

Understanding the Evidence Definitions Used for U. S. Department of Education Programs

Slide 1

This presentation is on understanding the evidence definitions used for U. S. Department of Education programs.

My name is Jonathan Jacobson, and I work for the National Center for Education Evaluation and Regional Assistance in the Institute of Education Sciences. IES is the research, statistics, and evaluation arm of the Department of Education.

Slide 2

The goals for this presentation are to help you to:

- Understand the meaning of the terms, “project component”, “logic model”, and “relevant outcome”
- Understand what the term “evidence-based” means in the context of federal education law and regulations
- Understanding the distinction, in the Department’s evidence definitions, between “strong evidence”, “moderate evidence”, “promising evidence”, and “evidence that demonstrates a rationale”
- Understand how the What Works Clearinghouse, an initiative of IES, reviews studies to assess the quality of evidence
- Be aware of evidence-related resources available online at the Department of Education’s website

Slide 3

The Department defines three terms that are foundational for the other evidence definitions in the Department’s general administrative regulations (known by the acronym “EDGAR”). These three terms are “project component”, “logic model”, and “relevant outcome.”

A “project component” means an activity, strategy, intervention, process, product, practice, or policy included in a project.

A “logic model” (also referred to as a *theory of action*) means a framework that identifies key project components of the proposed project (that is, the active “ingredients” that are hypothesized to be critical to achieving the relevant outcomes) and describes the theoretical and operational relationships among the key project components and relevant outcomes.

A “relevant outcome” means the student outcome or other outcome the key project component is designed to improve, consistent with the specific goals of the program.

Slide 4

This slide, based on a presentation prepared by the Regional Educational Laboratory (REL) Pacific, shows the four major elements of a project's logic model.

First there are Resources, which are the materials to implement the project, such as facilities, staff, stakeholder support, funding, and time.

Second there are Activities, which are the steps for project implementation, including the key project components that are critical for the project's success.

Third there are Outputs, which are the immediate products of the project, such as the levels of enrollment and attendance in a course of instruction.

Fourth there are Impacts on Outcomes, which are changes in project participants' knowledge, beliefs, or behavior. If influencing a student outcome or other relevant outcome is a goal for a project, then that outcome is a relevant outcome for the project.

Evidence to inform the design of a project, including the selection of project components, comes from studies that relate a project component--or a combination of project components, up to and including the project as a whole--to at least one relevant outcome.

Slide 5

Both the Elementary and Secondary Education Act, as amended by the Every Student Succeeds Act in December 2015, and the Department's General Administrative Regulations, as updated in July 2017, define "evidence-based" activities, strategies, or interventions.

In the context of federal education law and regulations, "evidence-based" means the proposed project component is supported by one or more of

- *strong evidence*,
- *moderate evidence*,
- *promising evidence*, or
- evidence that *demonstrates a rationale.*

"Demonstrates a rationale" means a key project component included in the project's *logic model* is informed by research or evaluation findings that suggest the *project component* is likely to improve *relevant outcomes*.

Slide 6

Both federal education law and ED regulations define 4 tiers of evidence. These tiers are distinguished by the sorts of studies providing evidence that a project component has a positive (that is, favorable) effect on a student outcome or other relevant outcome.

--Strong evidence, the highest tier, needs to be based on at least 1 well-designed, well-implemented *experimental study* demonstrating a statistically significant and positive effect of a project component on a relevant outcome. ED regulations require this study to meet What

Works Clearinghouse evidence standards without reservations, when assessed using Version 2.1 or Version 3.0 of the WWC Handbook.

--Moderate evidence needs to be based on at least 1 well-designed, well-implemented *quasi-experimental design study* demonstrating a statistically significant and positive effect of a project component on a relevant outcome. ED regulations require this study to meet What Works Clearinghouse evidence standards with or without reservations, when assessed using Version 2.1 or Version 3.0 of the WWC Handbook.

--Promising evidence needs to be based on at least 1 well-designed, well-implemented *correlational study with statistical controls for selection bias* demonstrating a statistically significant and positive effect of a project component on a relevant outcome. It is not necessary for this study to meet What Works Clearinghouse evidence standards or be reviewed by the WWC.

