

Medical Education Evaluation and Measurement



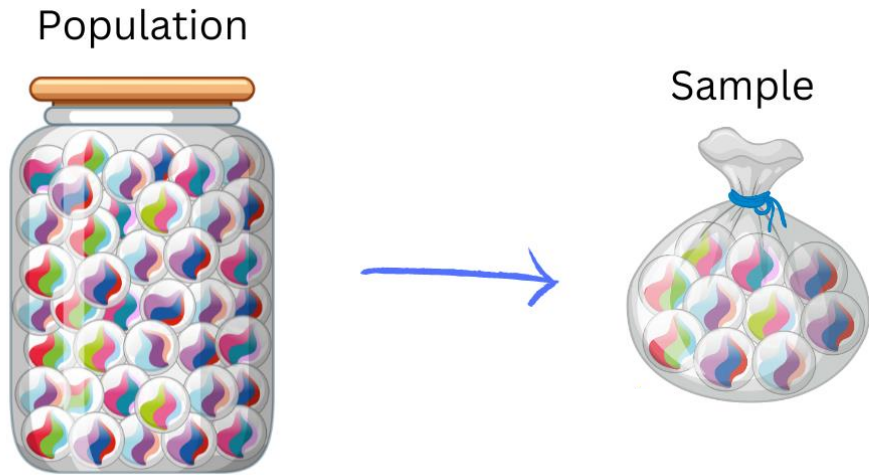
What is evaluation?



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.



What limitations does each example have in terms of generalizability?

