



Early Planning Needed to Know Your Tutoring Program's Effectiveness

What your District Needs to Do NOW to Sustain your Tutoring Program Post-ARPA Funding

Substantial new federal funds, such as those from the American Rescue Plan Act (ARPA), are allowing districts to provide students with services such as tutoring that were not financially feasible in the past.

Are these new programs cost-effective enough to merit allocating other funds to sustain them, such as Title I and Title IV funding, after ARPA funding runs out in 2024?

To answer this question well, districts will need to design their implementation of tutoring from the beginning with evaluation in mind. **The single most important design factor in being able to determine the effectiveness of the tutoring program is having a believable comparison group of students.** This student comparison group should be similar enough to those receiving tutoring that they will give a convincing picture of what the students receiving tutoring would have looked like if they did not have access to tutoring. Without a credible comparison group, no analysis will be able to identify the learning that is specifically attributable to the tutoring program.

Several approaches to implementation can create particularly convincing comparison groups:

- The most rigorous approach is to use a lottery to select which students participate in tutoring from a group of eligible (in-need) students. This approach allows a district to compare students who were randomly selected for tutoring through the lottery to those who were not, giving districts a picture of the benefit the tutoring provided with a high degree of confidence.
 - This approach may reduce any stigma associated with receiving tutoring
 - The approach may be combined with a small group of students exempted from the lottery to definitely receive tutoring.
 - The approach can also be used to determine the *timing* in which they receive tutoring. For example, in a group of 200 students eligible for tutoring, 100 students are randomly selected to receive tutoring in the first semester. The remaining 100 students receive tutoring in the second semester. After the first semester, the first group can be compared to the students who haven't received tutoring yet, isolating the impact of the tutoring.
- If a lottery approach is not feasible, the district can also create a convincing comparison group by choosing which students get tutoring based on a cut score - such as a particular test score. If students

with scores lower than the specific cut score receive tutoring and those with higher scores do not, the district will have a comparison group of students with scores just above the cut score that are very similar to those just below the cut score who do receive tutoring.

- *Having students opt-in to tutoring or having teachers choose which students get tutoring will make it very difficult to find a convincing comparison group* because those students differ in many unobserved ways, such as motivation, from students who do not receive tutoring.

In addition to establishing a comparison group, a district must also:

- **Identify who will conduct the analysis or research.** Larger districts often have internal capacity to assess programs. If district capacity is insufficient, State Education Agencies may have support available.
- **Clarify the specific question(s) to be answered.** In most cases the questions are something like: How have students benefited from tutoring in terms of academic performance, school engagement, and overall well-being? And are the student benefits from tutoring enough to justify the expense?
- **Determine data needed to answer the question.** Based on the questions above, a district would need to data on:
 - which students were assigned to tutoring, which students were assigned to the comparison group, and how did the assignment work?
 - which students participated in tutoring and how much did they participate?
 - what are the valued outcomes (test performance, grades, attendance, other measures of well-being) of students in the tutoring group and in the comparison group?
- **Establish systems to collect data.** Build a technology system to support the collection and dissemination of the needed data. This system should be as automatic as possible, with minimal reliance on educators.

While these steps may seem obvious, they can be tricky to implement and they are critical for making informed decisions about sustained funding for your district's tutoring program for the benefit of students.

Please see more information on how to construct comparison groups in the [Appendix](#).

Appendix:

Constructing A Control Group: Options and Considerations

There are three methods to construct a comparison or control group:

A randomized controlled trial (RCT): One method is to select a group of eligible students (such as those with low assessment scores or grades, or teacher nomination) and then randomize within that group (e.g. flip a coin) to identify those who get tutoring and those in the comparison group. This method of creating a comparison group allows researchers to directly compare outcomes for those who received the program to those who did not. Differences in outcomes are attributed to program participation, not to other differences such as difference in motivation or supports at home. An RCT is considered the most accurate approach to assessing the impact of a tutoring program

Some district administrators may be concerned about running an RCT because of the history of exploitation in research. These concerns are real, however, RCTs also have equity benefits. When students opt into programs or teachers choose which of their students receive programs, the potential for bias is much stronger than when assignment is based on chance. Moreover, RCTs can be used in a way that all eligible students get access to tutoring, just at different times. The comparison group can receive tutoring once the original tutoring group has completed the program and the outcome data has been collected.

- Example: Randomly assigning access to a math tutoring program to half of the students in a school who are behind grade level in mathematics for the first year of the program and the other half to tutoring during the second year of the program.

A quasi-experimental study (i.e., a regression discontinuity): A second approach is to construct a comparison group of students that resembles the treated group, but differs because of small differences in an observable characteristic. A common example of this approach is to select students into tutoring based whether their test score is above or below a specific cut score. The comparison between students who scored just below the cut score, and thus receive tutoring, with those who scored just above the cut score, and thus do not receive tutoring, provides a clean comparison for estimating the effects of the tutoring program. This method is considered valuable, but less accurate for estimating the impact of a program than an RCT.

A correlational study: A common, though often not convincing, approach to estimating the effects of a program is to compare students who received the program to students who appear similar based on measured characteristics such as prior test scores, attendance, or demographic characteristics. The problem with this approach is that the students could be different on other attributes such as motivation. This approach works better if students are assigned to tutoring because of their schedules (e.g. timing of electives) than it does if students opt into tutoring or if teachers choose students. This method is considered far less accurate than RCTs or quasi-experimental studies.