--The lowest tier, evidence that demonstrates a rationale, does not need to be based on research with a statistically significant finding or that has been reviewed by the WWC, but should indicate that the project component is likely to improve a relevant outcome.

Slide 7

Studies that can provide *strong evidence* for the effectiveness of a project component must have certain characteristics to meet the evidence definitions in federal law and U. S. Department of Education regulations.

First, the study must demonstrate a statistically significant and positive (that is, favorable) effect of the project component on a relevant outcome. A statistically significant and positive effect is an estimate of a favorable effect of the project component [that is, the intervention or treatment condition] on a relevant outcome for which the probability of observing an effect that is at least as large as the measured effect, under the hypothesis that the intervention had no true impact, is less than one in 20 (that is, has a p-value under 0.05 using a two-tailed t-test).

Second, to provide strong evidence, the study must be a well-designed and well-implemented *experimental* study. “Experimental study” means a study that is designed to compare outcomes between two groups of individuals (such as students) that are otherwise equivalent except for their assignment to either a treatment group receiving the project component or a control group that does not.

Slide 8

The Department’s evidence definitions recognize several types of research designs as experimental studies, each of which is eligible to receive the What Works Clearinghouse’s highest evidence rating, “Meets What Works Clearinghouse Standards without Reservations.” A randomized controlled trial (or RCT) employs random assignment of, for example, students, teachers, classrooms, or schools either to receive the project component being evaluated [that is, to serve as the treatment group] to not receive the project component [that is, to serve as the control group].

A regression discontinuity design (or RDD) study assigns the project component being evaluated using a measured variable (for example, assigning students reading below a cutoff score to tutoring or developmental education classes) and controls for that variable in the analysis of outcomes.

A single-case design (or SCD) study uses observations of a single case (for example, a student eligible for a behavioral intervention) over time in the absence and presence of a controlled treatment manipulation to determine whether the outcome is systematically related to the treatment [that is, to the project component].

Note that the What Works Clearinghouse has separate standards for reviewing RCTs, RDDs, and SCDs, which standards are described in detail in the WWC Procedures and Standards Handbooks. The Department's regulations require studies providing strong evidence to meet WWC standards without reservations, and to meet certain other requirements described in a separate presentation on "Using the What Works Clearinghouse to Identify Strong or Moderate Evidence of Positive Effects from Education Interventions."

Slide 9

Studies that can provide **moderate evidence** for the effectiveness of a project component (that is, the intervention or treatment condition) must have certain characteristics to meet the evidence definitions in federal law and U. S. Department of Education regulations.

First, the study must demonstrate a statistically significant and positive (that is, favorable) effect of the project component on a relevant outcome.

Second, to provide moderate evidence, the study must be **either** a well-designed and well-implemented **quasi-experimental design** (or QED) study, or an **experimental study** that is considered as good as or better for making causal inferences than a well-designed, well implemented QED. A QED study is a study using a design that attempts to approximate an experimental study by identifying a comparison group that is similar to the treatment group in important respects. Depending on its design and implementation, a QED study can meet What Works Clearinghouse Group Design Standards with Reservations. A randomized controlled trial receiving the same rating from the WWC would be an example of an experimental study that could also provide moderate evidence. The Department's regulations require studies providing moderate evidence to meet WWC standards with or without reservations, and to meet certain other requirements described in a separate presentation on "Using the What Works Clearinghouse to Identify Strong or Moderate Evidence of Positive Effects from Education Interventions."

Slide 10

Studies that can provide **promising evidence** for the effectiveness of a project component (that is, the intervention or treatment condition) must have certain characteristics to meet the evidence definitions in federal law and U. S. Department of Education regulations.

First, the study must demonstrate a statistically significant and positive (that is, favorable) effect of the project component on a relevant outcome.

Second, to provide promising evidence, the study must be *either* a well-designed and well-implemented *correlational study with statistical controls for selection bias*, or a quasi-experimental design study or experimental study that is considered as good as or better for causal inferences than a well-designed, well-implemented correlational study with statistical controls for selection bias. An example of such a correlational study would be a study using regression methods to account for differences between a treatment group and a comparison group. A quasi-experimental design study using statistical matching, or an experimental study with random assignment but high rates of sample attrition, would be examples of studies that could also provide promising evidence, even if they were not reviewed by the What Works Clearinghouse or were reviewed but did not meet WWC evidence standards.

