MATHEMATICA Policy Research

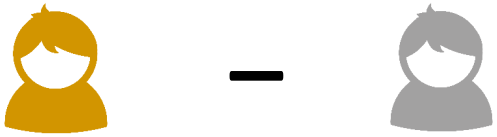
SEPTEMBER 2017

THINGS to CONSIDER

Choosing a Good Comparison Group

One way to test whether an educational technology is moving the needle is to conduct a **matched comparison design** evaluation. This method matches technology users to similar nonusers and then compares the differences in outcomes between the two groups. That difference is an estimate of the effect of the technology.

Technology's impact =



User outcomes **Nonuser outcomes**

However, to be confident in the results, your treatment and comparison groups (users and nonusers) must be similar. If they are not similar, the difference in outcomes may be due to differences between the groups rather than to the technology. Here are some things to think about when selecting two similar groups to compare.

To increase the likelihood of getting results you can trust…

Create groups that had similar characteristics before the technology was implemented (balanced)

Consider whether the method used to assign participants to groups will balance characteristics that can't be measured

Match on characteristics that you can measure and that are likely to influence the outcome

Include users and nonusers from several classrooms or schools

Conduct your own rapid cycle evaluation at http://www.edtechrce.org.

MATHEMATICA Policy Research

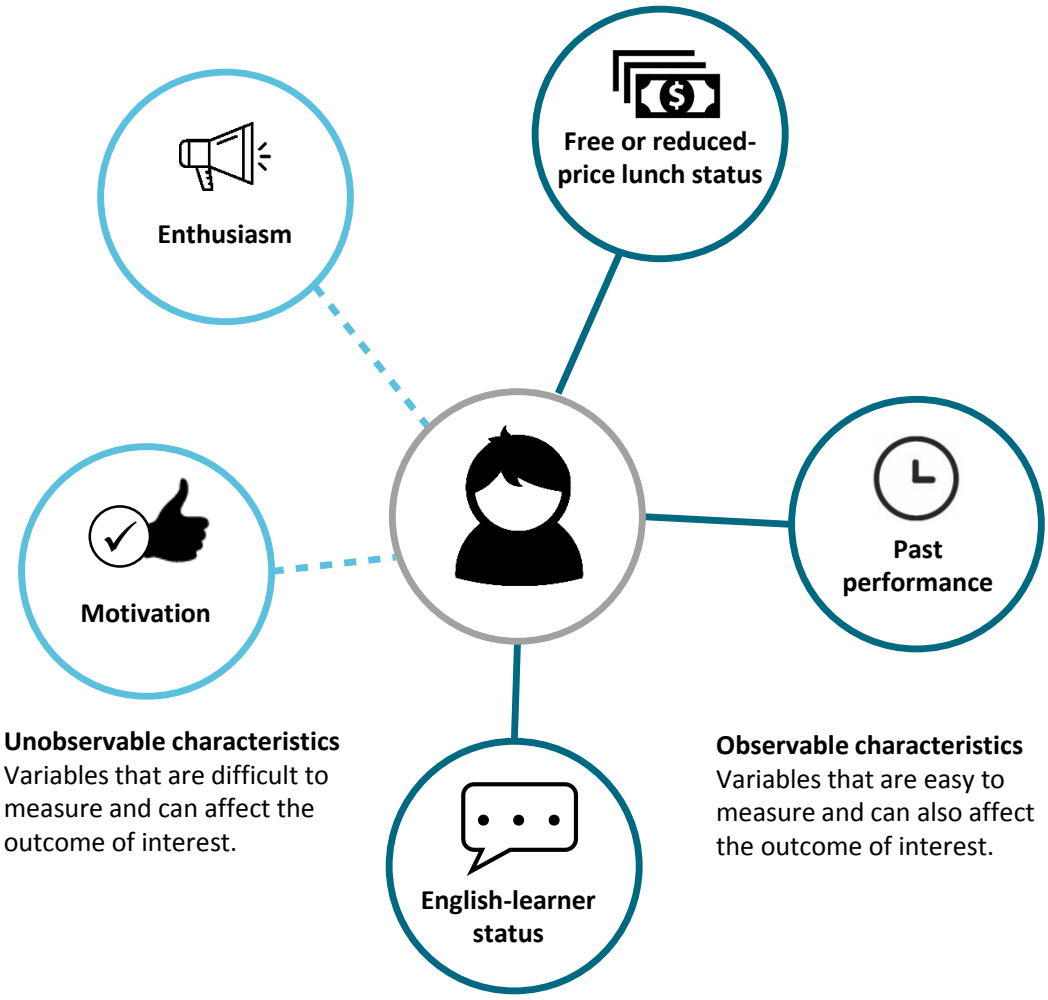
SEPTEMBER 2017

THINGS to CONSIDER

What makes my treatment and comparison groups similar?

