



CTE | Career & Technical Education
RESEARCH NETWORK



Introduction to Causal Research

August 22, 2020

Trey Miller, Principal Researcher

Stephen Ross

The work of the CTE Research Network Lead is supported by the Institute of Education Sciences at the U.S. Department of Education with funds provided under the *Carl D. Perkins Career and Technical Education Act* through Grant R305N180005 to the American Institutes for Research (AIR). The work of the Network member projects is supported by the Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education



Introduction

- The goal of causal research is to provide evidence of the effectiveness of a program, intervention, or policy change on one or more desired outcomes.
- We can't directly observe the difference between:
 - Outcomes for those receiving a program, and
 - The **counterfactual**: the outcomes *that would have happened* had the same people not received the program during the same timeframe.
- Well-designed causal research studies provide a credible simulation of the counterfactual with an unbiased comparison group.



An Example

- An intensive mentoring program for students at risk of not graduating high school might compare graduations rates as evidence of its effectiveness.
 - Graduation rate for students in mentoring program = 85%
 - Graduation rate for students in the comparison group = 75%
- How confident are you that difference in graduation rates (+10%) is due to the mentoring program?
- Depends, of course, on how well the comparison group provides an unbiased representation of the counterfactual.



Validity

- **Statistical conclusion validity** asks if a finding is accurate. Is there adequate power, sample size? Are assumptions of statistical tests violated? Are measures reliable?
- **Construct validity** asks how well the inferences made about higher level constructs represented in the study match what was implemented and what was measured as an outcome.
- **External validity** asks whether a cause-effect relationship holds under variation in persons, settings, treatment variables, and measurement variables.
- **Internal validity** asks whether changes in the treatment condition actually impact changes in the outcome. Did something other than the treatment explain variation in the measured outcome?



Threats To Internal Validity

- **Selection:** systematic differences in the kind of students in the treatment and comparison groups
- **Attrition:** observed effect might be caused by a systematic differential drop out between the groups
- **History:** observed effect might be caused by an influential event or policy change taking place during the study period.



Threats To Internal Validity

- **Maturation:** observed effect might be caused by subjects getting older or more experienced between pre- and post-test
- **Testing:** observed effect caused by subjects doing better on post-test simply because of experience of taking the pretest
- **Regression:** if inclusion in study is based cut-score, “regression to the mean” can look like an effect between pre- and post-test
- **Instrumentation:** difference between groups in how outcomes are measured groups



Non-experimental Designs: Pre-post

Before we review casual research designs, let's briefly consider some nonexperimental research designs.

Pre-post design studies compare the performance of a group of individuals before and after exposure to the treatment.

- No guards against the internal validity threats of history, maturation, or testing.
- There is no way to know if an improvement is caused by the treatment.



Non-experimental Designs: Correlational

- **Correlational designs** look at the relationship between two variables.
- For example, a positive relationship between visits to a school counselor in the first 2 years of high school and number of advanced mathematics courses by the end of grade 12 doesn't offer evidence of a causal relationship between the two.



Non-experimental Designs: Descriptive

- The value of well-done **descriptive studies** should not be overlooked.
- Research that clearly describes features, variations, and trends about an education topic can be critical to building understanding, detecting patterns, generating hypotheses, or contributing to the interpretation of a causal study.



Experimental Design

- **Randomized control trials (RCTs)** are considered the gold standard in evaluation research.
 - Random assignment ensures that the treated and untreated groups will be similar on both observable and unobservable characteristics, with the treatment being the only systematic difference between the two groups.
 - In a well-done RCT with a large enough sample size and low attrition, the difference in outcomes between the treatment and control groups can confidently be attributable to the treatment.



Quasi-experimental Design

- **Quasi-Experimental Designs (QED)** studies are designed to ensure that the treatment and comparison groups are equivalent, but the equivalence can be verified on only observed, measured variables.
 - QEDs are designed to reduce the possibility that there may be important preexisting or contextual differences between the treatment and comparison groups.
 - When done well, QEDs offer evidence of a causal relationship, but the evidence is not as strong as evidence from an RCT.



Quasi-experimental design: Regression Discontinuity

- **Regression Discontinuity (RD):** students or schools receive an intervention because they scored above or below a cut-point on some measure.
 - Groups can be compared on outcomes while controlling for differences on the qualifying measure.



Quasi-experimental design: Comparative Interrupted Time Series

- **Comparative Interrupted Time Series (CITS):** groups are measured across time.
 - Program impacts are evaluated by looking at whether the treatment group deviates from its baseline trend after the intervention differently than the comparison group.



Quasi-experimental design: Matched Comparison Groups

- **Matched Comparison Groups** design includes a treatment and comparison that have been designed to be similar on important characteristics.
 - There are multiple ways to use data to inform group assignment including propensity score methods to create a comparison group by matching intervention and comparison use on characteristics that are likely to be related to the outcome of interest.



High-quality Quasi-experimental Designs

- **Distinct groups:** There should be a treatment and control group of separate students (unlike pre-post designs)
- **Baseline equivalence:** You need evidence that groups are similar on key variables at the start of the study, so that differences in outcome measures won't be mistaken for an effect of the intervention.
- **Free of confounding factors:** You want to make sure there is no factor that may influence the outcome and is aligned with only one group.
- **Appropriate outcome measure:** Instruments used to measure outcomes must be reliable and valid, and not be “over-aligned” to the intervention.



What Works Clearinghouse

- The WWC uses a systematic review process to examine intervention research.
- To be eligible for the WWC's highest rating for group design studies, **Meets WWC Group Design Standards Without Reservations**, the study must be an RCT with low levels of sample attrition.
- A QED or a high-attrition RCT is eligible for the rating **Meets WWC Group Design Standards With Reservations** if it satisfies the WWC's baseline equivalence requirement that the analytic intervention and comparison groups appear similar at baseline.
- A QED or a high-attrition RCT that does not satisfy the BLE requirement receives the rating **Does Not Meet WWC Group Design Standards**.



Implementation Research

- In addition to examining causal evidence of the effectiveness of a program, you will also want to verify a program was delivered according to plan. You want to verify **fidelity of implementation**. If the intervention was not implemented as it was designed or intended to be used, the conclusion of its effectiveness will not be valid.
- Implementation research also includes the investigation of what happens when a program or intervention is adopted – how a program is developed and implemented locally, what elements are enacted in a real-world setting, and why certain contexts, settings, and factors influence how the program or policy is implemented.



Mixed Methods

- A well-designed quantitative study can offer evidence of the program's effectiveness, it may leave some important questions unanswered.
- **Mixed-method designs** incorporate qualitative research methods – for example, interviews, focus groups, case studies, or observations – to provide a richer understanding of a program's implementation and effectiveness.



CTEResearchNetwork.org



CTEResearchNetwork@air.org