Evaluating Your AMEP

What is evaluation?

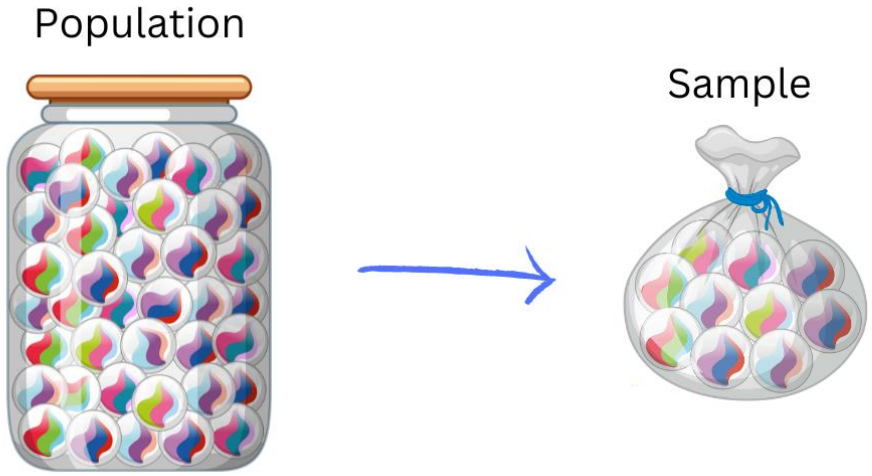
All evaluation is research, not all research is evaluation

Evaluation research answers questions about the effectiveness of programs, interventions etc.

Type of Evaluation	Purpose
Formative or Process 	Formative evaluation helps identify what's working well and what needs adjustment (e.g., fidelity of implementation). Think of this type of evaluation as a chef tasting and seasoning soup as it is prepared.
Summative or Outcome 	Summative evaluation assesses the overall effectiveness and outcomes of a project or program once it's completed, determining its success. Think of this type of evaluation as the restaurant critics rating of the soup the chef prepared.

Quantitative Evaluation Design

- Experimental
- Quasi-Experimental
- Within-Subjects/Longitudinal



What limitations does each example have in terms of generalizability?

Population: the entire group of learners that possess the characteristic being studied

