Design and Analytic Features for Reducing Biases in Skill-Building Intervention Impact Forecasts
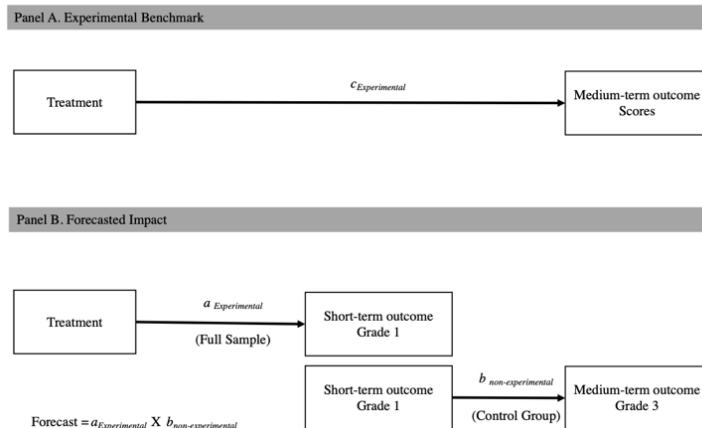
Daniela Alvarez-Vargas, Sirui Wan, Lynn S. Fuchs, Alice Klein, & Drew H. Bailey

Despite policy relevance, long term evaluations of educational interventions are rare relative to the amount of end of treatment evaluations. A common approach to this problem is to use statistical models to forecast the long-term effects of an intervention based on the estimated shorter term effects. Such forecasts typically rely on the correlation between children's early skills (e.g., preschool numeracy) and medium-term outcomes (e.g., 1st grade math achievement), calculated from longitudinal data available outside the evaluation. This approach sometimes over- or under-predicts the longer-term effects of early academic interventions, raising concerns about how best to forecast the long-term effects of such interventions. The present paper provides a methodological approach to assessing the types of research design and analysis specifications that may reduce biases in such forecasts.

What did we do?

We use the within-study design shown in Figure 1 to assess various approaches to forecasting the impacts of early math skill-building interventions 2 years after treatment end. We treat the observed experimental impact estimates on medium- term outcomes (2 years after treatment end) as a benchmark. We compare this benchmark to forecasts calculated using short-term experimental impact estimates in combination with nonexperimental longitudinal associations between early and later skill measures from the same studies. We compare various study design and analytic features for forecasting medium-term impacts. Specifically:

(1) the measurement of baseline covariates (types of pretests)

(2) the types of short-term skills assessed (*proximal:* skills directly taught by the intervention v. *distal:* broad and comprehensive skills)

(3) the specification of forecasting models (regressions assuming independent v. overlapping mediational pathways)

What did we find?

The forecasts closest to the experimental benchmark were obtained by including comprehensive baseline controls and experimentally estimated effects on conceptually proximal and distal short-term outcome measures. The specification of forecasting models did influence the accuracy of forecasts.

Three Applications

The approach can be applied for calculating power to detect medium-term effects, for choosing a set of short-term outcome measures, for funding organizations interested in forecasting the effects of proposed interventions on student achievement years after the end of treatment, and for researchers and policy analysts attempting to forecast future program benefits. Overall, our work highlights the importance of considering multiple competing biases including, but not limited to, omitted variables bias, measure over-alignment, and measure under-alignment.

Figure 1
Conceptual Framework of Within Study Comparison of Number Knowledge Tutoring