Evaluation Designs in Medical Education

Melissa McNeil, MD, MPH
Importance of Evaluation

- A good evaluation closes the loop in the curriculum development cycle.
- Helps curriculum developers determine if the goals and objectives of the curriculum were met and provides information for ongoing improvements.
- Can be used to maintain and garner support.
- Serves as a marker of accomplishment for the curriculum developer.
Designing an Evaluation

Helpful to be precise and systematic in designing an evaluation
- Ensures that important questions are answered and relevant needs are met

Objective today is to review a task-oriented approach that helps to ensure a systemic approach to evaluation
Identify Evaluation Users

- First step is to identify the users
- Can include: learners (evaluation of their own performance), faculty (evaluation of their own contributions) and curriculum developers (evaluation of the goals and objectives of the curriculum)
- Don’t forget other stakeholders: administrative personnel (dean’s office, department chair, hospital administrators)
Purpose of the Evaluation

Determine whether the evaluation is to measure the performance of individuals, the performance of an entire group, or both.

Also determine whether the evaluation is to be used for formative purposes (to improve performance) or summative purposed (to judge performance) or both.
Summative Evaluations for Individuals

- Evaluation of an individual learner or faculty member that is used for judgments or decisions about the individual

- Can include:
  - Verification of achievement
  - Motivation of individual to maintain or improve performance
  - Certification of performance for others
  - Grades
  - Promotions
Formative Evaluations for Individuals

Evaluation of an individual learner or faculty member that is used to help the individual improve performance; should include all of the following:

- Identification of areas for improvement
- Specific suggestions for improvement
- Serves as an educational method
Formative Elements for Programs

Evaluation of a curriculum that is used to improve program performance; should include all of the following:

- Identification of areas for improvement
- Specific suggestions for improvement
- Typically takes the form of surveys of learners to obtain feedback
- Quantitative ratings highlight areas that need improvement
- Qualitative feedback provides information in areas that have not been anticipated
Summative Evaluations for Programs

- Evaluation of a program that is used for judgments or decisions about the curriculum developers:
  - Judgments regarding efficacy
  - Motivation/recruitment of learners and faculty
  - Influencing attitudes regarding value of curriculum
  - Satisfying external requirements
  - Prestige, power, influence, promotion
  - Dissemination: presentations/publications
Think About Your Resources

- The most carefully planned evaluation will fail if resources are not available
- Limits in resources may require prioritization of evaluation updates
- Resources include: time, personnel equipment, facilities and funds
- Remember medical schools and residency programs have summative evaluations in place – e.g. board examination
Identify Evaluation Priorities

- Often resources will limit the number of objectives that can be evaluated.
- In that case will need to prioritize and select key evaluation questions.
- Remember that not all evaluation questions need to relate to explicit, written curricular objects – some are implicit.
- Always include open ended questions; allows you to capture unexpected strengths and weaknesses.
Choosing an Evaluation Design

Next need to decide which evaluation designs are most appropriate:

- To answer the evaluation questions
- Most feasible in terms of resources
- To provide *internal validity* (accurate assessment of the impact of the curriculum on the current subjects)
- To provide *external validity* (is it generalizable to other settings?)
Choosing an Evaluation Design

The choice of an evaluation design affects directly the internal validity and indirectly the external validity of the evaluation.

Need to consider factors that could threaten the internal validity of the evaluation which include: history, maturation, testing, instrumentation, selection bias, dropout bias, and statistical regression.
Choosing an Evaluation Design

- **History:** refers to events or other interventions that affect subjects during the period of an evaluation.

- **Maturation:** refers to changes within subjects that occur as a result of the passage of time rather than as a result of discrete external interventions.
Choosing an Evaluation Design

**Instrumentation:** refers to the effects that changes in raters or measurement methods, or that lack of precision in the measurement instrument, might have on obtained measures.

**Selection bias:** occurs when subjects in an intervention or comparison group posses characteristics that affect the results of the evaluation (e.g. affecting the response of the subjects on the intervention).
Choosing an Evaluation Design

- **Dropout bias**: occurs when two groups being compared differ in the incidence of subjects dropping out of the evaluation.

