likelihood that outcomes can be attributed to the program being evaluated.

- **Randomized control trial/ Systematic review/ Meta-Analysis:**

Randomized control trials are true experiments and are considered the most robust research design. Due to random assignment, they allow us to establish a cause-and-effect relation between the intended outcomes and the program. Individual randomized-controlled trial studies may be examined further in a systematic review or meta-analysis, which provide even more rigorous information on program effectiveness.

Well-conducted systematic reviews are a collection of high-quality studies (i.e., studies using randomized control trials). Researchers conducting these reviews, evaluate any study on the specific topic for quality. Given the high quality of a study, the researchers synthesize the findings across those high-quality studies. A meta-analysis is a type of systematic review that statistically summarizes the data from individual studies and uses these combined findings to provide statements about the overall effect. The strongest meta-analyses consist of independent studies (conducted by different researchers) and that are comparable (e.g., similar samples, methods, procedures) and of high-quality.

- **Quasi-Experimental Design:**

Quasi-experimental designs employ at least two groups of participants but without random assignment to the groups (e.g., treatment and comparison groups). Optimally, the study includes multiple measurement points (e.g., longitudinal design). Such designs are considered rigorous designs, albeit not as rigorous as randomized control trials, because the different groups may not be equivalent prior to the intervention/programming.

- **Single-Group Design:**

Single-group designs are designs in which one group of participants' data is collected at either one (i.e., post-test) or two time-points (pre- and post-test). The participants experienced the intervention/program/treatment. Such designs are not as rigorous as randomized control trials or quasi-experimental designs because they do not include a control (i.e., randomized control trials) or comparison group (i.e., quasi-experimental designs).

- **Exploratory Study:**

Exploratory studies aim to explore the programs they address. Exploratory studies are based on sound theory/previous research and/or expert opinion. Information taken from exploratory studies may point to important factors to consider in developing or refining a program. Types of exploratory studies included ethnographies (i.e., the study of people in their own environment), focus groups (i.e., a collective interview of a group of people from similar background),

sociometric (i.e., qualitative method that measures aspects of social relationships), and narrative analyses (i.e., a qualitative method in which a researcher collects and analyzes individual stories regarding various life situations).

- **Anecdotal information/Needs Assessment:**

Anecdotal information (e.g., “stories”, testimonials) is information not supported by data. Anecdotal evidence cannot inform program effectiveness as it is not empirical evidence. Although often based on empirical evidence, needs assessments (i.e., a type of assessment used for determining and addressing needs) cannot inform program effectiveness as they are not used for program effectiveness evaluation purposes.

INTERNAL VALIDITY: Refers to the extent to which the outcomes of a program can truly be attributed to the program or if these outcomes could have been caused by something else. The higher the internal validity, the more confidently a claim can be made that programming is truly producing the effects.

- **True Experimental Design:**

True experimental designs (i.e., randomized-controlled trials) are considered to be highly internally valid as participants are randomly assigned to the treatment (i.e., program) and control (i.e., some other experience or activity that is not the program) conditions. This design helps establish that changes in intended outcomes can be attributed to the program rather than other influences. The most optimal experimental designs measure the outcomes at multiple time points and have at least two groups: a treatment and a control group. Such designs allow for causal inferences about program effectiveness.

- **Quasi-Experimental Design:**

Quasi-experimental designs are also highly internally valid, although less internally valid than true experimental designs. Quasi-experimental designs have comparison groups, but no random assignment to the comparison versus treatment groups. Thus, participants in the two groups may differ on important characteristics related to the outcomes prior to programming, which impacts inferences about program effectiveness.

- **Non-Experimental Design:**

Non-experimental designs are weaker than true-experimental and quasi-experimental designs in terms of internal validity because they do not have a

control or comparison group. Thus, it is impossible to attribute observed changes in intended outcomes to the program. An example of a non-experimental design would be a single-group design where participants experienced the program and data was collected before and after the program.

- **Sound Theory Only:**

The program is based on sound theory; there is no empirical evidence of its effectiveness.

- **No Research/No Sound Theory:**

Programs not based on research or sound theory are considered the weakest of all in terms of potential program effectiveness. Without research or theory, there is no evidence to suggest that the program should lead to changes in intended outcomes.

INDEPENDENT REPLICATION: Involves duplicating the implementation of programming with another group of participants to determine whether the same effects are achieved. Replication should be independent, meaning replications should be implemented and evaluated by researchers/practitioners who are unaffiliated & have no conflicts of interest.

- **Program Replication with Evaluation Replication**

The most reliable programs (i.e., the ability to reproduce the original program effects) are replicated by different researchers/practitioners at least once, in a similar setting as the original program implementation (e.g., different students in a different school), using a rigorous research design (i.e., randomized control trial or quasi-experimental design), and are implemented and evaluated in the same way or very similarly to the original program.

- **Program Replication without Evaluation Replication:**

Moderately reliable programs (i.e., the ability to reproduce the original program effects in certain occasions) are implemented in a similar way and in similar settings as the original program (e.g., a different school with different students). Replications may or may not be conducted by different researchers/practitioners and are not evaluated the same way as the original evaluation of the program.

- **Partial Program Replication without Evaluation Replication:**

Partially replicated programs (i.e., partial implementation of the original program) that have not been evaluated are weakest in terms of reliability (i.e., there is no

means to evaluate the consistency of the original program effects). Such replications may or may not be conducted by different researchers/practitioners.

IMPLEMENTATION GUIDANCE: Includes any materials that aid in the implementation of a program/strategy/intervention. Implementation guidance is essential because programming is more likely to be adequately administered and thus achieve its effect when appropriate guidance and directions are provided.

- **Comprehensive:**
Comprehensive guidance entails the availability and accessibility of any products, services, or activities (i.e., training, coaching, technical assistance, support materials, manuals, guides) that facilitate proper program implementation in a new setting. Comprehensively guided programs can be implemented with high fidelity.
- **Partial:**
Partial guidance entails limited access and availability of products or services that help in the implementation of a program in a different setting. Limited guidance and support may lead to implementation issues which limits causal inferences about program effectiveness.
- **None:**
No guidance entails no access or availability of products or services that help in the implementation of a program in a different setting. Such programs are at a high risk of experiencing implementation issues.

EXTERNAL VALIDITY/ECOLOGICAL VALIDITY: Whether programming can demonstrate effects (positive/negative) among a wide range of populations and contexts.

- **Applied Studies – Different Settings (2+):**
Programs high in external and ecological validity have been implemented in two or more applied (“real world”) settings that differ from the original setting and are independent of each other in terms of population and geographical locations.
- **Applied Studies – Same Settings (2+):**
Programs moderate in external and ecological validity have been implemented in two or more applied (“real world”) settings that are similar to one another with similar populations.

- **Not Real-World Informed:**

The least externally and ecologically valid programs are those with a structure not consistent with an applied setting and whose implementation do not reflect the “real world”. Hence, nothing can be said regarding program effectiveness in applied, real world settings.

- **Somewhat Real-World Informed:**

Somewhat real-world informed programs have not been implemented in applied settings nor are structured/implemented in ways fully consistent with an applied setting. However, such programs may still be considered externally and ecologically valid if parts of the programs’ structure/implementation approximate “real-world” conditions.

- **Real-World Informed:**

Real-world informed programs may not have been implemented in applied settings. However, programs may still be considered externally and ecologically valid if they are implemented using materials and resources that would be used in applied settings or delivered in a manner as it would be delivered in the “real world”.