Considerations when choosing a comparison group:
Similar characteristics

Every person has "observable" characteristics (which can be easily measured) and "unobservable" characteristics (which cannot be easily measured). Both types of characteristics can affect the outcome you are measuring. If your groups are not similar, then part of the difference you see could be caused by the differences in characteristics, not by the educational technology.



Unobservable characteristics
Variables that are difficult to measure and can affect the outcome of interest.

Observable characteristics
Variables that are easy to measure and can also affect the outcome of interest.

In a matched comparison design, first consider the potential unobservable characteristics that may affect the outcome. Try to find a potential comparison group that is as similar as possible to the group of technology users. Once users and nonusers are matched, you can confirm that the treatment and comparison groups are similar on observable characteristics by checking for "baseline equivalence" (a measurement of whether characteristics across groups are similar before the technology is implemented).

MATHEMATICA Policy Research

SEPTEMBER 2017

THINGS to CONSIDER

Are there differences between my treatment and comparison groups that I can't measure?

Considerations when choosing a comparison group:
Unobservable characteristics

Even if your treatment and comparison groups are equivalent on observed baseline characteristics, your users may have differences in unobservable characteristics. This is particularly likely if you created your groups based on who volunteered to use a technology or who your frequent users are because the volunteers may have unobserved characteristics related to the outcome measure (see the Volunteers box below).

When is the measured effect too small?

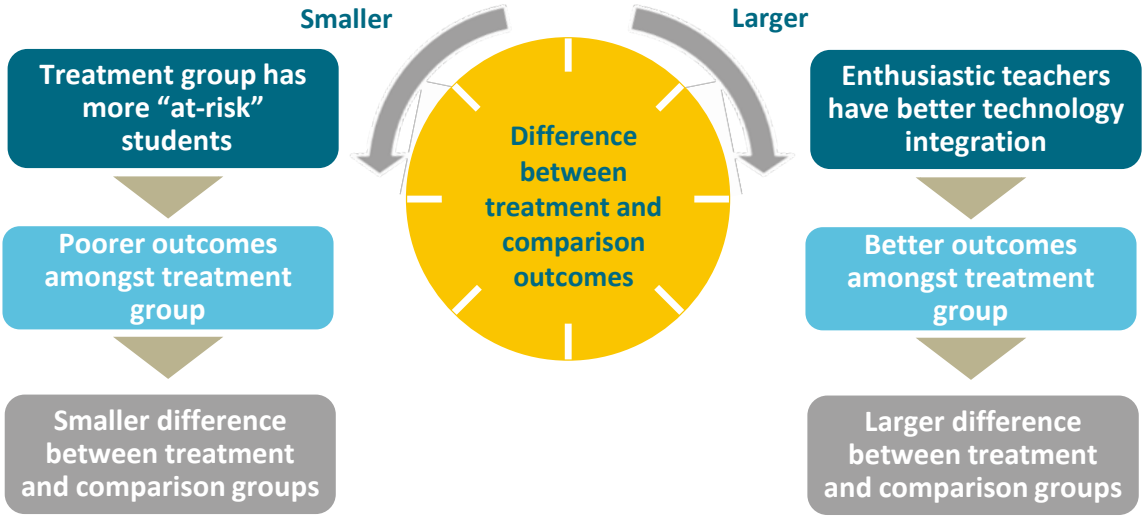
Sometimes unobservable characteristics can make the technology's effects look smaller (negative bias) than they actually are. For example, if an administrator selects students to take part in a program based on personal knowledge of an adverse home or family situation, such out-of-school factors might decrease the performance of those "at-risk" students, and the technology might seem to have a smaller impact on the treatment group.

When is the measured effect too large?

Sometimes unobservable characteristics can make the technology effects look larger (positive bias) than what they should actually be. For example, if an enthusiastic teacher is more likely to fully integrate the technology in his or her class, or if a motivated student is more likely to give the technology his or her full effort, you might see an impact on the treatment group that is too large

Volunteers

When you assign technology use based on who volunteers, you may introduce bias. Those who volunteer are likely different in unobservable ways from those who don't. They may be more motivated teachers or students with more engaged parents. These differences could also influence the outcomes you are measuring.



When there are unobservable characteristics that you cannot control for, you cannot determine whether differences in your outcome of interest are due solely to the technology or to unobservable characteristics. Think about what characteristics could be present and whether your method for selecting your treatment and comparison group, such as using volunteers, could have introduced unobservable differences between the groups.

MATHEMATICA Policy Research

SEPTEMBER 2017

THINGS to CONSIDER

What characteristics should I use for matching?

Considerations when choosing a comparison group:
Important observable characteristics

When conducting a matched comparison, you will have to determine which variables to include when creating a similar comparison group. You may have more variables in your dataset than you need.

