***Resource for this paper:*** <https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf>

While the ESEA definition of “evidence-based” states that “at least one study” is needed to provide *strong evidence*, *moderate evidence*, or *promising evidence* for an intervention, SEAs, districts, and other stakeholders should consider the entire body of relevant evidence. Additionally, when available, interventions supported by higher levels of evidence, specifically *strong evidence* and *moderate evidence*, which describe the *effectiveness of an intervention*[[1]](#footnote-1) through *causal inference*[[2]](#footnote-2), vii should be prioritized. Stakeholders should also consider whether there is evidence that an intervention has substantially improved an important education outcome (e.g., credit accumulation and high school graduation). The What Works Clearinghouse (WWC), an initiative of USED’s Institute of Education Sciences, is a helpful resource for locating the evidence on various education interventions.[[3]](#footnote-3)

|  |
| --- |
| WHAT IS AN “EVIDENCE-BASED” INTERVENTION? (from section 8101(21)(A) of the ESEA) “…the term ‘evidence-based,’ when used with respect to a State, district, or school activity, means an activity, strategy, or intervention that –   * demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on –   + evidence from at least one well-designed and well-implemented experimental study;   + moderate evidence from at least one well-designed and well-implemented quasi experimental study; or   + promising evidence from at least one well-designed and well-implemented correlational study with statistical controls for selection bias; or   demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes; and includes ongoing efforts to examine the effects of such activity, strategy, or intervention. |

***Strong Evidence.*** To be supported by *strong evidence*, there must be at least one well-designed and well-implemented experimental study (e.g., *a randomized control trial*[[4]](#footnote-4)) on the intervention. The Department of Education considers an experimental study to be “well-designed and well-implemented” if it meets [*What Works Clearinghouse* (WWC)](http://ies.ed.gov/ncee/wwc/) *Evidence Standards without reservations[[5]](#footnote-5)* or is of the equivalent quality for making *causal inferences*. Additionally, to provide *strong evidence*, the study should:

1. Show a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other *relevant outcome*;[[6]](#footnote-6)
2. Not be overridden by statistically significant and negative (i.e., unfavorable) evidence on the same intervention in other studies that meet *WWC Evidence Standards with or without reservations*[[7]](#footnote-7) or are the equivalent quality for making *causal inferences*;
3. Have a *large sample*[[8]](#footnote-8) and a *multi-site sample*;[[9]](#footnote-9) and
4. Have a sample that overlaps with the populations (i.e., the types of students served)[[10]](#footnote-10) and settings (e.g., rural, urban) proposed to receive the intervention.

***Moderate Evidence.*** To be supported by *moderate evidence,* there must be at least one well-designed and well-implemented *quasi-experimental study*[[11]](#footnote-11) on the intervention. The Department considers a quasi-experimental study to be “well-designed and well-implemented” if it meets *WWC Evidence Standards with reservations* or is of the equivalent quality for making *causal inferences*. Additionally, to provide *moderate evidence*, the study should:

1. Show a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other relevant outcome;
2. Not be overridden by statistically significant and negative (i.e., unfavorable) evidence on that intervention from other findings in studies that meet *WWC Evidence Standards with or without reservations* or are the equivalent quality for making *causal inferences*;
3. Have a *large sample* and a *multi-site sample*; and
4. Have a sample that overlaps with the populations (i.e., the types of students served) OR settings (e.g., rural, urban) proposed to receive the intervention.

***Promising Evidence.*** To be supported by *promising evidence*, there must be at least one well-designed and well-implemented correlational study with statistical controls for selection bias[[12]](#footnote-12) on the intervention. The Department considers a correlational study to be “well-designed and well-implemented” if it uses sampling and/or analytic methods to reduce or account for differences between the intervention group and a comparison group. Additionally, to provide *promising evidence*, the study should:

1. Show a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other *relevant outcome*; and
2. Not be overridden by statistically significant and negative (i.e., unfavorable) evidence on that intervention from findings in studies that meet *WWC Evidence Standards with or without reservations* or are the equivalent quality for making *causal inferences*.

***Demonstrates a Rationale.*** To demonstrate a rationale, the intervention should include:

1. A well-specified *logic model*[[13]](#footnote-13) that is informed by research or an evaluation that suggests how the intervention is likely to improve *relevant outcomes*; and
2. An effort to study the effects of the intervention, ideally producing promising evidence or higher, that will happen as part of the intervention or is underway elsewhere (e.g., this could mean another SEA, LEA, or research organization is studying the intervention elsewhere), to inform stakeholders about the success of that intervention.