Slide 11

Consistent with federal law and U. S. Department of Education regulations, studies that demonstrate a rationale need to include high-quality research or evaluation findings indicating that a project component (that is, the intervention or treatment condition) is likely to improve a student outcome or other relevant outcome.

Evidence that demonstrates a rationale can include favorable findings from an experimental study, a quasi-experimental design study, a correlational study with statistical controls for selection bias, or some other high-quality research study or evaluation.

The findings in question need to be *positive* (that is, favorable) but do *not* need to be statistically significant. These findings do not need to be reviewed by the What Works Clearinghouse or meet WWC evidence standards.

Slide 12

The role of the What Works Clearinghouse for the Department of Education and for other stakeholders is to identify well-designed and well-implemented experimental and quasi-experimental design studies that can inform education decisions.

Established in 2002, the WWC is an initiative of the Department's Institute of Education Sciences.

The WWC reviews, rates, and summarizes *original* studies of the *effects* of education interventions, which include education policies, programs, practices, or products—that is, project components as defined in the theory of action for an education project.

The WWC does NOT rate qualitative studies, descriptive studies, or studies that re-analyze or synthesize others' data. The WWC focuses on that subset of all education research that consists of original studies of the effects of education interventions.

In addition, the WWC does not rate the overall quality of interventions or endorse interventions. Rather, the WWC rates the quality of *studies* of interventions, which studies can inform decision makers' choices of whether to adopt certain policies, programs, practices, or products. The WWC's reviews of studies are documented on the WWC website, whatworks.ed.gov. The WWC only reports findings from reviewed studies that meet WWC standards.

Slide 13

What Works Clearinghouse standards have been developed by panels of experts for different types of impact study designs, including experimental designs such as RCTs, RDDs, and SCDs, as well as quasi-experimental design studies.

U. S. Department of Education programs using “strong evidence” or “moderate evidence” as defined in ED regulations rely on *previous* WWC reviews completed using Version 2.1 or Version 3.0 of the WWC Procedures and Standards Handbook. New reviews of studies during Fiscal Year 2018 are being conducted using Version 3.0 of the WWC Handbook. Version 4.0 Handbooks were released by the WWC in October 2017 but are not being used for the Department's evidence determinations in Fiscal Year 2018.

WWC standards focus on the *internal validity* of impact estimates, that is, whether the estimate is likely to be unbiased.

These standards are applied by teams of certified reviewers using a study review protocol to give eligible studies one of 3 possible ratings: Meets WWC Design Standards without Reservations (the highest possible rating), Meets WWC Design Standards with Reservations, or Does Not Meet WWC Design Standards.

Slide 14

Study review protocols define studies eligible for WWC review according to the eligibility of

- the intervention (that is, the policy, program, practice, or product being studied)
- the population or populations targeted for the intervention
- the research defined by topic, time frame, language, and location
- the outcomes included in the impact analysis

Besides being eligible under the applicable review protocol, the outcomes on which a study reports findings need to meet the standards in the WWC Handbook, including validity, reliability, not being over-aligned with the intervention, and being measured in the same way for the comparison group as for the intervention group.

Primary findings are prioritized for WWC review over secondary findings and are used by the WWC to characterize the effectiveness of interventions as positive, potentially positive, indeterminate, not discernible, mixed, potentially negative, or negative. Primary findings focus on confirmatory research questions instead of exploratory research questions, rely on the full study sample instead of subgroups, and make use of composite measures rather than subtests or

subscales. The time period when primary findings are measured is defined in the applicable WWC study review protocol and depends on the characteristics of the intervention and student population receiving it. The review protocol also defines which secondary findings may be included in a corresponding WWC review.

Protocols define studies eligible for WWC review by topic area, for example studies of Beginning Reading; English Language Learners; Secondary Mathematics; Supporting Postsecondary Success; or Teacher Training, Evaluation, and Compensation. If a study does not fit in any topic area reviewed by the WWC, the WWC uses the Review of Individual Studies

Protocol to assess the study's eligibility for WWC review and to guide that review, if the study is eligible. If studies are expected to be reviewed by the WWC for a grant competition sponsored by the Department, the applicable study review protocol will be determined by the Department prior to the publication of the Notice Inviting Applications and can be confirmed during the pre-application period.

