

CPP 524: Eval II – Research Design

Research Design Project Requirements

You are the lead evaluator for an organization or agency of your choice (hypothetical or real). You are submitting a grant to secure sustained funding for a real-world program or policy that your organization is responsible for administering. “Real-world” program can mean an actual program that you are familiar with, a program that is described in enough detail in existing literature that you can complete the assignment, or a new program that you would like to create and is based upon existing models that are similar enough to use as benchmarks (it is rare to think up a program that doesn’t already exist in some form already).

As part of the grant submission process the foundation requires an impact evaluation of the program in its current state. The grant will cover program costs and evaluation costs during the study period. Depending upon the results of the evaluation the foundation may or may not make a long-term commitment to the program, so the quality of your evaluation can impact a significant future funding stream and the financial sustainability of your organization.

Your job is to design the evaluation and provide details in your proposal, but to avoid conflicts of interest and the potential for bias in the study the foundation will hire a third-party evaluation firm to implement the data collection and analysis using the framework that you provide. The evaluation needs to be rigorous to receive serious consideration by the foundation and it needs to be clear enough for the third-party firm to implement your design.

You need to submit a memo that includes the following:

1. A clear program description
 - a. Basic program details including the target population and program eligibility
 - b. The theory of change and intended impact
2. A rigorous evaluation design
 - a. Valid and reliable measure(s) of impact
 - b. A rigorous and feasible estimator
 - c. The approach used to create a valid counterfactual
 - d. The most appropriate study time-frame
 - i. duration of the intervention
 - ii. length of observation after the intervention ends
 - e. Which type of estimate will be reported (treatment on the treated or intention to treat)
3. Justification for the proposed design
 - a. Does your proposed measure capture the most important dimensions of impact?

- b. Why is your estimator better than the two you did not choose?
 - c. What are costs of a shorter time-frame, of a longer time-frame?
- 4. Threats to validity
 - a. Apply the Campbell Score framework to your proposed study
 - b. Outline other implementation challenges the evaluation firm should anticipate

The goal is to demonstrate that you can apply principles of research design that you have learned in the class. It needs to be done in such a way that an external team of evaluation experts have enough information to (1) make sense of the program model and intended impact, and (2) implement the evaluation using the details you have provided.

Your instructor will serve as a proxy for the external evaluation specialist. Your evaluation design must be clear enough that a professional evaluator can read the proposal and move forward with the study without consulting you at later dates. In other words, you need to provide details on all of the major design choices for the study (type of estimator, how the outcome is measured, time-frame, etc). The third-party team is responsible for logistics of implementation.

You will start with a literature review to find at least three evaluation or performance studies that examine a similar program model. Use the background research to see which evaluation designs others have used and how they have measured program impact.

You will need to select at least one latent construct that can be used as an outcome or a control variable in the study, even if it is not your primary outcome of interest. You will need to identify a valid instrument (one that has been developed and tested) and report the reliability scores.

Selecting an Evaluation Design

- Your primary task is to identify which design is most appropriate for your program:
 - Pretest-posttest control group design (diff-in-diff)
 - Posttest only control-group design
 - Pre-post reflexive design
- **You are not allowed to use a pure experiment where you have full control over assignment into treatment and control groups.** Very few real-world evaluations have the opportunity to randomly assign individuals into categories, so I want you to think creatively about designing an evaluation when you don't have that privilege.
- How will you ensure that the research question is framed around a plausible counterfactual as the point of reference, and what steps need to be taken to ensure it is a valid comparison group? Specifically, does your comparison group represent what the world would have looked like if the program had not been implemented? Do your comparison group demographics need to be identical to the treatment group given the estimator you have selected for the study (T2-C1, T2-T1, or the diff-in-diff). Think hard about questions like selection and non-random attrition. Explain what other studies have done to address these issues and lay out a plan for how you reevaluation will address the issue.

- How will data be collected? And over what period?

Your paper should have the following structure:

PART 1A – Program Overview and Research Question (1-page memo form – see example below)

The first page of the evaluation should be a succinct explanation of the program in memo form:

- What is the problem addressed by the program?
- What is the program, policy or intervention that will be evaluated?
- Explain the program theory in a couple of sentences?
- What is the primary outcome of interest?
- Who is eligible to participate in the program (program inclusion criteria) and who is excluded?
- Optional: motivate the study if there is a specific context that makes the research especially timely or salient.

