Selecting Districts and Schools for Impact Studies in Education: A Simulation Study of Different Strategies

Journal of Research on Educational Effectiveness

Daniel Litwok, Austin Nichols, Azim Shivji, and Robert Olsen

Experimental studies of educational interventions are rarely designed to produce impact evidence, justified by statistical inference, that generalizes to populations of interest to education policymakers. This simulation study explores whether formal sampling strategies for selecting districts and schools improve the generalizability of impact evidence from experimental studies.

Which selection strategies produced samples with the greatest generalizability to the target population?

Our simulation study evaluates a hypothetical intervention targeting K–5 schools. We construct a national target population of schools from the Common Core of Data and generate impacts of the intervention for the entire population. From this population, we simulate samples for the evaluation by selecting districts, simulating district decisions about whether to participate, selecting schools in those districts, simulating school decisions about whether they agree to participate, and replacing districts and schools that decline to participate until the target sample size is reached. We calculate the average school-level impact for the resulting sample of schools and compare it to the average impact for the target population. The simulation repeats this procedure many times, each time selecting a different sample.

The selection strategies we test include: (1) a stylized version of purposive selection—sometimes referred to as convenience sampling—that recruits districts and schools in order from largest to smallest; (2) random selection with probabilities proportional to district size, as used in some surveys; and (3) balanced selection, which prioritizes the most typical districts and schools based on their characteristics. We test all combinations of these three approaches for both districts and schools.

Our study finds that random selection of districts with either balanced or random selection of schools produced samples with the most consistently strong generalizability.

How do selection strategies affect recruiting burden?

The cost of conducting an experimental study depends on the number of districts that are actively recruited to participate. By prioritizing the largest districts, purposive district selection minimized recruiting burden. Recruiting burden increased by 20 percent with random district selection, in which large districts were oversampled, and more than doubled with balanced district selection, in which districts were chosen independent of their size. Studies facing budget constraints may need to favor larger school districts, at least to some extent, to constrain recruiting burden.

When using random selection, how should declining districts be replaced?

When using random selection, we tested two alternatives for replacing districts that decline to participate: replacement with another randomly selected district ("random replacement") and replacement with another district with similar values of key characteristics ("nearest neighbor replacement"). In terms of generalizability, random replacement outperformed nearest neighbor replacement in many but not all scenarios.

Were the findings sensitive to simulation parameters?

We found that the study's main findings were not sensitive to the values of key simulation parameters, such as the variance of the intervention's impacts across schools and ability of observed variables to explain that variation. This suggests that that the conclusions of the study may apply to range of experimental studies of different educational interventions.

Full Article Citation:
Litwok, D., Nichols, A., Shivji, A., & Olsen, R. (2022). Selecting Districts and Schools for Impact Studies in Education: A Simulation Study of Different Strategies. *Journal of Research on Educational Effectiveness*. DOI: https://doi.org/10.1080/19345747.2022.2128952.