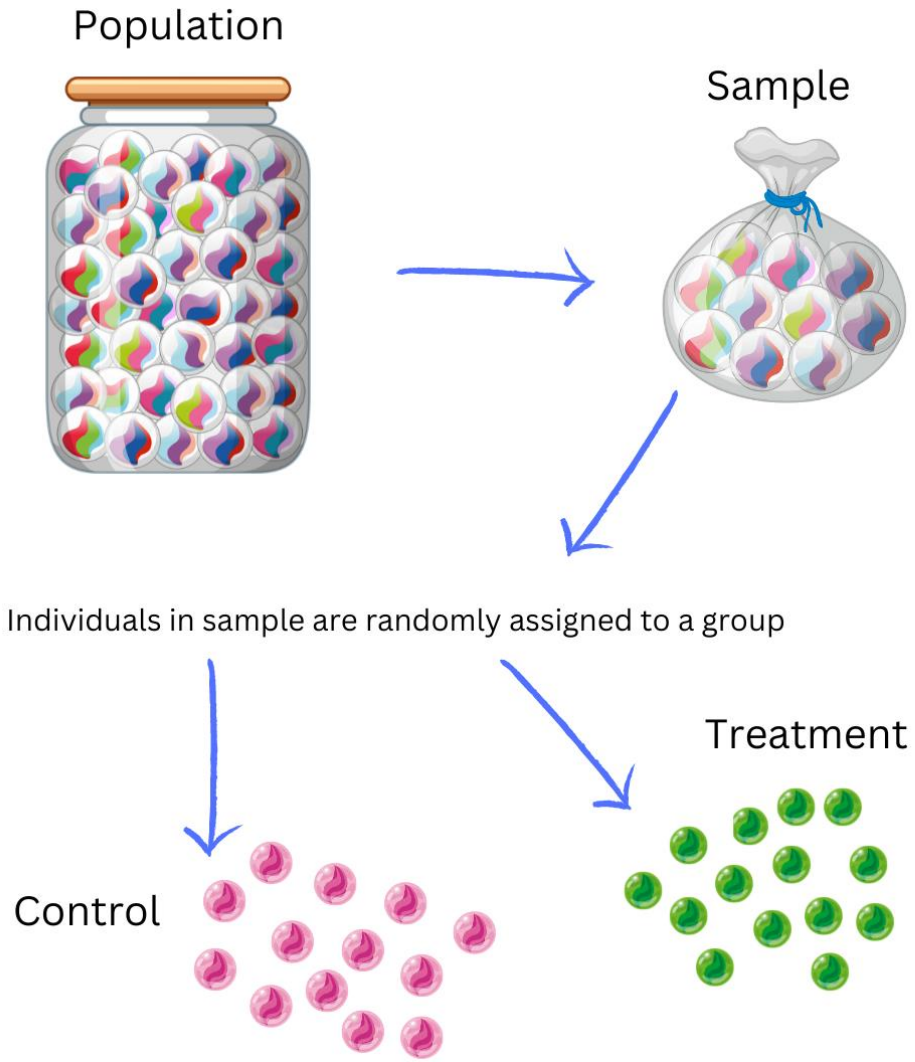
Example: All IM residents in the US

Sample: the group of learners you are going to collect data from to make a general inference about the population. Ideally, your sample should be randomly drawn from your population

Example A: 30% of IM residents at Boston University

Example B: 92% of IM residents who attended a workshop at a national meeting

Example C: ~18% of IM residents in the US who responded to a survey that was sent out to them by their PDs (which is a list of 100 emails based on your own personal network and tedious google searches)



True experimental design includes:

Random Sampling: Every member of the population has an equal chance of being selected for the study. (rarely possible)

Random Assignment: Participants are assigned to treatment or control groups randomly to reduce bias.

Controlled Conditions: Experimental conditions are carefully managed to isolate the effects of the independent variable (Group: Control or Treatment).

Challenges to Experimental Design in Med Ed

- Resource intensive
- Challenging to execute logistically alongside other education and clinical responsibilities
- Limited sample sizes (e.g., fellows) do not provide sufficient statistical power

Quasi-Experimental Design

Quasi-Experimental Design

- Like experimental design, but without the random assignment

- Treatment and comparison groups are based on pre-existing characteristics

Quasi-Experimental Design

- **Year of Residency**: Residents in odd-numbered years (e.g., first-year, third-year) could be assigned to the treatment group, while residents in even-numbered years (e.g., second-year, fourth-year) could be assigned to the comparison group.

- **Previous Addiction Medicine Training**: Residents with prior training or experience in addiction medicine could be assigned to the treatment group, while residents without such experience could be assigned to the comparison group.

- **Performance on Pre-Education Assessment**: Residents who scored above a certain threshold on a pre-education assessment of addiction medicine knowledge could be assigned to the treatment group, while residents who scored below that threshold could be assigned to the comparison group (regression discontinuity)

Best Practices for Quasi-Experimental Design

- Try to identify groups based on things that are not related to the outcome(s) you are studying

- Before starting, considering the pros/cons of using different grouping characteristics

Example

To conduct a quasi-experimental evaluation of a stand-alone two-hour didactic session within an internal medicine (IM) residency program using x+y scheduling, you could:

- **Identify Treatment and Comparison Groups**:
  - **Treatment Group:** Residents who attend the stand-alone two-hour didactic session during an ambulatory clinic week (y) in the x+y scheduling model.
  - **Comparison Group:** Residents who do not attend the stand-alone two-hour didactic session during an ambulatory clinic week and instead continue their regular activities in the x+y scheduling model.
- **Determine outcome measures to assess the impact of the stand-alone didactic session. These could include:**
  - Pre- and post-tests to evaluate knowledge acquisition.
  - Resident feedback surveys to assess satisfaction and perceived effectiveness of the session.
  - Performance evaluations during subsequent clinical rotations to measure application of knowledge.