This cover page should be exactly one page in length, single-spaced.

PART 1B – Theory of Change

You will diagram **the essential elements** of your program model (the most parsimonious and abstract version of your program). Each causal link in the model represents an assumption. You will enumerate the list of assumptions and discuss your level of confidence in each, discussing which parts of the program might be most sensitive to implementation integrity. In other words, conduct a brief and informal pre-mortem analysis of the most likely causes of program failure.

PART 2 – Evaluation Design

2A – Measurement of the Outcome

How will you measure the dependent variable in the analysis?

You must select at least one latent construct and identify an appropriate "instrument" that has been validated i.e. how do we know that we are measuring what we think we are measuring? Don't try to reinvent the wheel – 99 out of 100 times you will be able to find existing instruments that can be adapted for the study.

Unless you have the instructor's permission you are NOT allowed to use the latent construct measures used in the warm-up lab (happiness, grit, SF36 health). The goal is to learn how to search for validated instruments within the literature, identify alpha scores, and cite your resources properly.

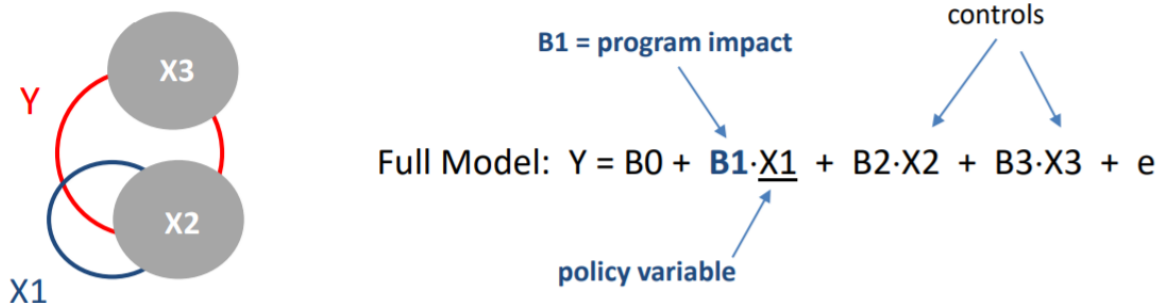
If the most natural dependent variable is something tangible and concrete, you can instead suggest adding a latent construct as a control variable. For example, if you want to see whether a job training program in prison reduces the likelihood of recidivism the outcome (reincarceration) can be measured concretely without error.¹ You could add a personality construct like IQ, motivation, or tolerance for change as a control variable to improve the study.

Refresher on the taxonomy of controls:

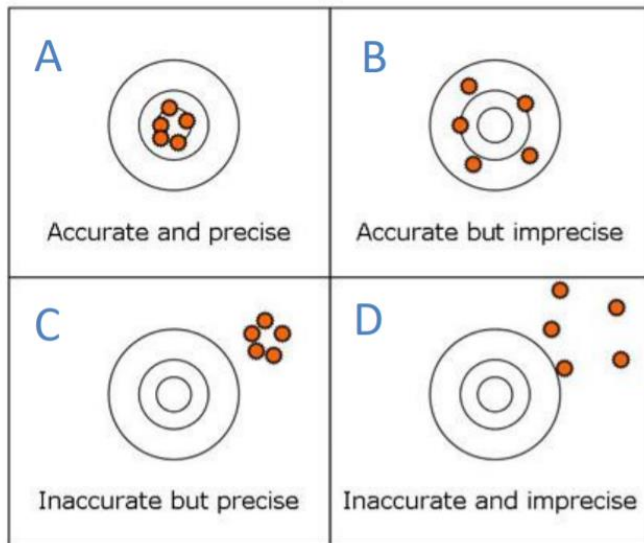
Controls that are independent of the assignment to treatment are thus independent of the policy variable but can explain additional variance, thus reducing model residual and decreasing standard errors, giving your study additional power without increasing sample size.

