Non-Regulatory Guidance:
Using Evidence to Strengthen Education Investments

September 16, 2016

Purpose of the Guidance

The U.S. Department of Education (Department) has determined that this guidance is significant guidance under the Office of Management and Budget's Final Bulletin for Agency Good Guidance Practices, 72 Fed. Reg. 3432 (Jan. 25, 2007). See www.whitehouse.gov/sites/default/files/omb/memoranda/fy2007/m07-07.pdf. *Significant guidance is non-binding and does not create or impose new legal requirements.*

The Department is issuing this guidance to provide State educational agencies (SEAs), local educational agencies (LEAs), schools, educators, and partner organizations with information to assist them in selecting and using "evidence-based" activities, strategies, and interventions, as defined in Title VIII of the Elementary and Secondary Education Act of 1965 (ESEA), as amended by the Every Student Succeeds Act of 2015 (ESSA). If you are interested in commenting on this guidance, please email us your comment at OESEGuidanceDocument@ed.gov *or write to us at the following address:*

> U.S. Department of Education
> Office of Elementary and Secondary Education
> 400 Maryland Avenue, S.W.
> Washington, D.C. 20202

For further information about the Department's guidance processes, please visit www2.ed.gov/policy/gen/guid/significant-guidance.html.

The Department does not mandate or prescribe practices, models, or other activities in this non-regulatory guidance document. This guidance contains examples of, adaptations of, and links to resources created and maintained by other public and private organizations. This information, informed by research and gathered in part from practitioners, is provided for the reader's convenience and is included here to offer examples of the many resources that educators, parents, advocates, administrators, and other concerned parties may find helpful and use at their discretion. The Department does not control or guarantee the accuracy, relevance, timeliness, or completeness of this outside information. Further, the inclusion of links to items and examples do not reflect their importance, nor are they intended to represent or be an endorsement by the Department of any views expressed, or materials provided.

Introduction

Using, generating, and sharing evidence about effective strategies to support students gives stakeholders an important tool to accelerate student learning. ESEA[1] emphasizes the use of evidence-based activities, strategies, and interventions (collectively referred to as "interventions"). This guidance is designed to help SEAs, LEAs, schools, educators, partner organizations and other stakeholders successfully choose and implement interventions that improve outcomes for students. Part I of this guidance reviews steps for effective decision-making and Part II of this guidance recommends considerations, resources, and criteria for identifying "evidence-based" interventions based on each of ESSA's four evidence levels in Section 8101(21)(A) of the ESEA.

---

[1] Throughout this document, unless otherwise indicated, citations to the ESEA refer to the ESEA, as amended by the ESSA.

Part I: Strengthening the Effectiveness of ESEA Investments

Ways to strengthen the effectiveness of ESEA investments include identifying local needs, selecting evidence-based interventions that SEAs, LEAs, and schools have the capacity to implement, planning for and then supporting the intervention, and examining and reflecting upon how the intervention is working. These steps, [i] when taken together, promote continuous improvement and can support better outcomes for students. Links to resources, definitions for *italicized words,* and other relevant information are included in endnotes.
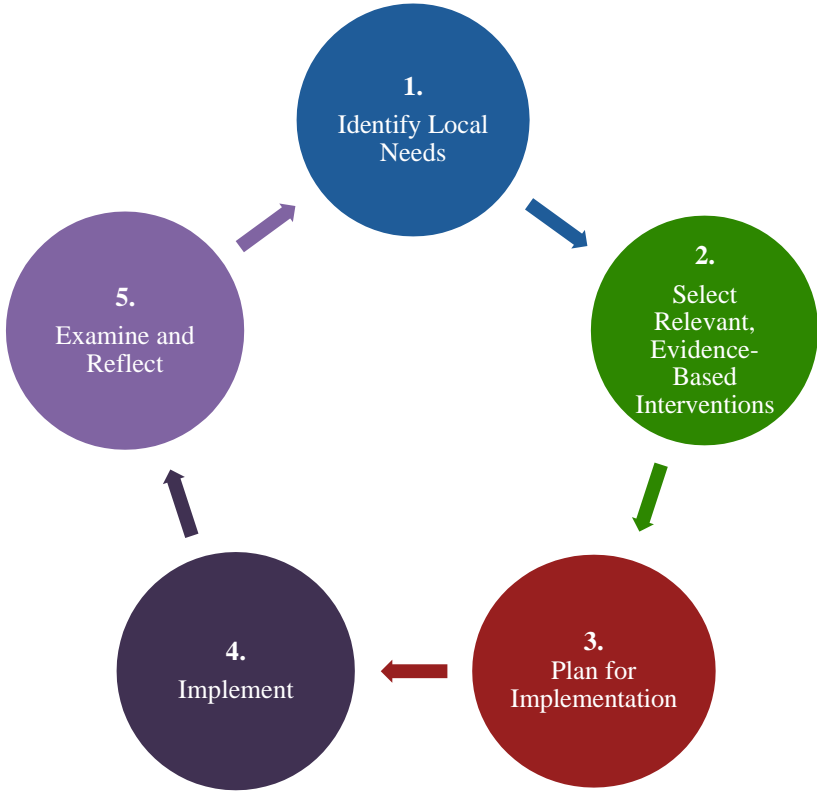
Step 1. IDENTIFY LOCAL NEEDS