Within-Subjects or Longitudinal Design

Within-Subjects/Repeated Measures Design

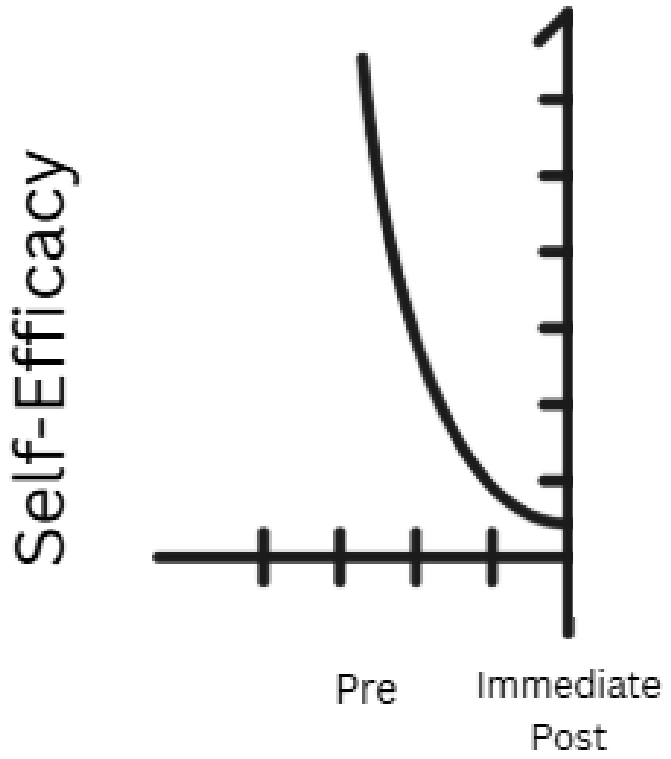
- No separate comparison group — individuals serve as their own control (past-self as counterfactual).
- All learners receive the intervention **OR** the research question focuses on change over time.
- Measures changes within individuals, such as:
  - Knowledge
  - Attitudes
  - Behaviors

Self-Efficacy versus Confidence

- Imagine you are treating a patient with opioid use disorder (OUD) who is hesitant to start medication for OUD due to concerns about withdrawal symptoms and stigma. How would you approach this situation?
- **Self-Efficacy (specific):** Please rate your confidence in your ability to effectively educate the patient about the benefits of MOUD and address their concerns about withdraw al symptoms and stigma on a scale of 1 to 5, with 1 being not confident at all and 5 being extremely confident.
- **Confidence (general)**: Please rate your overall confidence in your ability to manage patients with OUD and provide comprehensive addiction medicine care, on a scale of 1 to 5, with 1 being not confident at all and 5 being extremely confident.