X2 is something that is correlated with the intervention, which typically means selection into a program or non-random attrition out of a program. Thus, is it a competing hypothesis and needs to be “controlled” to avoid bias. Randomization of participants into study groups (or similar quasi-experimental alternatives) breaks the correlation between X1 and X2, eliminating bias in the study. You “control” competing variables through study design.

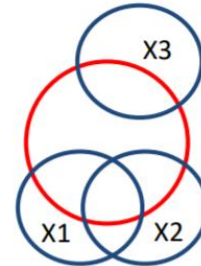
X3 type variables, on the other hand, can be very useful to your study because they can be added to any of your regression models and they will improve the standard errors associated with your study variable X1. It's always good to add these types of controls to your study.



¹ Reincarceration is a proxy for what we care about – whether job skills reduce the likelihood that the formerly incarcerated participate in illegal activities – innocent people can be profiles and arrested, and guilty people may not get caught. But their actual behavior is a much harder thing to measure so reincarceration is a decent proxy.



$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$



2B – The Estimator

Explain your proposed estimator for the study and why it is appropriate:

- Pre-post reflexive design (T2-T1)
- Posttest only control-group design (T2-C2)
- Pretest-posttest control group design (diff-in-diff)

2C – Construction of the Counterfactual

How will you ensure that the research question is framed around a plausible counterfactual as the point of reference, and what steps need to be taken to ensure it is a valid comparison group?

Specifically, does your comparison group represent what the world would have looked like if the program had not been implemented?

What assumptions need to be true for your estimator to be valid?

- Do study group demographics need to be equivalent given the estimator you have selected for the study?
- Is the estimator sensitive to secular trends? If so, do you have a way to test for trends in the study to show that they do not exist and thus will not introduce bias?

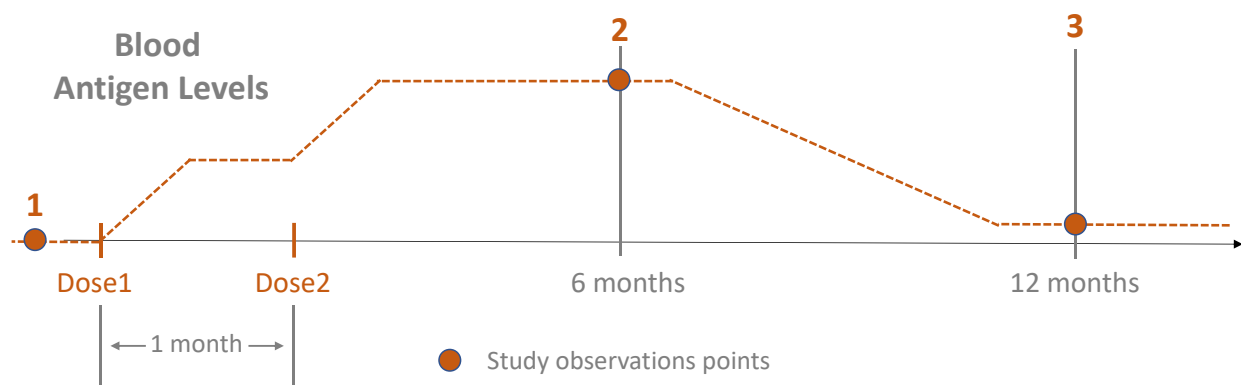
2D – Study Timeframe

Report your study period:

- What is the first observation period?
- What is the duration of the treatment?
- How long should you wait after the treatment ends before taking your final measure of the outcome?

Your justification for the time frame should come from previous research or descriptive data from studies of similar interventions. For example, “previous studies show that antigen levels for influenza vaccines decline in the period 6 months to 12 months after the last vaccine is administered, so we use that as the benchmark time-frame for a coronavirus vaccine study.” Only conjecture if you absolutely must.

Here is where being clear about your research question will pay dividends. For example, “Is the COVID vaccine effective” seems like a straight-forward statement of the evaluation problem. However, consider the following scenario:



The treatment period is only a month, but the study lasts for approximately 13 months (one for the treatment period, and 12 more months until the last observation is taken).