Study review protocols can be viewed by the public on the same webpage as the WWC procedures and standards Handbooks.

Slide 15

Here is a detailed presentation of the outcomes eligible for WWC review under two review protocols relevant for the Department's Fiscal Year 2018 grant competitions. Please consult the corresponding Notice Inviting Applications for official information on each competition. The first protocol, the Review of Individual Studies Protocol, version 3.0 from April 2016, is being used for reviews of studies proposed as strong or moderate evidence for the Education Innovation and Research (EIR) program grant competitions. Eligible outcomes under this protocol include:

- standardized tests and assessments of student academic readiness, academic knowledge, or academic skills
- Student grade point averages from final grades or credits earned at the secondary (Grade 6-12) or postsecondary level
- Student educational enrollment, educational attendance, or educational attainment
- Labor market outcomes of students, including employment and earnings
- Behavioral ratings of students and student behavioral and social outcomes
- Teacher outcomes as defined in the Teacher Training, Evaluation, and Compensation evidence review protocol

The Teacher Training, Evaluation, and Compensation evidence review protocol, version 3.2 from July 2016, is being used for reviews of studies proposed as moderate evidence for the Supporting Effective Educator Development (SEED) program grant competition. Eligible outcomes under this protocol include

- Student achievement in English Language Arts, mathematics, science, social studies, or general academic achievement
- Student promotion or graduation
- Quality of teacher instruction

- Teacher attendance
- Teacher retention
- Measures of teacher or school effectiveness, such as value-added measures

More information, including the full text of each study review protocol, is provided on the WWC website at the links provided on this slide.

Slide 16

This flowchart shows the process by which an eligible randomized controlled trial or eligible quasi-experimental design study is assessed to see if meets the WWC's group design standards. If the study determines membership in the intervention group through a random process, and the rates of overall and differential sample attrition are low, then the study can receive the highest rating: Meets WWC Group Design Standards without Reservations.

If the study is an RCT with high rates of overall or differential sample attrition, or is a QED study, then it needs to establish equivalence at baseline of the intervention group and the comparison group for the sample used in the analysis of impacts. If that equivalence is established for key baseline characteristics specified in the study review protocol, then the study can Meet WWC Group Design Standards with Reservations. If the study does not demonstrate baseline equivalence, then it receives the rating, Does Not Meet WWC Group Design Standards. Note that the sign, size, and statistical significance of the estimated effect of the intervention are reported by the WWC, if the study meets WWC standards, but this information does not affect the WWC study rating. Consequently, knowing a study's rating by the WWC (that is, whether a study meets WWC design standards with or without reservations) is NOT sufficient to establish whether a study provides strong evidence, moderate evidence, or even promising evidence as defined by the Department. To provide strong or moderate evidence as defined by the Department for its programs, individual studies need to satisfy additional requirements besides meeting WWC standards. We describe these requirements in a separate presentation, "Using the What Works Clearinghouse to Identify Strong or Moderate Evidence of Positive Effects from Education Interventions."

Slide 17

Before concluding this presentation, I want to make sure you are aware of several pages on the U. S. Department of Education website that relate to evidence.

First, to help you understand how various terms related to evidence are defined for the Department's programs, you can consult the Education Department General Administrative Regulations (or EDGAR), specifically 34 CFR Part 77.

Second, to find evidence from experimental and quasi-experimental design studies reviewed by the What Works Clearinghouse, you can consult the WWC website, whatworks.ed.gov. As I mentioned previously, a separate presentation provides details on how to use the WWC website to find Strong or Moderate Evidence of positive effects of education interventions.

Third, if you are interested in using evidence to design education projects in the context of programs supported under the Elementary and Secondary Education Act, you can examine the document, “Non-Regulatory Guidance on Using Evidence to Strengthen Education Investments.”

Finally, if you are interesting in building evidence through the design of a project evaluation to meet What Works Clearinghouse standards, you can examine the “Technical Assistance Materials for Conducting Rigorous Impact Evaluations”, available on the website of the National Center for Education Evaluation and Regional Assistance at the Institute of Education Sciences.

Slide 18

Thank you for your time and interest in this topic.

We welcome your comments and questions on this presentation, which you can send to me at jonathan.jacobson@ed.gov.