Variables in your dataset

You should select only the variables that could influence your outcome of interest if they were different between the two groups. For example, if the outcome you are measuring is reading achievement scores, you may only want to select past performance in reading, free or reduced-price lunch status, and English learner status.

	Past performance	Birth date	Free or reduced-price lunch status	English learner status	Student athlete	Outcome of interest
Student ID	Fall reading test score	Birth date	Free lunch	English learner	Student athlete	Spring reading test score
167954	180	11/8/07	1	0	1	164
135497	105	9/23/07	0	1	0	137
157948	153	12/15/07	0	1	1	156
134957	124	4/15/08	1	0	0	146
165297	162	2/7/08	0	0	1	175
146852	141	10/31/07	1	1	0	167

Only include the variables that could be related to your outcome of interest because including more variables will make it harder to find a similar comparison group.

MATHEMATICA Policy Research

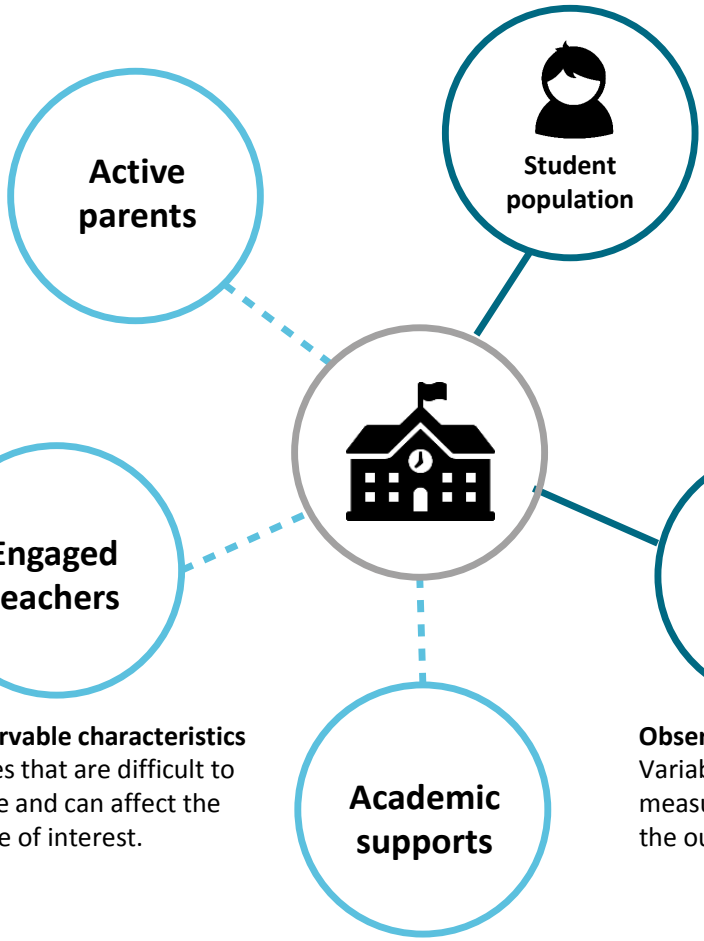
SEPTEMBER 2017

THINGS to CONSIDER

Do school or classroom characteristics matter?

Considerations when choosing a comparison group:
School and class characteristics

Just like individuals have observable and unobservable characteristics, so do schools and classrooms – and these characteristics could affect the outcome you are measuring. When the treatment and comparison schools or classrooms are similar, we assume that their unobservable characteristics are comparable and therefore do not bias the results. But if your schools or classrooms are not similar, then part of the difference you see could be caused by the differences in characteristics, not by the educational technology.



Unobservable characteristics
Variables that are difficult to measure and can affect the outcome of interest.

Observable characteristics
Variables that are easy to measure and can also affect the outcome of interest.

Think about unobservable characteristics as early as possible in your evaluation. If you suspect that there are differences across schools or classrooms, you may want to modify your evaluation design to reduce the likelihood that those differences will bias your results. You should also include contextual details about the potential differences in unobservable characteristics in your findings brief, report, or stakeholder presentation.

When you are assigning technology at the school or classroom level and you are interested in its effect on student outcomes, be sure to check baseline equivalence of both school or classroom characteristics *and* student characteristics. For example, the students in the schools may have similar pretest scores and background characteristics, but differences in information technology (IT) capacity across treatment and comparison schools could influence the outcomes. Or, IT capacity could be similar across schools, but student differences could skew the outcomes.