1. The effectiveness of the intervention is measured in a rigorous study (e.g. one that allows for causal inference) as the difference between the average outcomes for the two groups in the study. [↑](#footnote-ref-1)
2. Causal inference is the process of drawing a conclusion that an activity or intervention was likely to have affected an outcome. [↑](#footnote-ref-2)
3. WWC is available at <http://ies.ed.gov/ncee/wwc/> [↑](#footnote-ref-3)
4. An experimental study is designed to compare outcomes between two groups of individuals that are otherwise equivalent except for their assignment to either the intervention group or the control group. A common type of experimental study is a randomized control trial or RCT. A randomized controlled trial, as defined by Part 77.1 of the Education Department General Administration Regulations (EDGAR), is a study that employs random assignment of, for example, students, teachers, classrooms, schools, or districts to receive the intervention being evaluated (the treatment group) or not to receive the intervention (the control group). The estimated effectiveness of the intervention is the difference between the average outcomes for the treatment group and for the control group. These studies, depending on design and implementation, can meet What Works Clearinghouse Evidence Standards without reservations. An RCT, for example, may look at the impact of participation in a magnet program that relies on a lottery system for admissions. The treatment group could be made up of applicants admitted to the magnet program by lottery and the control group could be made up of applicants that were not admitted to the magnet program by lottery. If an RCT is well-designed and well-implemented, then students in the treatment and control groups are expected to have similar outcomes, on average, except to the extent that the outcomes are affected by program admission. The comparability of the two groups could be compromised if there are problems with design or implementation, which may include problems with sample attrition, changes in group status after randomization, and investigator manipulation. [↑](#footnote-ref-4)
5. WWC Evidence Standards without reservations is the highest possible rating for a group design study reviewed by

   the WWC. Studies receiving this rating provide the highest degree of confidence that an observed effect was caused

   by the intervention. Well-implemented randomized controlled trials (i.e., RCTs that are not compromised by

   problems like attrition) may receive this highest rating. These standards are described in the WWC Procedures and

   Standards Handbook, which can be accessed at http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19. [↑](#footnote-ref-5)
6. A relevant outcome, as defined by Part 77.1 of EDGAR, means the student outcome(s) (or the ultimate outcome if not related to students) the proposed process, product, strategy, or practice is designed to improve; consistent with the specific goals of a program. [↑](#footnote-ref-6)
7. WWC Evidence Standards with reservations is the middle possible rating for a group design study reviewed by the WWC. Studies receiving this rating provide a lower degree of confidence that an observed effect was caused by the intervention. RCTs that are not as well implemented or have problems with attrition, along with strong quasi experimental designs, may receive this rating. These standards are described in the WWC Procedures and Standards Handbook, which can be assessed at http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19. [↑](#footnote-ref-7)
8. A large sample, as defined by Part 77.1 of EDGAR, is an analytic sample of 350 or more students (or other single analysis units), or 50 or more groups (such as classrooms or schools) that contain 10 or more students (or other single analysis units). As EDGAR provides, multiple studies can cumulatively meet the large sample requirement and the multi-site sample requirement, as long as each study meets the other requirements corresponding with the specific level of evidence. [↑](#footnote-ref-8)
9. A multi-site sample, as defined by Part 77.1 of EDGAR, consists of more than one site, where site can be defined as an LEA, locality, or State. As EDGAR provides, multiple studies can cumulatively meet the large sample requirement and the multi-site sample requirement, as long as each study meets the other requirements corresponding with the specific level of evidence. [↑](#footnote-ref-9)
10. In order to demonstrate overlap with the population, the study or studies should show that the intervention has a statistically significant and positive effect on the specific population and/or subgroup of interest being served by the intervention. xvi A quasi-experimental study (as known as a quasi-experimental design study or QED), as defined by Part 77.1 of EDGAR, means a study using a design that attempts to approximate an experimental design by identifying a comparison group that is similar to the treatment group in important respects. These studies, depending on design and implementation, can meet What Works Clearinghouse Evidence Standards. An example of a QED is a study comparing outcomes for two groups of classrooms matched closely on the basis of student demographics and prior mathematics achievement, half of which are served by teachers who participated in a new mathematics professional development (PD) program, and half of which are served by other teachers. This study uses a nonequivalent group design by attempting to match or statistically control differences between the two groups. Another type of QED is a regression discontinuity design (RDD), which uses a cutoff or threshold above or below which an intervention is assigned to individuals. [↑](#footnote-ref-10)
11. A quasi-experimental study (as known as a quasi-experimental design study or QED), as defined by Part 77.1 of EDGAR, means a study using a design that attempts to approximate an experimental design by identifying a comparison group that is similar to the treatment group in important respects. These studies, depending on design and implementation, can meet What Works Clearinghouse Evidence Standards. An example of a QED is a study comparing outcomes for two groups of classrooms matched closely on the basis of student demographics and prior mathematics achievement, half of which are served by teachers who participated in a new mathematics professional development (PD) program, and half of which are served by other teachers. This study uses a nonequivalent group design by attempting to match or statistically control differences between the two groups. Another type of QED is a regression discontinuity design (RDD), which uses a cutoff or threshold above or below which an intervention is assigned to individuals. [↑](#footnote-ref-11)
12. A correlational study with statistical controls for selection bias is designed to examine the strength of the relationship (not the causal relationship) between an intervention and a student outcome by comparing two similar groups. In an example correlational study, researchers may look at how two classrooms with similar characteristics perform on a reading assessment after one of the classes (the treatment group) participates in a new reading program. While the researcher is looking at outcomes in classrooms that look similar, there may be other important differences between the classrooms (e.g. previous reading assessment scores) that are not accounted for, but would be in more rigorous studies like experimental studies or QEDs. These types of studies cannot meet WWC standards. [↑](#footnote-ref-12)
13. A logic model (also known as a theory of action), as defined by Part 77.1 of EDGAR, means a well-specified conceptual framework that identifies key components of the proposed process, product, strategy, or practice (i.e., the active “ingredients” that are hypothesized to be critical to achieving the relevant outcomes) and describes the relationships among the key components and outcomes, theoretically and operationally. More information on logic models can be found at http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2015057. [↑](#footnote-ref-13)