Applying Generalizability Index Method to Examine the Representativeness of Cost Study Samples

Authors: Anyi Wang[1], Atsuko Muroga[2] and Brooks Bowden[3]

Context

This study explores a way to examine sample representativeness and generalizability of cost analysis of educational interventions. Educational cost studies are often not designed along with effectiveness evaluation, but conducted retrospectively. Retrospective cost studies do not necessarily have the privilege to include all samples in the evaluation. This is especially true for rigorous cost studies, which involve iterative data collection through interviews and meticulous documentation of resources used in educational interventions (Levin & McEwan, 2001). This process is labor intensive and relies on the cooperation of participants. Therefore, drawing a subsample from the evaluation as a convenient sample becomes an appealing option, despite at the cost of losing representativeness. Cost studies based on non-representative samples may still be internally valid, but the findings may not be generalizable and can be less meaningful to draw policy implications for a larger population.

Research question

Given this context, two important questions arise for cost studies using such convenient samples: how representative is the sample to the larger population of policy relevance, and what methodologies can be applied to examine the representativeness of the study sample?

We focus on one recent educational intervention in Boston, *City Connects*, which provides comprehensive student support services to students in participating schools. The intervention involves the assessment of student needs and strengths, the referral of students to appropriate services, and following up. The whole process utilizes resources from the schools, community partners, as well as parents and volunteers. The ideal approach to select cost study sample is to include all sites from an effectiveness evaluation because these sites are already a representative sample from a larger population. However, when the effectiveness study sample is large, given the labor intensive and iterative nature of cost data collection, including all effectiveness evaluation sites in the cost sample is challenged. In that case, a subsample must be selected ideally with some sampling method (e.g., probability sampling, stratification, etc) to ensure that the subsample is as representative as the original sample. The problem faced with our cost study sample (i.e., two treatment schools and four counterfactual schools) is that the selection of subsample was purposive—based on the program implementation fidelity and/or the willingness to participate in the study—rather than statistical. The concern is that the two treatment schools with high implementation fidelity may not be representative to all treatment schools, and that the four counterfactual schools may be different from a typical school in Boston Public Schools in

[1]Ph.D., Senior Research Assistant, Center for Benefit-Cost Studies of Education (CBCSE), Teachers College, Columbia University (Email: aw2544@tc.columbia.edu)
[2]M.A., Research Assistant, CBCSE, Teachers College, Columbia University (Email: am3994@tc.columbia.edu)
[3]Ph.D., Assistant Professor of Methods and Policy, North Carolina State University & Assistant Director and Director of Training, CBCSE, Teachers College, Columbia University (Email: brooks_bowden@ncsu.edu)

general. Bearing in mind these challenges, we set out to empirically examine to what degree our subsample of treatment and counterfactual schools are representative of their own groups.

Data

In our case study, we use school-level data for all BPS schools extracted from the *Massachusetts Department of Elementary & Secondary Education (DESE) - School and District Profile* database from 2001 to 2016 (i.e., from School Year 2000-2001 to 2015-2016). The school level variables we gathered covers student demographics, attendance, class size, and academic achievement. Additionally, we obtained information on years of implementation from the central program office.

The original data we obtained from DESE contains missing values for certain observations and inconsistent measures across years. To mitigate loss of sample size and potential biases due to these reasons, we supplement our dataset with multiple imputation using chained equations.

Method

We present three pairs of comparisons to examine our sample representativeness. The first pair is between the two sampled treatment schools and all *City Connects* participant schools. The second pair of comparison is between the four sampled counterfactual schools and all BPS schools that are not part of *City Connects*. These two comparisons provide evidence on whether our sampled schools represent the treated and control population. Finally, the third comparison is between our sampled schools and all BPS schools, which may indicate whether the results from the cost analysis can be generalized to BPS.

We follow Tipton (2014) by using a propensity-score-based generalizability index to quantify the extent to which units in a particular sample are similar to or different from the larger population. The advantage of the generalizability index is that it takes into account the variation between schools across an array of observable characteristics and combine them into a single measure of overall similarity. The method was originally developed to assess the generalizability of samples in randomized controlled trials or other outcome evaluations, and our study is the first application in the context of cost studies. However, this single index is not enough to uncover which specific characteristics in the sample are different from the larger population, i.e., the difference in student demographics and academic achievement. These differences are also important contextual information for cost studies and policy deliberation, beyond knowing whether the sample schools are representative. Therefore, to examine the specific differences between schools under each pair of comparison, we also calculate standardized mean differences (SMD) for selected school characteristics.

Implications

Educational policy makers often face the choice of multiple alternative programs with similar educational objective. Cost analysis coupled with effectiveness measure informs policy makers of the efficiency of the alternatives, and assists them in choosing the ones that provide the best results for any given resource or the ones that utilize the least resource for any given outcome.

Ideally, cost analysis should be designed *ex ante* along with the evaluation of program effectiveness, and should include all sites within the evaluation. In reality, many cost studies are conducted retrospectively and cannot afford to collect cost data from all units in the sample for effectiveness analysis. In this paper, we illustrate with a specific example of *City Connects* program that the methods we apply can help determine whether and to what degree the results of a cost study using a convenient sample can be generalized to a larger population to inform policy decision. These methods can provide more clarity around generalizability and representativeness when the constraints compromise the design. We believe that our study can shed light on future cost analysis of educational interventions, especially those with causal evidence that support expansions.

(997 words, excluding title and sub-headings)

Preliminary reference list

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (Vol. 4). Sage.

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *Journal of Educational and Behavioral Statistics, 39*(6), 478-501.