



## How to Gather Rigorous Evidence of Your Program's Effectiveness

Ideally every tutoring initiative leads to positive student outcomes. The evidence is very strong that tutoring can benefit students. However, some programs are likely to be more effective than others. How can we be sure that an investment in tutoring is paying off? Well-designed evaluations can provide definitive evidence about how much a tutoring initiative is driving student learning and other valued outcomes.

Whether you are a tutoring provider looking for evidence of your program's effectiveness or a school or district leader interested in understanding the benefits for your students, conducting a high-quality study, particularly a randomized controlled trial (RCT), might be the right choice for evaluating your tutoring program.

### What is a randomized controlled trial?

RCTs involve randomly selecting students to receive tutoring from among a preselected group of students with need for tutoring. Random assignment is like using a lottery or a coin flip to determine which students receive tutoring. Each student who is eligible to receive tutoring is assigned to either the Tutoring Group or to a Comparison Group that does not receive tutoring.

### Why does random assignment matter?

When random assignment is done correctly and the sample is large enough, the Tutoring Group and Comparison Group will be similar on both measured traits (like gender or economic status) and unmeasured characteristics (like motivation or parent engagement). As a result, if you see differences in outcomes, such as assessment scores, between those who receive and do not receive tutoring, you can be sure that the differences are driven by the tutoring program. If you didn't randomize students to the Tutoring Group and Comparison Group, they might look similar on measured outcomes such as prior test scores but might not be similar on unmeasured characteristics, such as interest in the subject or engagement in school. With an RCT, when you follow the groups over time, you can confidently attribute the difference in student outcomes to the tutoring program, rather than to other factors.

### Why should you consider conducting an RCT?

#### 1) You want data-driven evidence for the effectiveness of a program or an intervention.

RCTs are widely accepted as the "gold standard" of evaluation designs.<sup>1</sup> They allow you to say whether a program causes changes in outcomes. RCTs give robust and credible estimates of a program's effects because it makes it possible to determine what would have happened if students did not receive tutoring.

<sup>1</sup> In order for an intervention to meet the Every Child Succeeds Act (ESSA) standards for Tier 1 (Strong Evidence), interventions must be evaluated using at least one randomized controlled trial (RCT).

## 2) You want to increase program demand and buy-in.

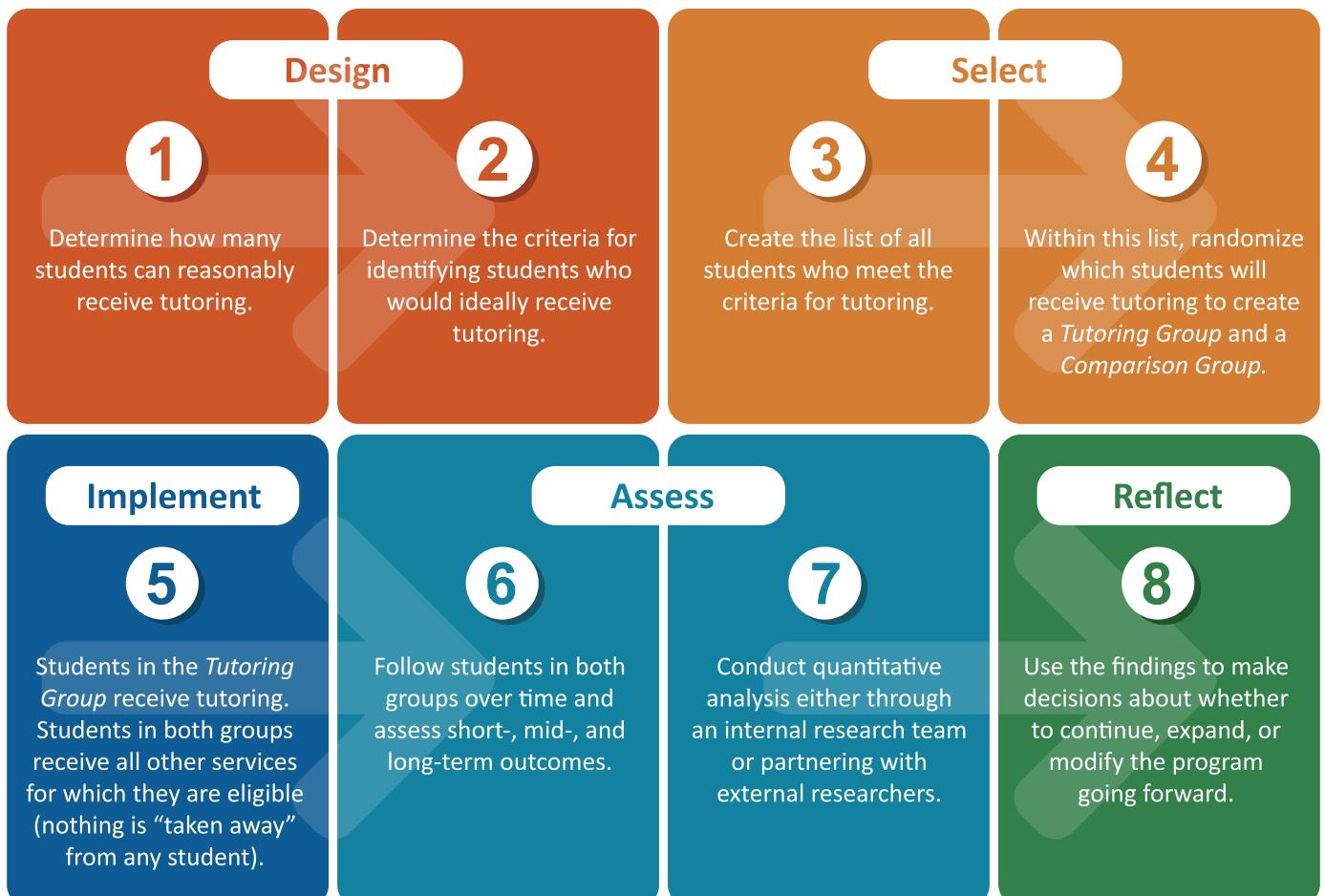
Many districts, schools, and funders increasingly require evidence of a program’s effectiveness as a prerequisite for dedicating resources to it. Convincing evidence that a tutoring program leads to student learning can both increase demand for a program and promote buy-in among key stakeholders. Programs that have been evaluated by RCTs (or include an RCT as a component of their proposal) are also more likely to receive external funding.