SEAs and LEAs should engage in timely and meaningful consultation with a broad range of stakeholders (e.g., families, students, educators, community partners) and examine relevant data to understand the most pressing needs of students, schools, and/or educators and the potential root causes of those needs. Interviews, focus groups, and surveys as well as additional information on students (e.g., assessment results, graduation rates), schools (e.g., resources, climate) and educators (e.g., effectiveness, retention rates) provide insights into local needs. Some questions to consider:

➢ Which stakeholders can help identify local needs and/or root causes?[ii]
➢ What data are needed to best understand local needs and/or root causes?
➢ How do student outcomes compare to identified performance goals? Are there inequities in student resources or outcomes within the State or district?

➢ What are the potential root causes of gaps with performance goals or inequities?
➢ How should needs be prioritized when several are identified?

Step 2. SELECT RELEVANT, EVIDENCE-BASED INTERVENTIONS

Once needs have been identified, SEAs, LEAs, schools, and other stakeholders will determine the interventions that will best serve their needs. By using rigorous and relevant evidence and assessing the local capacity to implement the intervention (e.g., funding, staff, staff skills, stakeholder support), SEAs and LEAs are more likely to implement interventions successfully. Those concepts are briefly discussed below (also see Part II of this guidance for more information on evidence-based interventions):

- While ESEA requires "at least one study" on an intervention to provide *strong evidence*, *moderate evidence*, or *promising evidence,* SEAs, LEAs, and other stakeholders should consider the entire body of relevant evidence.
- Interventions supported by higher levels of evidence, specifically *strong evidence* or *moderate evidence*, are more likely to improve student outcomes because they have been proven to be effective. When *strong evidence* or *moderate evidence* is not available, *promising evidence* may suggest that an intervention is worth exploring. Interventions with little to no evidence should at least *demonstrate a rationale* for how they will achieve their intended goals and be examined to understand how they are working.
- The relevance of the evidence – specifically the setting (e.g., elementary school) and/or population (e.g., students with disabilities, English Learners) of the evidence – may predict how well an evidence-based intervention will work in a local context (for more information, also see Part II and endnotes). SEAs and LEAs should look for interventions supported by *strong evidence* or *moderate evidence* in a similar setting and/or population to the ones being served. The [What Works Clearinghouse™ (WWC)](#) uses rigorous standards to review evidence of effectiveness on a wide range of interventions and also summarizes the settings and populations in the studies.[iii]
- Local capacity also helps predict the success of an intervention, so the available funding, staff resources, staff skills, and support for interventions should be considered when selecting an evidence-based intervention. SEAs can work with individual and/or groups of LEAs to improve their capacity to implement evidence-based interventions.

Some questions to consider about using evidence:
➢ Are there any interventions supported by *strong evidence* or *moderate evidence*?
➢ What do the majority of studies on this intervention find? Does the intervention have positive and statistically significant effects on important student or other *relevant outcomes*, or are there null, negative, or not statistically significant findings?
➢ Were studies conducted in settings and with populations relevant to the local context (e.g., students with disabilities, English Learners)?
➢ If *strong evidence* or *moderate evidence* is not available, is there *promising evidence*?

- ➢ Does the intervention *demonstrate a rationale* that suggests it may work (e.g., it is represented in a *logic model* supported by research)?
- ➢ How can the success of the intervention be measured?

Some questions to consider about local capacity:
- ➢ What resources are required to implement this intervention?
- ➢ Will the potential impact of this intervention justify the costs, or are there more cost-effective interventions that will accomplish the same outcomes?
- ➢ What is the local capacity to implement this intervention? Are there available funds? Do staff have the needed skills? Is there buy-in for the intervention?
- ➢ How does this intervention fit into larger strategic goals and other existing efforts?
- ➢ How will this intervention be sustained over time?