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

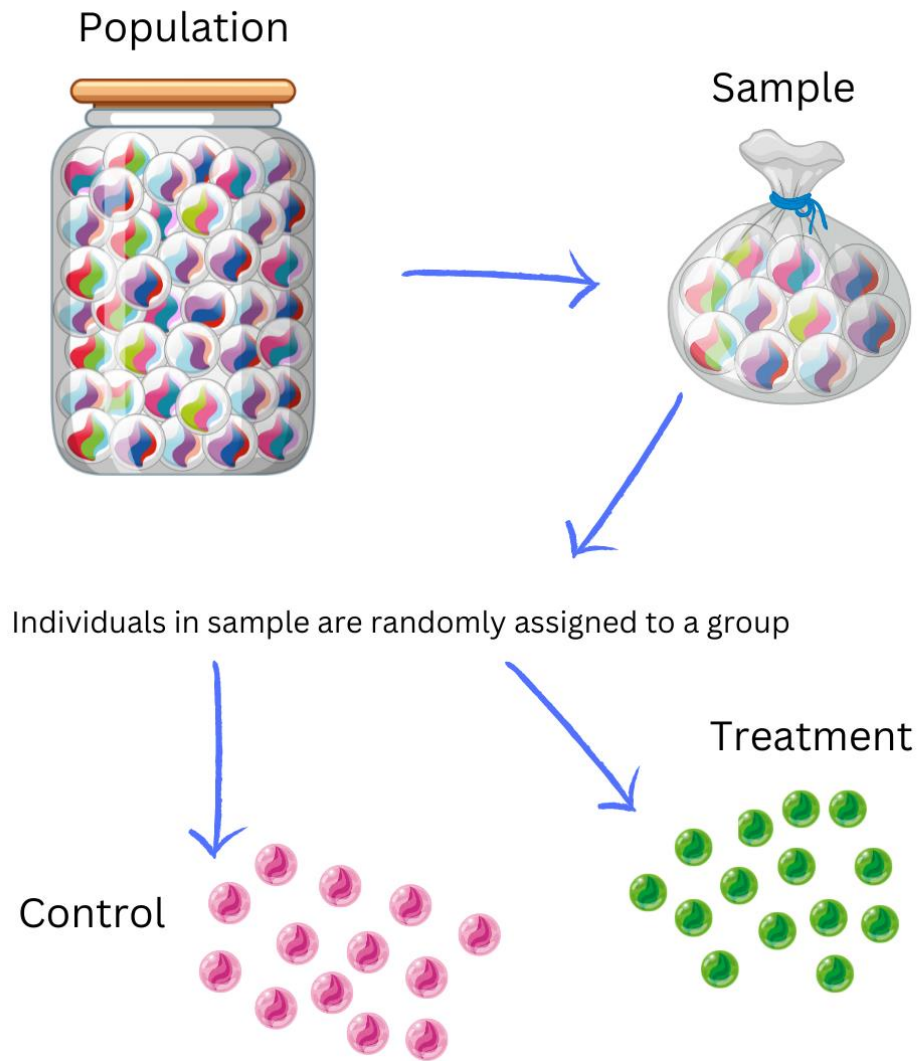
Example: All IM residents in the US

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

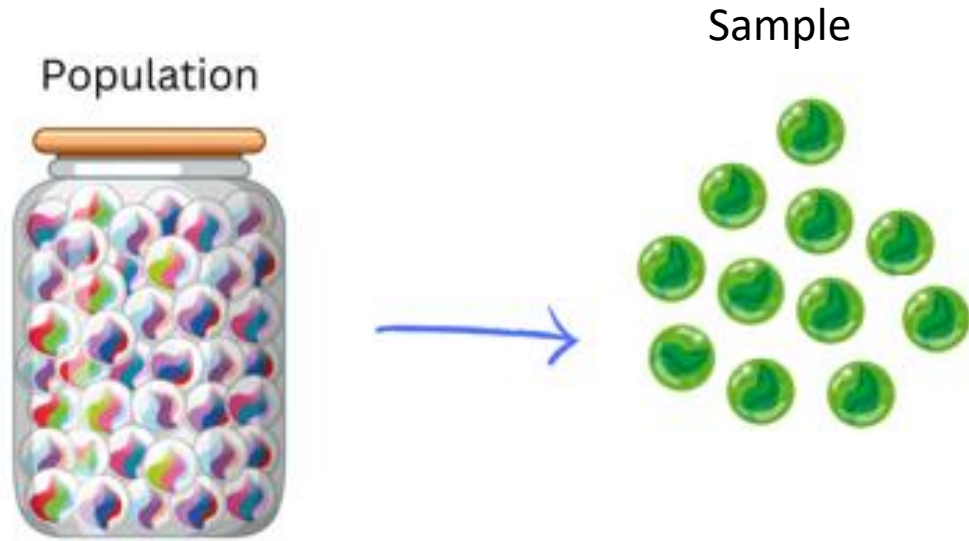
- Resource intensive
- Challenging to execute logistically alongside other education and clinical responsibilities
- Limited sample sizes do not provide sufficient statistical power

Alternative Quantitative Designs (see handout)

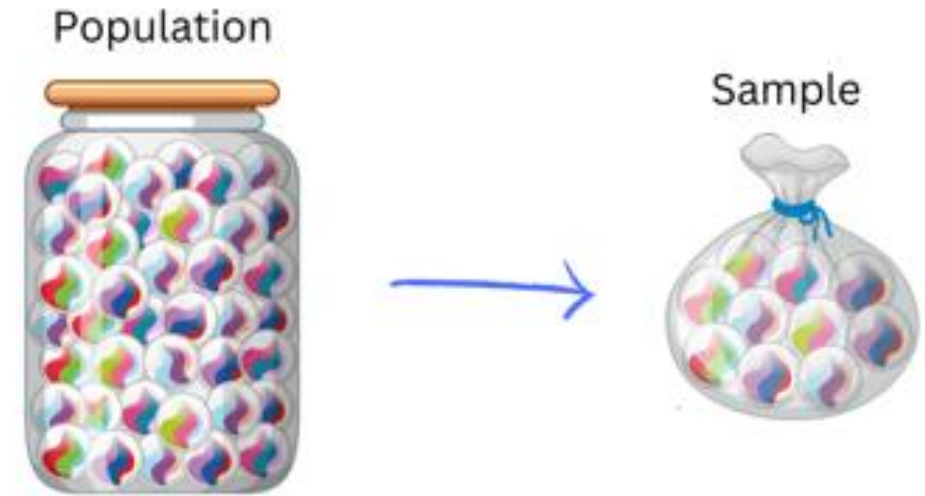
- Quasi-Experimental
- Within-Subjects/Longitudinal

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.



Qualitative sampling is often purposeful, like carefully sifting through the marbles to select specific ones whose unique experiences or characteristics you want to understand.



Random sampling is also used 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 generalizability

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 (see handout)

- 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

Writing Evaluation Questions

Quantitative Questions



Reliability and Validity

Reliability is how consistently you hit the dartboard

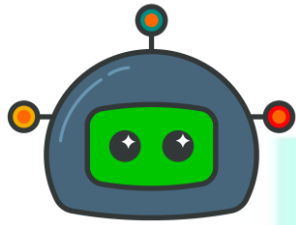


Reliability and Validity

Reliability is how consistently you hit the dartboard

Validity is how close you are to the bullseye





Reliability & Validity

Reliability



Measure	Reliability	Validity
Weight	Very good. Scales may have some element of user error but it is very manageable	Excellent. The measure is exactly as intended: the amount of gravitational force being exerted onto the body
Blood Pressure	Good. Cuffs have some element of user error but can be reduced with training	Good. Measure can be affected by situational factors but generally works as intended,
BMI	Excellent. Assuming no calculation error	Bad. Two people with completely different body compositions can have the same BMI

Qualitative Questions

- **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

Activity

- Draft at least 1 quantitative survey question and 1 qualitative question in AuxEd (10 min)
- Group Share (10 min)
 - What was your experience of writing questions?
 - How do you feel about what you were able to accomplish in 10 minutes?
 - Did you identify any “snakes” in this process?