## 3) You want to learn how to make your tutoring program more effective.

RCTs can be used for more than program evaluation — they can also be used to explore what characteristics of tutoring programs makes them more or less effective. Randomly assigning which students receive certain versions of a tutoring program may illuminate the most effective (and cost effective) delivery models. Random assignment of eligible students to the Tutoring or Comparison Groups or to different forms of Tutoring not only helps provide good information on tutoring effectiveness, it is also fair to students. By randomly selecting which students receive tutoring based on predetermined criteria (e.g., students who are more than one grade-level below standards or those receiving special services), you are taking a systematic approach to selecting students for tutoring as opposed to other selection methods (e.g., time of day available, teacher selection, or students selecting in) that might unintentionally disadvantage some groups of students.

### How to Conduct a Randomized Controlled Trial of Your Tutoring Program

Here are the steps needed to conduct an RCT of a tutoring program.



## Conducting a Randomized Controlled Trial of Your Tutoring Program: A District Example

Here is a practical example of the steps a district might take to conduct an RCT of a tutoring program.

### 1

**Determine how many students can reasonably receive tutoring.**

A School District has the resources to tutor 500 students in literacy across 10 elementary schools.

### 2

**Determine the criteria for identifying students who would ideally receive tutoring.**

The School District decides to focus on 1st - 3rd grade students who:

- Are struggling with literacy (i.e., at least one grade level behind)
- Are in their five Title I schools
- Can be scheduled during the block the school has designated for tutoring

### 3

**Create the list of all students who meet the criteria for tutoring.**

1,500 students in the School District meet these criteria across the five schools:



### 4

**Within this list, randomize which students will receive tutoring to create a *Tutoring Group* and a *Comparison Group*.**

The District randomizes 500 students to the *Tutoring Group* and 1000 students to the *Comparison Group*. They choose to do the randomization separately for each school so that each school provides tutoring to one third of eligible students:



### 5

**Students in the *Tutoring Group* receive tutoring. Students in both groups receive all other services for which they are eligible (nothing is “taken away” from any student).**

All students in the *Tutoring Group* and in the *Comparison Group* continue to have the same educational experiences and receive the same services as they would have except that the *Tutoring Group* also receives tutoring for the duration of the tutoring program.

### 6

**Follow students in both groups over time and assess short-, mid-, and long-term outcomes.**

The School District assesses the outcomes of students in the *Tutoring Group* and in the *Comparison Group*. These outcomes may include, but are not limited to:

- Student achievement (e.g., grades, test scores, course completion)
- Attendance
- Discipline
- Dropout
- Social-emotional skills (e.g., self-regulation, relationship skills)
- Belonging
- Motivation (e.g., valuing of subject, attitudes towards school)

### 7

**Conduct quantitative analysis either through an internal research team or partnering with external researchers.**

The School District has a small internal research team, so they partner with external researchers for support conducting the statistical analyses. Together, the researchers create a final report that details the key results.

### 8

**Use the findings to make decisions about whether to continue, expand, or modify the program going forward.**

Now, key stakeholders in the School District can review the findings and think about how to improve the program.