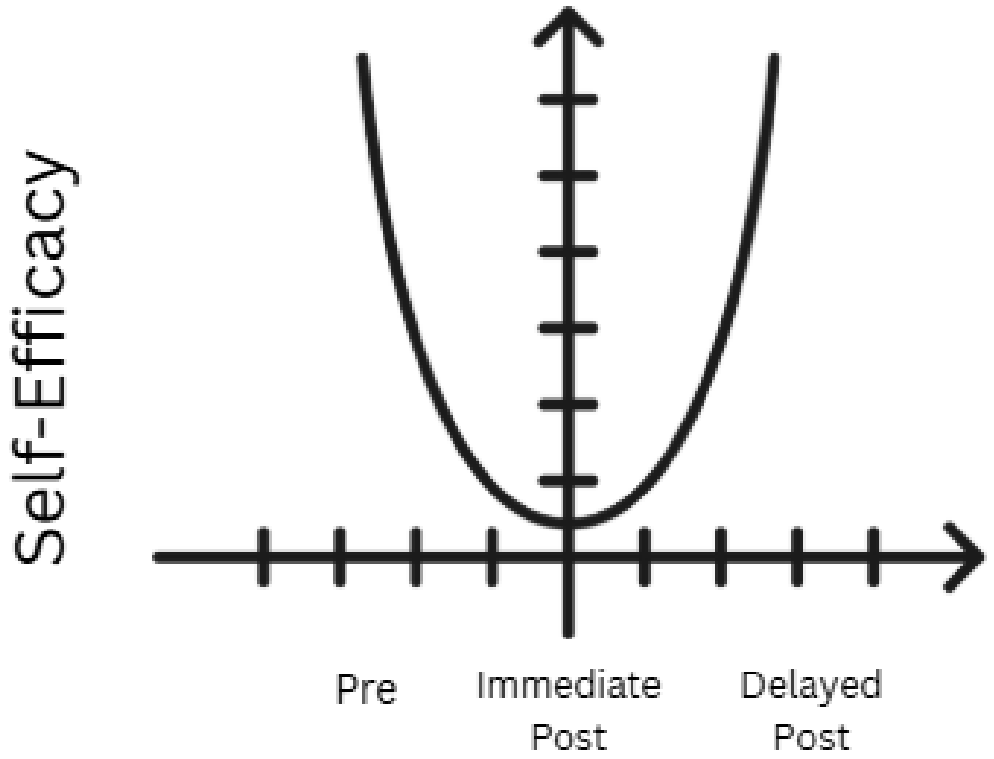
Within-Subjects Design

- Need at least two assessment points for pre-/post- comparison

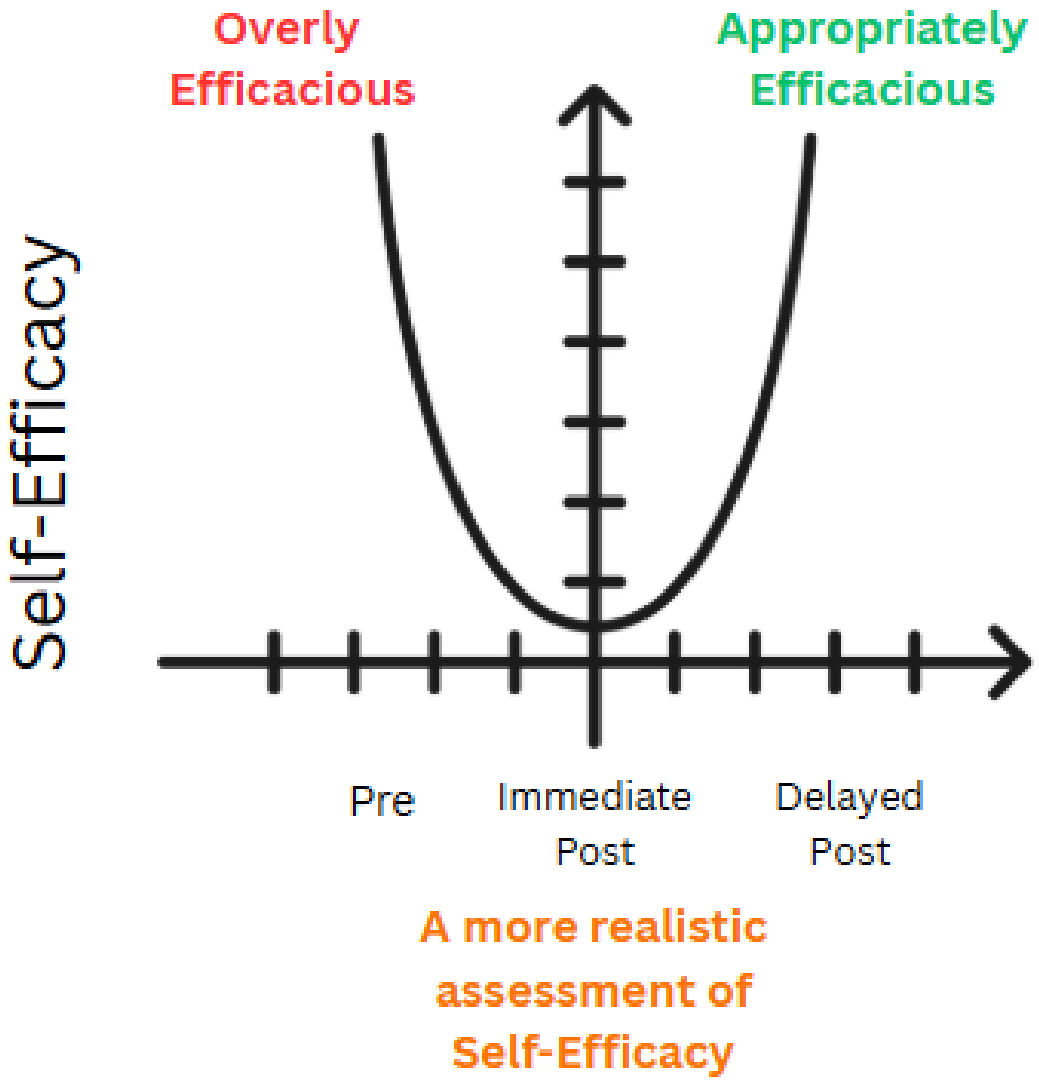


Within-Subjects Design

- If you want to look for longitudinal change, using 3+ assessment points is best



