

Ed Tech Rapid Cycle Evaluation Coach

Overview: Creating Groups with Random Assignment

If you select pilot participants randomly (by chance), you will create two groups that are similar on observed and unobserved characteristics. When random assignment is well implemented, you can be confident that any differences in outcomes are due to the educational technology you are testing. Because of this, random assignment is considered the best choice for evaluations of effectiveness and should be used whenever possible.

THE EVALUATION CHALLENGE

You want to test whether a technology is effective, but it is impossible to simultaneously observe what happens when an individual uses a technology and doesn't use that same technology. If you introduce a technology and watch what happens, you might notice improvements, for example, in student test scores. However, you cannot assume that the technology caused improved student outcomes. Many other factors (including regular classroom teaching, other programs, student maturation, and so on) could have contributed to the increases.

To overcome these challenges, it is important to compare a group of technology users to a group of nonusers, on the assumption that the only real difference between them is whether they are using the technology. However, comparing technology users and nonusers brings an additional set of challenges. When we make comparisons without trying to ensure similarities between groups, it is possible that those who use the technology differ in any number of ways from those who do not use the technology. For example, hard workers might be more likely to try a new technology, but they also perform better on tests. This might cause you to confuse the effect of the technology with the effect of working hard (because either could cause the technology users to outperform nonusers). Those differences can make an ineffective technology look effective, or vice versa.

RANDOMIZED PILOT

Solution

The best test of whether your technology works is to randomly select some of the potential users to pilot the technology and others to continue classroom practice as usual. When the groups are assigned by chance, you can be confident that you are comparing apples to apples—that the two groups are the same in every way except the technology use. Then, if you see differences in outcomes (such as student achievement scores) you can be convinced that the new technology is moving the needle.

Ed Tech Rapid Cycle Evaluation Coach

How it works

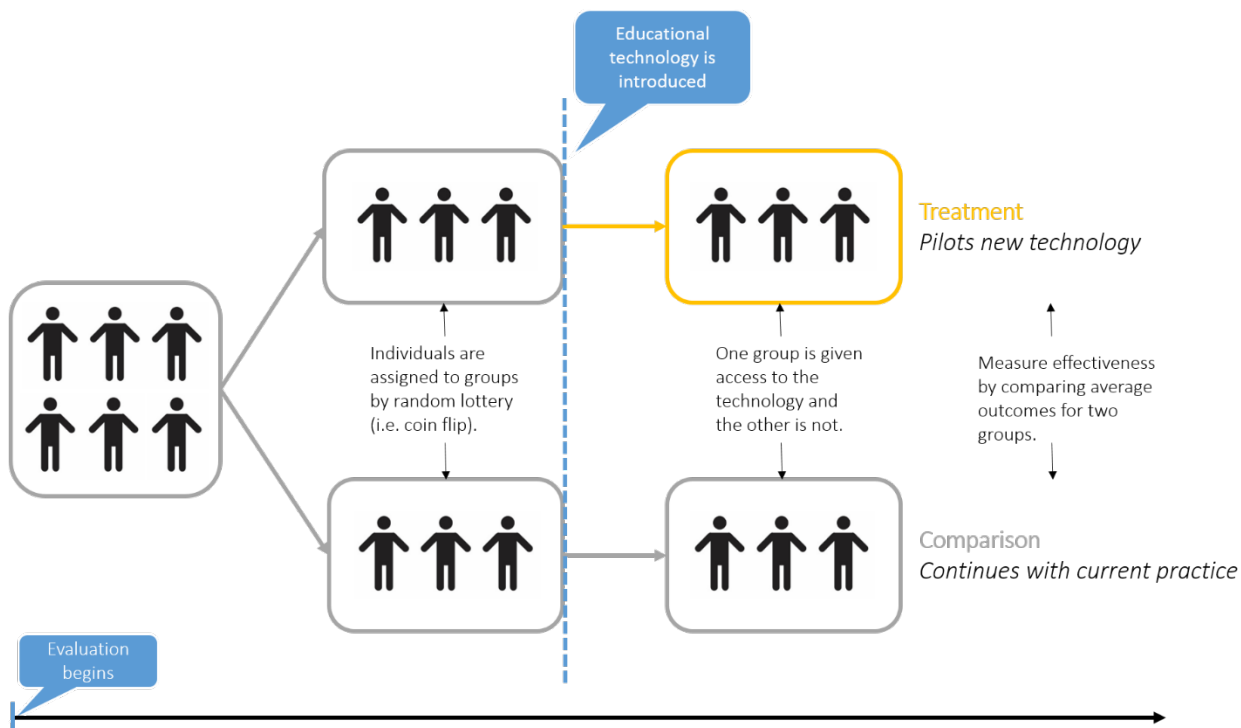
Suppose you want to determine the impact of the reading technology U-Read on 5th-grade reading test scores in your school. To know if U-Read is having an effect on end-of-year reading scores, you would like to be able to observe your users in a parallel universe. In this universe the parallel nonusers are exactly the same as the users, but have no access to U-Read. If this parallel group scored lower on the end-of-year reading test, you could conclude that U-Read moved the needle for users.

With random assignment

You can randomly assign students at the beginning of the year, using a method equivalent to a coin flip, to one of two groups: users or nonusers. You have two groups that are the same on observed and unobserved characteristics on average. Although they aren't really parallel users, the only difference between the groups in the aggregate is whether they use the technology.

Therefore, you can conclude that any differences in achievement are due to the technology, not other factors.

Exhibit 1. Random assignment



Ed Tech Rapid Cycle Evaluation Coach

BASELINE CHARACTERISTICS

An important question to ask is “Was the random assignment successful?” If random assignment was successful, the two groups will appear similar on measured background characteristics, such as test score results from previous years or demographic characteristics. A common way to assess this is to compare the average values of the groups’ background characteristics. We can quantify the difference between the two groups (users and nonusers) using a measure called an **effect size**¹. The RCE Coach’s random assignment dashboard automatically calculates the effect size for any variables you specify. The Coach’s random assignment dashboard will run and rerun its randomization until you have baseline equivalence on the characteristics you selected.

The final analysis also uses background characteristics, after you implement your technology and collect outcome data. It is necessary to include background characteristics in your analysis of the results if a lot of participants drop out of the study. The dropout rate in an evaluation is called attrition. For instance, if all students remain in the technology user group but 25 percent drop out of the comparison group, the two groups might no longer be similar. This attrition can occur for many reasons; for example, students might be absent on the day the test is administered or move to a different district or state, or teachers could decide not to participate in the study after random assignment. Accounting for background characteristics can rebalance the two groups, if the initial differences were not too large. However, it is a good idea to include some background characteristics in your analysis even if your evaluation went as planned. (To learn more about attrition, see the U.S. Department of Education’s What Works Clearinghouse overview of [attrition standards](#).)

© 2016, Mathematica Policy Research, Inc. This document carries a Creative Commons (CC BY) license which permits re-use of content with attribution as follows: Developed by Mathematica Policy Research, Inc. as part of the Rapid Cycle Tech Evaluations project funded by the U.S. Department of Education.



¹ An **effect size** measures different characteristics using the same yardstick. It is calculated by dividing the difference in means between the two groups by the standard deviation of the entire sample.