Note that you would need a measure at point 1 (before the first vaccine dose) to ensure you are not including people that already have antigens in their blood from exposure to COVID that created natural immunity through fighting the disease.

If antigen levels drop to almost zero after a year and patients are easily infected past that point was the vaccine effective?

Well, it completely depends on what we mean by effective. It was very effective at building up blood antigen levels and fighting disease for several months, but effectiveness wanes quickly in this hypothetical example. If we pick the 6-month timeframe we would conclude that vaccine works. If we pick the 12-month timeframe we might conclude that the vaccine does not work that well.

The distinction here is very important:

- Is the vaccine effective at inducing an immune response that is protective against COVID?
- Is the vaccine effective at providing long-term protection from disease?

If the goal is to prevent hospitalizations during a peak period in a pandemic, then short-term efficacy might be enough. If the goal is eradication of a disease entirely then long-term immunity is quintessential.

Be clear about how you operationalize effectiveness with regards to program goals and your study timeframe. “We are interested in immune response 6 months after the second dose is administered.” Or “we are interested in sustained immune response measured 12 months after the full vaccine regiment ends.”

2E – Treatment Effects

Report whether the **treatment-on-the-treated (TOT)** or the **intention-to-treat (ITT)** estimate is more appropriate for your study and explain why.

Make sure your proposed treatment measure is feasible in your study. For example, if you have no way of observing compliance it would be impossible to use the TOT measure.

If compliance is not binary (you got the vaccine or you did not) how would you define compliance in your study?

For example, if study participants are required to take a medication on a daily basis for 30 days but they forget one day, is that non-compliant? What if they forget a couple of days? What if they forget to take the medication on half of the days in the study? Absolute non-compliance is easy – the participant refuses to take ANY of the medication. But at what point do you place someone in the non-compliance category for failing to follow basic program protocols?

You can leave some of these details up to the third-party evaluator but they might make arbitrary decisions that make your program look less effective!

PART 3 – Justification for your Proposed Design

3A – Literature Review

Report your findings from the three studies you read. Specifically:

- What programs were evaluated?
- Which research designs (estimators) were used?
- How were outcomes measured?

3B – Justify your Design Decisions

Describe the counterfactual and you will use for the study and diagram your expectations about what the data will look like if your program is successful.

Use the diagram to explain the assumptions you are making about group equivalency and secular trends.

Justify your choice of an estimator. Explain why you believe it to be the best identification strategy for your research question by comparing it to the two options you did not choose.

Provide any important details about the construction of the counterfactual group within your study that the external evaluation team should consider when designing study logistics. What is the process used to assign study participants to groups? Do you expect attrition, and is there any reason to believe it would be non-random?

What assumptions must be met for your estimator to be valid (group equivalency and secular trends)? Consequently, what tests should be conducted to ensure the counterfactual is robust?

PART 4 – Threats to Validity

Using the Campbell Scores framework enumerate the potential competing hypotheses in the study:

- Address the ten competing hypotheses outlined in the Campbell Scores in your study. In essence, you will rate your own research design on a scale of 1 to 10. Note that it is very unusual for a study to receive a perfect ten, so I am looking for you to clearly identify issues, think through how they might be addressed, and if they cannot be addressed completely how would you report the limitations to ensure you are upholding high standards of transparency and ethics as an external evaluator. You can address additional competing hypotheses if there is something specific to your research domain that is worth noting.
 - Selection
 - Non-random attrition
 - Maturation (if applicable)
 - Secular trends (if applicable)
 - Study time-frame (how do you know it's long enough?)
 - Testing (is there reason to suspect it in your case?)
 - Regression to the mean
 - Seasonality
 - Intervening events
 - Measurement error
- The primary hypothesis in your study is, “You can explain the results (the estimated effect) because of the impact of the program or policy.” The term ***competing hypothesis*** refers to a statement, “You can explain the results through bias introduced by an omitted variable” or “You can explain the results as a result of maturation, selection bias, etc.”
- The goal of the evaluation design is to eliminate ALL competing hypothesis so the only remaining explanation comes from the impact attributed to the program. It is rare that a single evaluation is able to eliminate all competing hypotheses. For those that are not

addressed through your research design, articulate why the hypotheses cannot be eliminated.