Within-Subjects Design



Tips for using Within-Subjects Design

- Make sure you are thoughtful about *when* you are collecting data

- Make sure you'll be able to detect change if it occurs
  - Select measures that you're likely to see some variability on
  - Needs assessment data are very helpful here

BE SURE YOU HAVE A WAY TO LINK THE PRE- AND POST- DATA

Quick Recap

- **Experimental design**
  - Involves random assignment to treatment and control groups.
  - **Allows you to claim:** Your intervention caused the differences between groups.

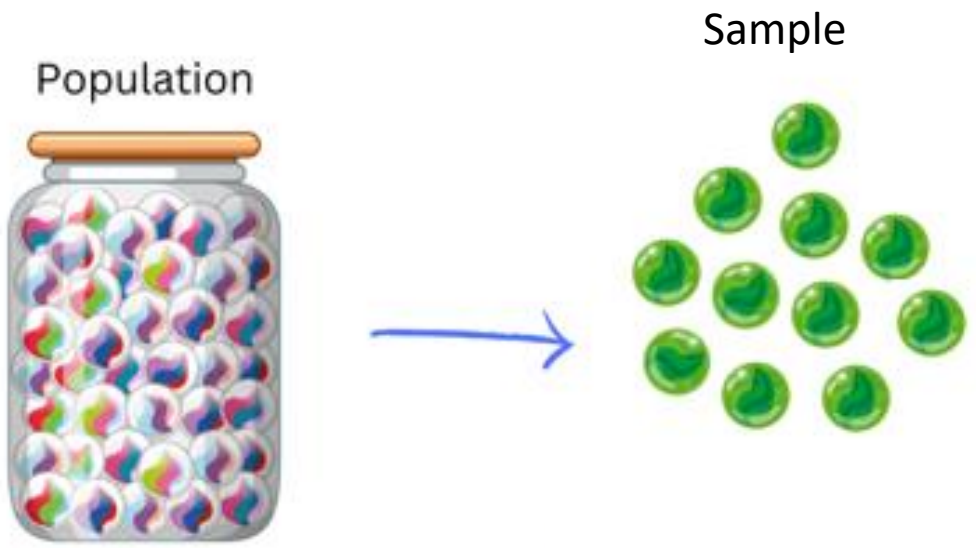
- **Quasi-experimental design**
  - Similar to experimental design but without random assignment; groups are based on pre-existing characteristics.
  - **Allows you to claim:** Your intervention is associated with differences between groups, but causality is harder to confirm – especially depending on what you used as your grouping variable.

- **Within-subjects/longitudinal design**
  - Involves assessing each individual at 2 or more time points, often before and after an intervention.
  - **Allows you to claim:** Changes over time are associated with the intervention, but other time-related factors may also contribute.