Step 3. PLAN FOR IMPLEMENTATION

An implementation plan, developed with input from stakeholders, sets up SEAs, LEAs and schools for successful implementation. [iv] Implementation plans may include the following components:
- ✓ A *logic model* that is informed by research or an evaluation that suggests how the intervention is likely to improve *relevant outcomes* (also see Part II of guidance for more information on logic models);
- ✓ Well-defined, measurable goals;
- ✓ Clearly outlined roles and responsibilities for people involved, including those implementing the intervention on the ground, those with a deep understanding of the intervention, and those ultimately responsible for its success;
- ✓ Implementation timelines for successful execution;
- ✓ Resources required to support the intervention; and
- ✓ Strategies to monitor performance and ensure continuous improvement, including plans for data collection, analysis and/or an evaluation (also see Step 5 in this guidance).

Step 4. IMPLEMENT

Implementation will impact the ultimate success of an intervention, so SEAs, LEAs, and schools should have ways to collect information about how the implementation is working and make necessary changes along the way. Some questions to consider:
- ➢ What information will be collected to monitor the quality of implementation? Is additional information needed to understand how the implementation is working?
- ➢ Is the implementation plan being followed? If not, why not? Are changes needed?
- ➢ Are more resources required? Do resources need to be realigned or timelines adjusted? Are stakeholders being engaged?
- ➢ What are unforeseen barriers to successful implementation?
- ➢ How is implementation working with other existing efforts?

- ➢ What does the information being collected suggest about the success of the implementation?
- ➢ Are changes needed to improve the implementation?
- ➢ Is the intervention ready to be scaled to more students or educators?

Step 5. EXAMINE AND REFLECT

As part of implementation and decision-making, there are different ways to examine how interventions are working. Performance monitoring involves tracking data about an intervention to see how performance compares to identified targets and goals. Rigorous evaluations measure the *effectiveness of an intervention*, answering questions about the impact of a specific intervention on *relevant outcomes*. These types of information are most valuable when shared with key stakeholders for decision-making. Both concepts are briefly described below (also see Part II of this guidance for information on the different levels of evidence):

- Performance monitoring involves regularly collecting and analyzing data in order to track progress against targets and goals. Performance monitoring can help identify whether key elements of a *logic model* are being implemented as planned and whether the intervention is meeting interim goals and milestones, and suggest ways the intervention could be changed for continuous improvement. Performance information can also provide insight into whether the expected outcomes are being achieved. This constitutes examining the effects of an intervention, as mentioned in evidence that *demonstrates a rationale*.
- Evaluations of effectiveness may be appropriate when SEAs and/or LEAs want to know if an intervention affected the intended student or educator outcomes. These types of evaluations may meet *strong evidence* or *moderate evidence* levels, as defined in ESEA section 8101(21) and clarified in Part II of this guidance.[v]

Some questions to consider:
- ➢ What are reasonable expectations of success and how can success be measured?
- ➢ What are interim progress and performance milestones that can be tracked?
- ➢ What have participants (i.e., students and educators) in the intervention shared about their experience and how the intervention was implemented?
- ➢ Is there the need and capacity to evaluate the *effectiveness of an intervention* through a study that could produce *strong evidence* or *moderate evidence*, or would *promising evidence* from a correlational study or performance data that *demonstrates a rationale* suffice?
- ➢ How could knowledge about this intervention be shared with others andincorporated into decision-making going forward?
- ➢ Based on information, should this intervention continue as is, be modified, or be discontinued?

Part II: Guidance on the Definition of "Evidence-Based"

Evidence is a powerful tool to identify ways to address education problems and build knowledge on what works. ESEA emphasizes the use of evidence-based activities, strategies, and interventions (collectively referred to as "interventions"). Section 8101(21)(A) of the ESEA defines an evidence-based intervention as being supported by *strong evidence, moderate evidence, promising evidence,* or evidence that *demonstrates a rationale* (see text box below). Some ESEA programs encourage the use of "evidence-based" interventions while others, including several competitive grant programs and Title I, section 1003 funds, require the use of "evidence-based" interventions that meet higher levels of evidence.

In order to help SEAs, LEAs, schools, educators, and partner organizations (collectively referred to as "stakeholders") understand and identify the rigor of evidence associated with various interventions, below are the recommended considerations, resources, and criteria for each of ESSA's four evidence levels. These recommendations are applicable to all programs in ESSA. This guidance does not address the specific role of evidence in each ESSA program and therefore should be used in conjunction with program-specific guidance. *Italicized* words are defined in the endnotes.

