External Validity in the Context of RCTs: Lessons from the Causal Explanatory Tradition

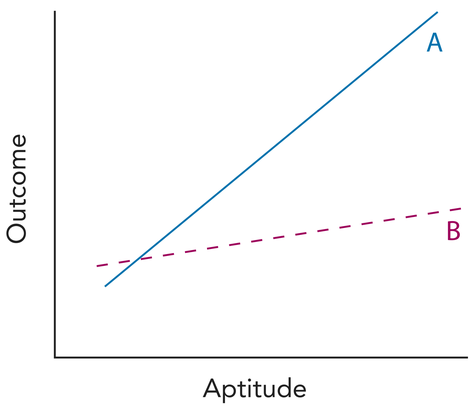
Andrew Jaciw, Denis Newman
Empirical Education Inc.

Empirical Education

Our experience designing and conducting a number of randomized experiments in the past few years has given us a new appreciation of some of the seminal writing of Lee J. Cronbach. We borrow the term 'causal explanatory tradition' from Professor Denis C. Phillips who explained the basic ideas to the first author, and conveyed that Cronbach's approach to program evaluation emphasized mechanisms and explanations of treatments in context.

We recommend in particular:
Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 116-127.
This is mostly about the psychology laboratory where the moderating mechanisms are psychological or cognitive but applies directly to our field experiments on K-12 instructional programs.

Regressions Within Two Treatments



Cronbach was concerned with aptitudes of individuals especially in the psychological laboratory, but in field experiments in education—when we focus on moderators—we are interested in attributes of context that can exist at different organizational levels that interact with the treatment of interest.

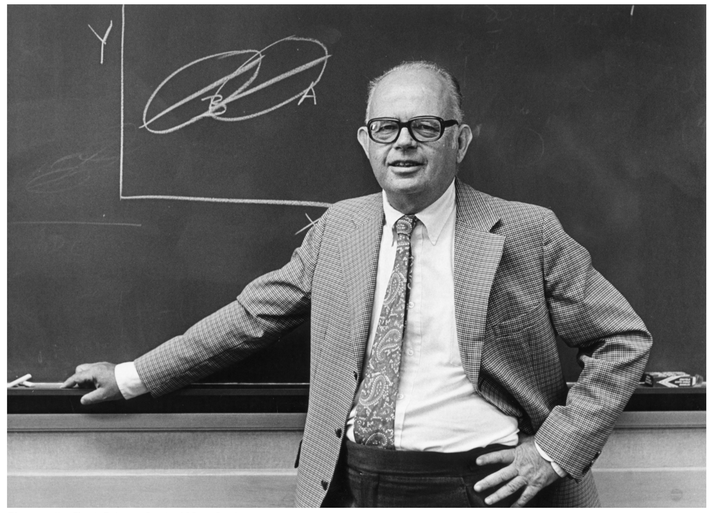


Photo credit: Chuck Painter, Stanford News Service

Three experiments as case studies:

1. ***The effects of small class size: Tennessee STAR experiment.***
A well-known case where results from one context, did not extrapolate to another (viz. California)
Cronbach: ***"When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion…positive results obtained with a new procedure for early education in one community warrant another community trying it. But instead of trusting that those results generalize, the next community needs its own local evaluation"*** (p. 125).

2. ***Studying the scale-up of an innovation.***
Scaling up means bringing innovations into new contexts with different conditions under which the innovation operates and with which it interacts.
Cronbach: ***"Instead of making generalization the ruling consideration in our research, I suggest that we reverse our priorities. An observer collecting data in one particular situation… will give attention to whatever variables were controlled, but he will give equally careful attention to uncontrolled conditions …. As results accumulate, a person who seeks understanding will do his best to trace how the uncontrolled factors could have caused local departures from the modal effect. That is, generalization comes late, and the exception is taken as seriously as the rule"*** (pp. 124-125).

3. ***Multi-year evaluation of a K-12 program.***
It can take four years for the results of a two-year impact analysis to be vetted and for the report to be released. Most interventions undergo continuous improvement so the need to maintain the same treatment condition in the experiment results in ecological invalidity.
Cronbach: ***"Generalizations decay"*** (p. 122).

Question:

What can we do to make better use of our experiments—in their planning, execution, and analysis; in creating a basis for external validity; and in producing results that are relevant, especially for primary stakeholders?

Generalization predicated on interactions
Cronbach: ***"Once we attend to interactions, we enter a hall of mirrors that extends to infinity"*** (p. 119).
Implication: Cronbach's famous quote expresses the realization that 2nd, 3rd, 4th, etc. order interactions might always be moderated at the next higher level. This does not make experimentation impossible and we are not rejecting experimental control or the need to establish internal validity. But the exploration of interactions can be more important than establishing an average causal impact in finding how best to apply the intervention the next time. The most interesting moderators are those that may be most productive in the particular locale.

Generalization as a process of developing a detailed account of treatment in context
Cronbach: ***"An observer collecting data in one particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context.... As he goes from situation to situation, his first task is to describe and interpret the effect anew in each locale, perhaps taking into account factors unique to that locale…"*** (p. 125).
Implication: A detailed account of how a treatment works in diverse contexts does more to inform the general picture than corroborating an average effect estimate across many contexts. Moderators that account for heterogeneity in the impact across contexts can give insight into general mechanisms; however, unaccounted-for heterogeneity may indicate how treatment is operating in relation to factors unique to locales. Details about site-specific effects serve generalizability by providing a track record that new locales can look to, in order to determine how the program is likely to play out in their cases.

Explanation as a basis for generalizability
Cronbach: ***"...systematic inquiry can realistically hope to make two contributions. One reasonable aspiration is to assess local events accurately….The other…is to develop explanatory concepts"*** (p. 124).
Implication: A central aim of experiments should be to corroborate or refine the theory of the causal explanatory mechanisms by which effects are achieved. This should be more than just running analyses of moderating and mediating effects as superfluous exploration using available 'convenience variables.' Rather, we should take seriously the tasks of theorizing moderators and mediators before the experiment, puting in the resources to measure them well, and assessing whether they produce theorized effects.

Generalization, as concerned with treatment under changing conditions, and with conditions for the evolution of the treatment
Cronbach: ***"Short-run empiricism is 'response sensitive.'...one monitors responses to the treatment and adjusts it, instead of prescribing a fixed treatment on the basis of a generalization from prior experience"*** (p. 126).
Implication: Treatments can mature in the life cycle of an experiment. They can trigger unexpected mediating processes. Implementation studies should be sensitive not just to the presence of predicted occurrences, but also, to occurrences of the unpredicted and conditions for those events. One should question the value of drawn out experiments where subjects are blinded to the effects for many years—this may not be in the subjects' best interest—and reviewers should be sensitive to unanticipated positive or adverse effects that call for termination of the study to minimize harm or deprivation.

Note. *Qualification:* In this work we consider the ideas of Lee Cronbach concerning problems of program evaluation in outlining an approach to the planning and conduct of randomized trials that is meant to be more responsive to context and that addresses the problem of generalizability. This work does not intent to represent Cronbach's philosophy of evaluation. For him experiments were limited in their application and he exhorted researchers to use multiple rigorous methods. We believe however that the experimental paradigm can be strengthened by incorporating some of Cronbach's main principles.