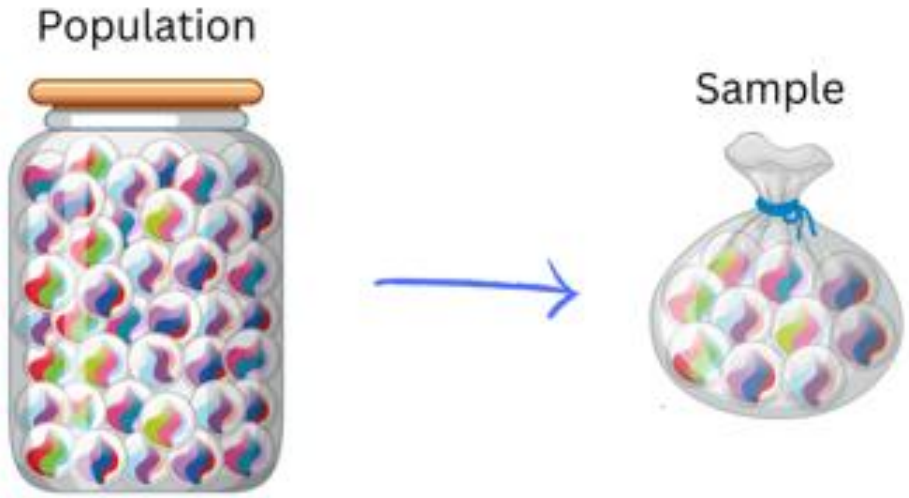
Qualitative Evaluation

- Whereas with quantitative evaluation the goal is to generalize findings to a larger population, qualitative research focuses on developing a deep, contextual understanding of specific experiences, perspectives, or phenomena.

- The goal is not to generalize, but to capture rich, detailed insights that reveal complexity and meaning.



If you think of quantitative sampling as grabbing a handful of marbles at random from a giant bag, qualitative sampling is more purposeful like carefully sifting through the marbles to select specific ones whose unique experiences or characteristics you want to understand.



You may also use random sampling in qualitative research when your goal is to capture a wide range of experiences or perspectives without introducing selection bias. However, even then, the focus remains on depth and richness of understanding, not on statistical generalization

Qualitative Evaluation

- Capture learners' experiences and perspectives in their own words
- Identify unexpected outcomes or hidden barriers that quantitative approaches may miss
- Amplify diverse voices and uncover nuances across different learner groups, especially learners who have been historically-excluded from medicine
- Foster a deeper, more-nuanced understanding of how and why educational interventions succeed or struggle
- Can contextualize and enrich quantitative findings

Qualitative Evaluation

- Interviews
- Focus groups
- Open-ended evaluation questions?

Qualitative Evaluation

- Interviews
- Focus groups
- ~~Open-ended evaluation questions?~~
  - Usually **- but not always -** open-ended survey responses are too short for a full qualitative analysis. In those cases, you can group the responses into categories and report simple descriptive statistics on how often each category appears

Balmer DF, Rama JA, Martimianakis AT, Stenfors-Hayes T. Using Data From Program Evaluations for Qualitative Research. *Journal of Graduate Medical Education*. 2016 Oct;8(5):773–774. doi:10.4300/JGME-D-16-00540.1

Qualitative Evaluation

- Interviews: One-on-one conversations that allow for deep, personal exploration
  - Best for sensitive topics and when individual perspectives are important

- Focus Groups: Group discussions that explore shared and differing experiences
  - Best for understanding group norms, dynamics, and a range of views
  - A focus group is a guided discussion where participants interact with each other, and those interactions are part of the data you're analyzing! It is not just a group interview

Interview vs Focus Group

- **Interview Question**:
  "What barriers have you encountered when applying the new clinical skills?"
  - Unique barriers that may not be visible to others; individualized context.
  - Responses may be more honest due to increased privacy

- **Focus Group**:
  "What barriers have you noticed when applying the new clinical skills?"
  - Common barriers, differences between groups, and how discussion reveals priorities (or social pressures)

Interview vs Focus Group

- **Interview Question**:
  "Can you share how your background or identity shaped your experience in this program?"
  - Personal reflections on belonging, support, challenges, and how their identity intersected with learning

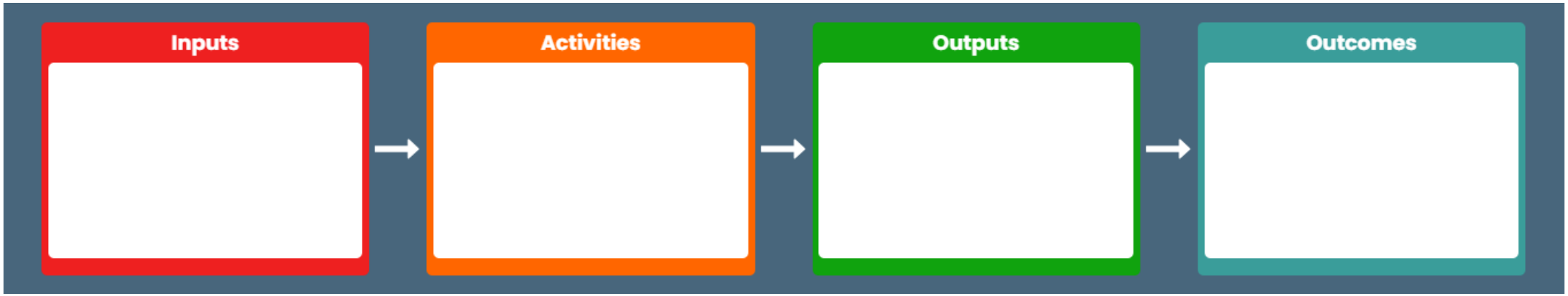
- **Focus Group**:
  "How do you think the program supported (or failed to support) learners from historically excluded backgrounds?"
  - Collective insights into structural barriers, peer experiences, shared themes of inclusion or exclusion, and how group dynamics influence perceptions of support