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, local educational agency, or school activity, means an activity, strategy, or intervention that –

   (i)   demonstrates a statistically significant effect on improving student outcomes or other *relevant outcomes* based on –

     (I)   *strong evidence* from at least one well-designed and well-implemented experimental study;

     (II)   *moderate evidence* from at least one well-designed and well-implemented *quasi-experimental study*; or

     (III)   *promising evidence* from at least one well-designed and well-implemented correlational study with statistical controls for selection bias; or

   (ii)   (I) *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

     (II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention.

Evidence Considerations, Resources, and Criteria for Levels

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, LEAs, 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*[vi] through *causal inference,*[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 ED's Institute of Education Sciences, is a helpful resource for locating the evidence on various education interventions.[viii] For a longer discussion of key steps and considerations for decision-making, including but not limited to the use of evidence-based interventions, see Part I of this guidance.

The criteria below represent the Department's recommendations for identifying evidence at each of the four levels in ESEA (also summarized in Table 1 on page 12).

❖ ***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*[ix]) on the intervention. The Department considers an experimental study to be "well-designed and well-implemented" if it meets *WWC Evidence Standards without reservations*[x] or is of the equivalent quality for making *causal inference*s. 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;*[xi]

    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*[xii] or are the equivalent quality for making *causal inferences*;

    3) Have *a large sample*[xiii] and *a multi-site sample*[xiv]; and

    4) Have a sample that overlaps with the populations (i.e., the types of students served) [xv] 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*[xvi] 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[xvii] 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*[xviii] 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.

---

[i] These steps largely draw from existing decision-making frameworks and take place as part of a continuous cycle.
[ii] See here for the Department's policy letter on stakeholder engagement and here for a communication and engagement rubric for information on how they can be engaged in meaningful ways.

iii In addition to the WWC, evidence resources like the Department's Regional Educational Laboratories (RELs) and other federally-funded technical assistance centers may provide summaries of the evidence on various interventions and guidance on how existing research aligns to the ESEA evidence levels discussed in the Part II of this guidance.

iv See here for an implementation planning and monitoring tool.

v In order to ensure these evaluations of effectiveness produce credible results, SEAs or LEAs can leverage Department of Education technical assistance, including working with local RELs to plan, implement, and conduct evaluations and/or by using supporting resources like this free software to simplify analysis and reporting of evaluation results.

vi 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.

vii *Causal inference* is the process of drawing a conclusion that an activity or intervention was likely to have affected an outcome.

viii WWC is available at http://ies.ed.gov/ncee/wwc/.

ix 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.

x *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.

xi 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.

xii *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.

xiii 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.

xiv 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.

[xv] 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.

[xvii] 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.

[xviii] 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.

Table 1. Summary of Recommended Study Criteria for Each Evidence Level

	Strong Evidence	Moderate Evidence	Promising Evidence	Demonstrates a Rationale
Study Design	Experimental study	Quasi-experimental study	Correlational study with statistical controls for selection bias	Provides a well-specified logic model informed by research or evaluation
WWC Standard	Meets WWC Evidence Standards <u>without</u> reservations (or is the equivalent quality)	Meets WWC Evidence Standards <u>with</u> or <u>without</u> reservations (or is the equivalent quality)	N/A	N/A
Favorable Effects	Shows a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other relevant outcome	Shows a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other relevant outcome	Shows a statistically significant and positive (i.e., favorable) effect of the intervention on a student outcome or other relevant outcome	Relevant research or an evaluation that suggests that the intervention is likely to improve a student outcome or other relevant outcome
Other Effects	Is not overridden by statistically significant and negative (i.e., unfavorable) evidence from other findings in studies that meet WWC Evidence Standards with or without reservations (or are the equivalent quality)	Is not overridden by statistically significant and negative (i.e., unfavorable) evidence from other findings in studies that meet WWC Evidence Standards with or without reservations (or are the equivalent quality)	Is not overridden by statistically significant and negative (i.e., unfavorable) evidence from other findings in studies that meet WWC Evidence Standards with or without reservations (or are the equivalent quality)	An effort to study the effects of the intervention, ideally producing promising evidence or higher, will happen as part of the intervention or is underway elsewhere
Sample Size and Overlap	Includes a large sample and a multi-site sample, overlapping with populations <u>and</u> settings proposed to receive the intervention	Includes a large sample and a multi-site sample, overlapping with populations <u>or</u> settings proposed to receive the intervention	N/A	N/A