ESSA and Evidence Claims:

A Practical Guide to Understanding What “Evidence-Based” Really Means

+39%

Average Growth

≥.25

n = 200K

i-Ready
Introduction

The implementation of the Every Student Succeeds Act (ESSA) represents America’s greatest shift in federal education law since the passage of No Child Left Behind (NCLB) in 2001. While there are a number of key differences between NCLB and ESSA, there is one key area in which the law has evolved considerably: focused use of federal education funds on programs with proven results. ESSA raises the bar of what qualifies as an “evidence-based” educational activity, and the new law better reflects the state of the art for efficacy research in general.

ESSA’s new approach gives districts the flexibility to adopt and evaluate new approaches, but also strongly encourages the adoption of activities with established evidence bases and proven results. Most notably, ESSA requires all Title I schools to use “evidence-based” interventions and strategies. Knowing and understanding not only the technicalities of what the law deems “evidence-based,” but also what common-sense questions to ask about that evidence-based designation, will help districts find the products and programs that best fit their needs with their Title I dollars.
How does federal law define “evidence-based”?  
ESSA defines four categories, or levels. These categories are defined solely by the type of study conducted, not by the strength of the study results.

<table>
<thead>
<tr>
<th>Evidence Tier</th>
<th>Type of Study</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Strong Evidence</td>
<td>A well-designed and implemented experimental or randomized control trial (RCT) Requires special effort and data collection. Presents unique logistical and technical challenges.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Moderate Evidence</td>
<td>At least one quasi-experimental study Can be performed using available data or data specifically collected for the study; common in educational research.</td>
</tr>
<tr>
<td>Level 3</td>
<td>Promising Evidence</td>
<td>At least one correlational study with statistical controls for selection bias Are usually performed using available data and are common in educational research.</td>
</tr>
<tr>
<td>Level 4</td>
<td>Demonstrates a Rationale</td>
<td>Relevant research or evaluation showing that the product will likely improve student outcomes; still needs other support that it has a favorable effect Gives districts the ability to adopt new, promising approaches not yet supported by high-quality efficacy research.</td>
</tr>
</tbody>
</table>

The evidence tiers propose a hierarchy of “evidence-based” with randomized control trial (RCT) research being the strongest at Level 1 and lacking formal research evidence being the weakest at Level 4. In general, it is the case that a higher level indicates a study may be more conclusive and will support stronger inferences. However, these evidence levels fail to clarify very important concerns: namely, how strong are the results, how credible are the results, and how relevant are they to a particular district’s goals and challenges?

We propose asking some illuminating questions to reveal the substance of the evidence in the next section.
What level is best?
ESSA’s evidence levels are determined by the type of study conducted and whether the results were statistically significant, but the type of study alone should not be the deciding factor in determining whether to select a product or activity. In many cases, other criteria may matter a lot more. A Level 1 study’s evidence is not necessarily of higher quality than a product backed by a strong Level 3 study; however, factors such as the content standards used in the study, timing, and having diverse students in the study sample could hold greater weight than the type of study.

In addition, a Level 1 study is hard to achieve, not because it means the results are stellar, but because successfully performing a randomized control study in education is difficult, if not unfeasible. In a randomized education study, some students do not receive the intervention so they may act as a control for the sake of the study. Level 1 studies often occur in a university setting or under the direction of a research institution. Additional questions are required to understand if a Level 1 study truly represents a superior product or if the product merely met the singular study criteria. A Level 2 or Level 3 study is usually faster and more practical for a study of broadly-used educational activities. Therefore, if researchers chose a different type of study to meet the evidence-based guidelines, it is often because another study type is better suited for meeting the more practical criteria that concern most districts.

Five Questions You Should Be Asking about Evidence-Based Interventions in ESSA
To assist schools and districts with the task of evaluating evidence claims for ESSA, we’ve created five questions districts should ask to cut through the evidence-based jargon.

When was the study conducted?
ESSA does not require that the research evidence be current or up to date. However, we know that a lot can change in education in a few short years. Educational standards change, instructional methods change, students’ preparation levels change, and even how we assess students changes. While most programs will highlight their most current, relevant evidence, it is worth learning more about when a study was conducted and how it related to your proposed uses. The rapid pace of change in education can render even the highest quality study obsolete. Make sure to understand when the study or studies were conducted and think about how the evidence relates to your current situation.
How large was the sample size for the study and did the sample include a diverse set of students including subgroups?

At the higher levels, for example, 1 and 2, ESSA guidance is that studies include a “large” sample size (typically meaning more than 350 students) and that the sample come from multiple sites. However, since evidence levels are self-reported, it’s important to ask how many students and which students were included in the study. Well-implemented studies should also include students of diverse background; gender, ethnicity, and socio-economic background among others. The subgroup designations used in ESSA reporting may also be important. Quality efficacy studies mirror diverse district demographics.

Studies can have general applicability, but relevance to different student populations and types of school gives districts more confidence that the product will work for all students, including English learners, diverse genders, multiple races, and various socio-economic backgrounds. To ensure the product shows gains among diverse students, districts will need to ask clarifying questions about the study methodology.

Ask specifically if the study mirrors the demographics and diversity of students served in your school district.

Was the study based on current content and standards?

The guidelines for ESSA research implicitly assume that studies are aligned with a state’s current standards, but it is up to districts to find out for sure. It is possible that a Level 1 study used older standards or even standards from another state. To ensure good standards alignment, ask about what content and standards alignment were used in the program or intervention being studied. Also, note the program’s purpose. Was the product designed as more of skills-based intervention or tied to standards-based curriculum? Knowing the purpose of the product will help to compare whether the product will fit the solution the district is looking for.

Were the results favorable?

ESSA defines favorable as “statistically significant and positive.” Level 1–3 studies should have statistically significant, positive results. However, many studies contain a mix of positive, negative, and indeterminate findings. It’s important to understand what each finding means and how it might affect your district. Evaluating effect sizes can provide a guide to the how strong the results were. Effect sizes are a common way of measuring the strength of an educational intervention. Larger effect sizes indicate a greater effect.
Because the outcomes are more challenging to influence with interventions, the average effect sizes in research fields such as education, medicine, and economics are smaller than in other fields of research. Specifically, effect sizes of .25 or greater are considered “large” in education research.\(^2\)

**Does the product truly help students?**

While the statistical results are an important part of a well-designed study’s specifications, the real question is: **does the product produce positive outcomes for students?** What evidence besides the studies required by ESSA shows positive, practical outcomes? Ask about other groups that have used or reviewed the product for content alignment, efficacy, and usability. Federal guidance agrees that “stakeholders should consider the entire body of relevant evidence.” Getting additional perspectives will give districts the best information possible to make the most effective use of federal funds.

In conclusion, there is no formal mechanism for approval of an intervention under the ESSA regulations. While states sometimes suggest districts rely on online clearinghouses, it’s really up to educators to ask the right questions to get the best interventions, curriculum, and activities with their Title I dollars. By asking these five questions, districts and schools will get a better picture of whether a product truly is the best fit for them.

---

Learn more about the research on *i-Ready*

Research based on data from over four million students who took the *i-Ready Diagnostic* during the 2016–2017 school year found that students who used *i-Ready Instruction* experienced greater learning gains than those who did not use the program. These learning gains were not only substantial, but were also significant across key student subgroups, including non-Caucasian students, students with disabilities, economically disadvantaged students, and English learners.

The strength of the evidence gathered means that *i-Ready Instruction* meets the criteria for ESSA Level 3: Promising Evidence with Favorable Effects.

CurriculumAssociates.com/EssaResearch-wp