Note that you won't know the answer to some of the items until you collect the data. For those items describe what you expect to happen, or a possible or likely scenario, and explain whether the study design can address the potential issue.

For example, you will not know whether intervening events *have actually occurred* until you implement the study. For this item list the **most likely** intervening event that one might anticipate beforehand and describe its impact on the study (would it effect both study groups equally or would it disproportionately impact one group?). Based the score for this item on whether the event would introduce bias if it were to occur.

For attrition, speculate on what non-random attrition might look like in your study. Are participants likely to leave the treatment and control groups at equal rates? Are the attriters likely to be high performers, low performers, or random? Explain why. And can your estimator / study design handle non-random attrition? What are scenarios where attrition would bias your study, and what are scenarios where non-random attrition would be ok?

Formatting

The first page (the program overview memo) will be single-spaced. Use 1.5 spacing for the rest. Your paper will be approximately 15 pages in length. In addition, be sure to include references for cited works.

The first page is written as a memo:

Evaluation of the Urban Naturalist Environmental Education Program

The problem: In a world that is increasingly dominated by technology, children are generally spending less time outdoors and have only limited interaction with nature. Lack of time outdoors has been associated with obesity, higher rates of asthma and allergy, and lack of concern for environmental issues.

Program: The Urban Naturalist program provides outdoor education to students in 15 public schools in Atlanta. Children meet at a local park once a week during the school year and participate in science lessons and outdoor wilderness activities.

Program theory: Case studies from a recent report have shown that students often learn more effectively within an environment-based context than within a traditional education framework and the benefits can include improved performance on standardized measures of academic achievement, reduced classroom management problems and increased enthusiasm for learning.² Among the five case studies examined in the report, the researchers observed performance improvements in reading, math, science and social studies and saw classroom discipline problems decline.

The dependent variable in the study: The study will examine overall classroom performance of 8th-graders that are part of an environment-based education program.

Motivation of the Study: Legislation has been proposed to provide \$500 million over five years to schools with approved “environmental literacy” plans for students in grades kindergarten through 12, and offer competitive grants to schools and non-profits for outdoor education projects. Strong evidence is needed for program effectiveness before this kind of support is provided.³

Program inclusion criteria: All 8th-graders in the 15 Atlanta public schools included in the program are eligible to participate. There is a fee structure based upon the free-lunch status of students. Those eligible for free lunches can participate for free while others have to pay a \$50 participation fee. There are no enrollment limits for this group. Students in the adjacent school districts can apply for the program if there are extra openings, but they are not guaranteed a spot and have a pay a nominal registration fee of \$65.⁴

² G. Lieberman, & L. Hoody, “Closing the achievement gap: using the environment as an integrating context for learning” (San Diego, CA, State Education and Environmental Roundtable, 1998).

³ Penny Starr, “No Child Left Inside Act Would Spend \$500M Teaching Environmental Literacy Starting in Kindergarten.” *CNSNews.com*, 14 May 2009, <http://www.cnsnews.com/news/article/48164> (accessed November 5, 2010).

⁴ Information from the program website at <http://fakewebsite.com>.

Grading Rubric

Part 1 – cover page (6)

- Problem (1): What is the problem addressed by the program?
- Program (1): What is the program, policy or intervention that will be evaluated?
- Program Theory (1): Explain the program theory in a couple of sentences?
- Outcome (1): What is the primary outcome of interest?
- Program Inclusion Criteria (1): Who is eligible to participate in the program (program inclusion criteria) and who is excluded?
- Motivate the study (1): if there is a specific context that makes the research especially timely or salient (optional).

Part 2 – outcome measure (6)

- Identification of appropriate latent construct (2)
- Justification (2)
- Proper citation of reliability (2)

Part 3 – theory of change (10)

- Coherent theory of change (4)
- Quality of the diagram (3)
- Explanation of assumptions (3)

Part 4 – model and competing hypotheses (14)

- Explanation of the counterfactual (3)
- Diagram of expectations (3)
- Justification of the design (3)
- Campbell Scores (5)

Writing quality (2)

Proper citations (2)

Total: 40 points