	Interviews	Focus Groups
Time	More time-consuming	More efficient
Scheduling	Easier to schedule (only 2 people)	Harder to coordinate (multiple schedules)
Scope of Data	Deep personal insights	Broader range of experiences
Social Influence	Private; no peer influence	Group interaction shapes responses
Sensitive Topics	Better for sensitive or personal topics	Sensitive topics may be harder to discuss openly
Analysis Complexity	Usually simpler (single-person narratives)	More complex (tracking individual + group dynamics)

Questions?

lindsay.b.demers@gmail.com

Logic Models



- Logic models keep planning focused, helping prevent scope creep.
- They also help clearly demonstrate the thru-line from Inputs to Outcomes

Logic Models

- Inputs
  - List all the resources needed to implement your curriculum. Consider funding, faculty, instructional materials, technology, learner assessment tools, partnerships, and any other necessary support.

- Example:
  - Faculty member with expertise in addiction medicine
  - Resident time (2-hour block during academic half-day)
  - Pre-test and post-test (knowledge quizzes)
  - Lecture slides and handouts
  - Classroom space

Logic Models

- Activities
  - Describe the instructional strategies or learning experiences that will be included in your curriculum. What lessons, workshops, assessments, or clinical experiences will be implemented? How will they be delivered?

- Example:
  - Deliver a 2-hour interactive lecture covering core topics
    - Introduction to substance use disorders (SUDs)
    - Screening and brief intervention techniques
    - Overview of treatment options and referral pathways
  - Administer a pre-test before the lecture to assess baseline knowledge
  - Use case-based discussion during the lecture to apply concepts
  - Administer a post-test immediately after the lecture to measure knowledge gained
  - Provide handouts summarizing key points and local resources

Logic Models

- Outputs
  - List the measurable products or experiences that will result from your curriculum. Be sure to include learner assessment methods such as quizzes, OSCEs, self-assessments, or performance evaluations.

- Example:
  - Number of residents who attend the lecture
  - Number of residents who complete the pre-test
  - Number of residents who complete the post-test
  - Pre- and post-test score comparisons to measure knowledge gain
  - Distribution of handouts and resource materials to all participants

Logic Models

- Outcomes: Think about changes in participants' knowledge, attitudes, skills, or behaviors both short- and long-term. How might those changes influence their practice or environment over time and connect to your program's broader goals?

- Example:
  - Residents gain foundational knowledge about addiction medicine.
  - Residents apply screening and brief intervention techniques in clinical settings.
  - Patients with substance use disorders are more frequently identified and referred to appropriate care.

- Could make this more granular into short-, mid-, and long-term outcomes

Logic Models – Outcomes cont’d

Short-Term Outcomes: (Immediately after the lecture)

- Residents demonstrate increased knowledge about SUDs, screening, and treatment (measured by improved post-test scores).
- Residents feel more confident initiating conversations about substance use.

Medium-Term Outcomes: (Weeks to months after the lecture)

- Residents incorporate routine screening for substance use into patient care.
- Residents initiate brief interventions or referrals for patients with positive screens.
- Increased documentation of substance use screening in clinical notes.

Long-Term Outcomes: (6–12 months and beyond)

- Increased identification of patients with substance use disorders in clinical practice.
- More patients accessing treatment services.
- Potential improvement in patient outcomes related to substance use (e.g., reduced hospital admissions).