

# 6-Point Efficacy Research Checklist

## How to determine the true effectiveness of your educational programs

Adopting a new educational program, but unsure if it's the right one? A critical factor to consider is the **proven effectiveness of the program**.

We want to help ensure that you have a good understanding of the quality of research behind a given program's reported effectiveness.

This checklist will help you determine:

- If an efficacy study meets standards presented by governing bodies such as Evidence for ESSA and/or the What Works Clearinghouse (WWC)
- If the quality of the research you are reviewing meets a good level of scientific rigor

- The study's research design ensures that the evidence enables you to effectively address the research problem as clearly as possible**

**ESSA Tier I & WWC (without reservations): Randomized controlled trial**

The study has an experimental and a control group, and students (and teachers) were randomly assigned into the groups

**ESSA Tier II & WWC (with reservations): Quasi-experimental design**

The study has an experimental and a control group, but students (and teachers) were not randomly assigned into the groups

**ESSA Tier III: Observational study**

There is only one group in the study, and all students in the group used the program

- The study uses state assessments or assessments developed by a recognized organization (e.g., NWEA) as a measure of student achievement**

Studies that use a developer's own benchmark tests as measures of student achievement are not eligible for approval by Evidence for ESSA or the WWC because such benchmark tests lack objectivity.

- The study sample's characteristics are similar to your school or district's student population

If the characteristics of participants are vastly different from your student population, the results found in the study may not be observed in your school/district. Characteristics include, but are not limited to, school grades studied, percentage of students in each ethnic group, socioeconomic background, and percentage of English learners.

- A reputable research organization conducted the study and handled the data

If the developers themselves conducted the study or handled the data, it may be biased and, ultimately, invalidate its findings. For example, students with low program usage may be arbitrarily removed from the study in order to statistically increase program effectiveness.

Examples of reputable research organizations are Harvard University's Center for Education Policy Research and Johns Hopkins University's Center for Research and Reform in Education.

- The statistical analysis is appropriate for efficacy research

In its statistical analysis, the study must control for pretest achievement and other relevant student characteristics (e.g., eligibility for Free/Reduced Lunch, English learner status, etc.) in order for its findings to be valid.

To achieve this, analysis of covariance (ANCOVA), multiple regression, and hierarchical linear modelling (HML) are commonly used analytic methods. Results from correlations and t-tests are not sufficient to demonstrate program effectiveness due to their lack of statistical control.

modelling (HML) are commonly used analytic methods. Results from correlations and t-tests are not sufficient to demonstrate program effectiveness due to their lack of statistical control.

- The study results are statistically significant

Statistically significant results commonly have a probability value of less than 5% (i.e.,  $p < 0.05$ ) or, in other words, a 95% confidence level. For example, if a study is conducted 100 times, 95 of them will produce winning results. The lower the  $p$ -value, the less probability there is for false-positive results.

However, *p*-value — no matter how small it is — is not an indication of how much impact the program has on student learning. To understand the magnitude of the program impact, you need to look at the effect size. Evidence for ESSA has set the standard for effect size at 0.2 or above. If an effect size is too small (e.g., less than 0.1) or too large (e.g., greater than 0.5), you need to interpret the result with caution. Also, if claims are made that X amount of program usage would boost student achievement by Y amount, such claims need to be substantiated by rationales and evidence from efficacy research.