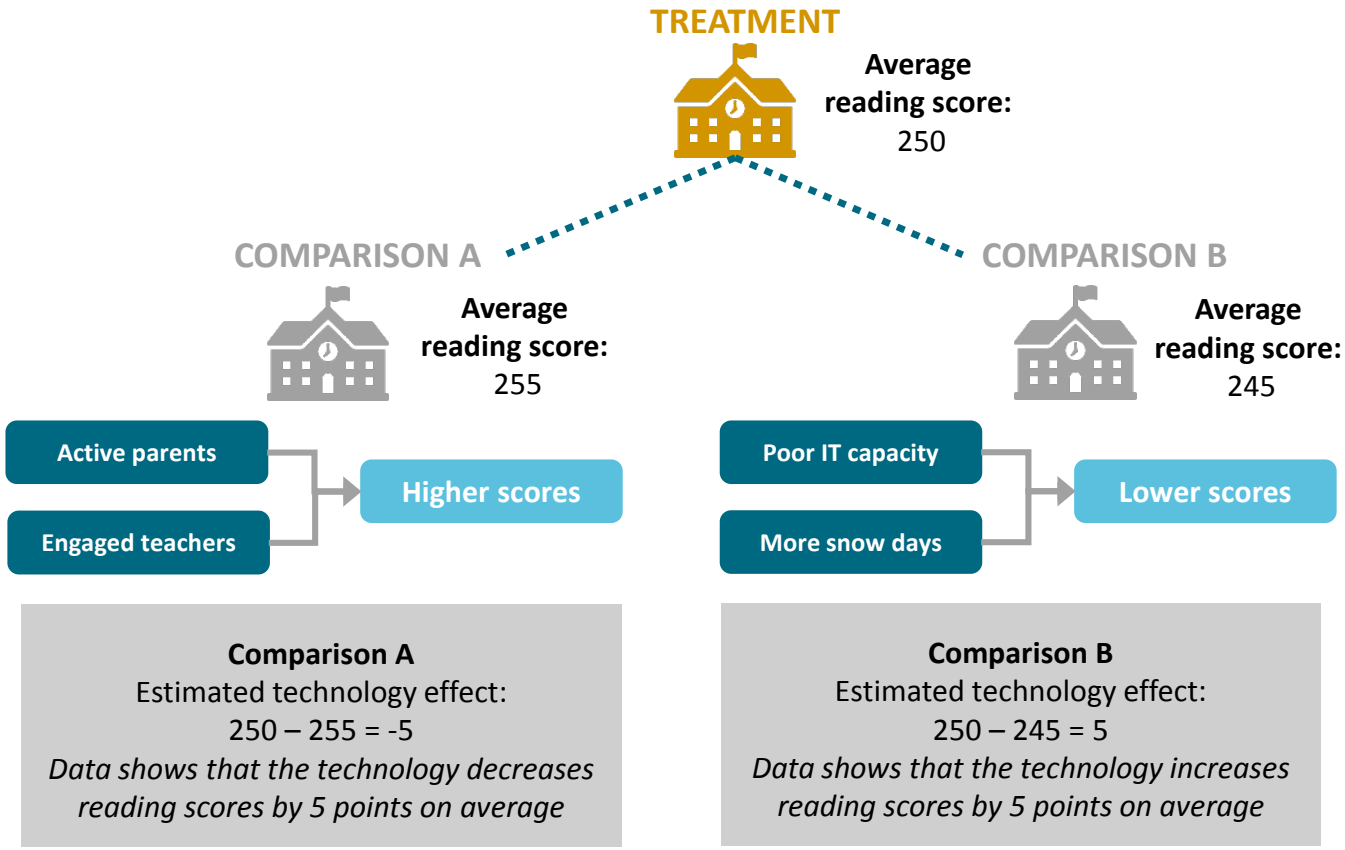
MATHEMATICA Policy Research

SEPTEMBER 2017

THINGS to CONSIDER

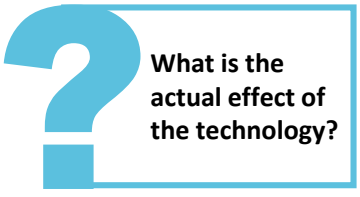
Considerations when choosing a comparison group:
The problem of one

When conducting evaluations, the more participants you include in your treatment and comparison groups, the more confident you can be in your results. This is because a larger number of participants reduces the likelihood that unobservable characteristics are biasing the outcomes. However, even with a large number of students in your evaluation, if the technology is only assigned to one school, then unobserved differences at the school level could influence the outcomes.



Comparison A
Estimated technology effect:
250 – 255 = -5
Data shows that the technology decreases reading scores by 5 points on average

Comparison B
Estimated technology effect:
250 – 245 = 5
Data shows that the technology increases reading scores by 5 points on average



When we have only one school or one classroom in the treatment or comparison group, we cannot know the actual effect of the technology. Comparing one school to another is equivalent to comparing one individual to another. This kind of comparison makes it hard to determine what portion of the difference is attributable to the technology and what portion is attributable to differences in school characteristics. Differences in characteristics could make results look larger or smaller, making it hard to determine whether the technology is moving the needle.