## Frequently Asked Questions

### I heard tutoring was an evidence-based practice – why do I need to evaluate my program?

While research shows that high-impact tutoring is one of the most cost-effective ways to accelerate student learning, not all tutoring programs are effective. As every educator knows, no context is the same, and providing students with tutoring is a complex, resource-intensive undertaking. Some programs work better than others; it is useful to know how your particular tutoring program is benefiting students.

### Can teachers play a role in selecting students?

Yes. If you want to incorporate teacher recommendations, give each teacher a list of students who meet the predetermined criteria and have them recommend a certain number of students for tutoring (ideally double the number of students who would actually be able to receive tutoring). Then, randomly select students from that list to receive tutoring.

### What if we want some students to definitely receive tutoring no matter what? Can we still evaluate the program?

Yes. You still have a few options:

**1. Create three “groups” of students based on need.** For example, Group 1 might include the lowest performing students and they automatically receive tutoring. Group 2 are students who are not the lowest performing, but still would greatly benefit from tutoring. Group 3 students are performing at grade level or above, and are not a priority for tutoring. If you have enough available tutoring slots remaining after assigning all Group 1 students to the tutoring group to make an RCT viable, you can randomly assign Group 2 students to the Tutoring Group or the Comparison Group and compare outcomes for these Group 2 students.

**2. Stagger the rollout of a tutoring program.** In this scenario, the process is much like a standard RCT. Students are randomly assigned to receive tutoring during the first term (or school year) or during the second term (or school year). The comparison at the end of the first period between the two groups will show how effective tutoring is, at least in the short-run. One limitation of this approach is that you can only assess short-term outcomes (before the Comparison Group starts tutoring), hindering your ability to assess the long-term impact of tutoring.

**3. Use a cut score to identify students.** To use a cut score, students who score below a predefined threshold on a test receive tutoring. The effect of tutoring is estimated from a comparison of the Tutoring Group that initially scored just below the cutoff and a Comparison Group that initially scores just above the cutoff. The logic behind using a cut score is that the students who score just below and just above the cut score are very similar to one another, so whether they receive tutoring or not is very close to being randomly assigned. However, for the cut score approach to work, you need a very definitive cut-off point with many students scoring just above and just below the threshold. Additionally, the specific cut score cannot be used for determining whether the students receive other educational services.

### Can I conduct an RCT if my district or program is relatively small (perhaps serves <50 students)?

Before conducting an RCT, it is important to determine whether your study will have enough students receiving tutoring (statistical power) to identify the effect of your program. Otherwise, you may conduct a study only to find that you can not precisely determine program impact and so gain little helpful information from the research. Small sample sizes can make it difficult to run powerful studies. However, strong data on prior student performance may make it possible to run impactful RCTs with relatively few students. Talk with a researcher about the specifics of your situation.

## What if we want to learn about how to make tutoring more effective?

For this type of evaluation, instead of randomly assigning students to receive tutoring or not, you randomly assign students to receive different types of tutoring. Students can be assigned to Tutoring Group A, Tutoring Group B, and so on (in addition to a Comparison Group, if desired).

### Making tutoring more effective: An example

Say you want to know the optimal number of tutoring sessions for students, which has implications for scheduling and cost-effectiveness. You might choose to randomize students who meet the predetermined criteria to receive a certain number of tutoring sessions per week:

- Tutoring Group A (two sessions per week)
- Tutoring Group B (three sessions per week)
- Tutoring Group C (four sessions per week)

Then, you can compare student outcomes at the end of the tutoring program to determine the benefits of students receiving additional tutoring sessions per week. Imagine students who received three tutoring sessions per week had meaningfully better outcomes than those who received two sessions per week. Then you can evaluate those gains relative to the costs associated with an additional session. Or perhaps you find no difference in the outcomes between students who received three and four sessions per week. At that point, it would be reasonable to move forward with providing three tutoring sessions, because it is less costly and no less effective. For more examples, see the Accelerator's research priorities focused on [Identifying the Characteristics of Effective Tutoring](#).

## Recommended additional reading

[Using Rigorous Designs](#)

[The Advantages of Randomized Evaluations](#)

[The narrow path to do it right: Lessons from vaccine making for high-dosage tutoring](#)

## Research Collaboration Interest

We are looking to partner with school districts and tutoring providers across the country to evaluate the efficacy of high-impact tutoring through rigorous research studies. [If interested, click here for more information.](#)

## Contact

### National Student Support Accelerator

164 Angell St., 2nd floor, Providence, RI 02906

Telephone: 401.863.7990

Email: [info@studentsupportaccelerator.org](mailto:info@studentsupportaccelerator.org)

Web: <https://studentsupportaccelerator.org>