- **Statistical regression**: can occur when subjects have been selected on the basis of extreme scores on tests; because of test taking characteristics, subsequent score are less likely to be as extreme.
Choosing an Evaluation Design

Commonly Used Evaluation Designs:
- Posttest only
- Pretest-posttest
- Nonrandomized controlled pretest-posttest
- Randomized controlled posttest only
- Randomized controlled pretest-posttest

As designs increase in rigor, they also increase in resources needed.
Choosing an Evaluation Design

Evaluation designs are classified as pre-experimental, quasi experimental, and true experimental.

- Pre-experimental designs lack controls.
- Quasi-experimental designs have controls.
- True experimental designs have both controls and randomization.
The Problem

- Third year residents have been recruited to teach medical students physical diagnosis.
- You have been asked to evaluate whether or not this initiative will improve the physical diagnosis skills of the residents.
- How will your structure your evaluation?
The Problem

1) What are the goals of your evaluation?

2) Who are the stakeholders that must be considered in designing your evaluation?

3) What information do you need to consider?
The Problem

- You decide that you will take all of your third year residents and run them through a standardized patient physical examination exercise with a checklist evaluation at the end of their residency.
- This is a posttest design: X----O
- Will this be an adequate evaluation?
Posttest Design

Design permits assessment of what learners have achieved after the educational intervention.

Limitations: achievements could have been present before intervention (selection bias), could have occurred as part of the temporal process of learning (maturation), or could have happened from other interventions (history).

Most appropriate when goal of evaluation is certification of proficiency; allows for assessment of participants perceptions of curriculum.
The Problem

Would there be any benefit to adding an earlier standardized patient examination for each resident?

What additional information would you obtain?

Would this be worth the effort?
Pretest-Posttest Design

Pretest-Posttest Design 01----X----02

- Can demonstrate that changes in proficiency have occurred in learners during the course of the curriculum
- Still could be due to factors other than the curriculum (maturation, history)
The Problem

What could you do to increase your assurance that the physical diagnosis skills of your residents were due to the opportunities to teach medical students physical diagnosis?
Consider Adding a Control Group

Controlled pretest-posttest
EO1-----X-----O2    CO1-----X-----O2

- Quasi-experimental design
- Controls for maturation and history if control group is equivalent
- Complex and resource intensive
- Denies the curriculum to some
The Problem

How can be increase the likelihood that the learners in your two groups (ie those who are teaching medical students vs those who are not) are comparable in their baseline physical diagnosis skills?
Randomized Controlled Posttest

Randomized controlled posttest:
X-----O1    RCX-----O1

- True experimental design
- Controls for maturation, history
- Complex and resource intensive but less so than randomized, controlled pretest-posttest
- Does not demonstrate changes in learners
- Curriculum denied to some
Randomized Pretest-Posttest

Randomized, Controlled, Pretest-Posttest

EO1-----X-----O2     randomized
CO1-----X-----O2

- True experimental design
- Controls for maturation, history, effects of testing
- Most complex and resource intensive
- Curriculum denied to some
Problem Number 2

- Your second year physical diagnosis course for second year medical students has performing write ups of the history and physical examinations that they perform.

- How will you determine if you are meeting your goals?
Address Ethical Concerns

Confidentiality:
- These concerns usually relate to those being evaluated
- Should evaluators have confidentiality (unknown to those being evaluated) or anonymity (unknown to anyone)
- Generally an issue for evaluators in subordinate positions where retaliation is a concern
- Anonymous evaluators may be more open and honest but also may be less responsive in their criticism
Address Ethical Concerns

Access

- Need to determine who has access to evaluations
- Especially important in sensitive areas such as attitudes, interpersonal skills, teaching ability
- Concerns magnified as the reliability and validity of the evaluation instruments become less precise
Address Ethical Concerns

Consent

- Does consent need to be provided for evaluation process?
- If no consent, need to decide how much information about the evaluation process will be shared
  - Strengths and limitations; potential users of the evaluations; uses the evaluations will be put to; confidentiality of evaluation results