Outline

• Introduction
• Technical quality
• The Kirkpatrick model
• Performance measures
• Research methods
• The take away
Introduction

• What are medical simulations?
  • Standardized patients, part-task simulators/anatomic models, computer simulations, and mannequins

• Why use them?
  • Cost and risk avoidance
  • Benefits to learning

• Why evaluate them?
  • Most evaluations have not been of sufficient technical quality to produce trustworthy results
    (Issenberg et al., 2005; Joint Task Force, 2001; Lurie, 2003; McGaghie et al., 2010)
Technical Quality

• Reliability, the consistency of measurement
  • Limitations of classical test theory (Li Cai)
  • Inter-rater reliability
  • The difficulty of designing equivalent scenarios
Technical Quality

• Validity, the degree to which evidence supports interpretations and uses of results
  • A property of the inferences based on results
  • A validity argument, not a statistic
• Sources of evidence
  – Content
  – Response processes
  – Internal structure
  – Relation to other variables
  – Threats to validity
  – Sensitivity to instruction and experience
# The Kirkpatrick Model

<table>
<thead>
<tr>
<th>Level</th>
<th>Evaluation Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Reaction</td>
<td>How did users react to the simulation? Do they value it? Are they motivated by it?</td>
</tr>
<tr>
<td>Level 2: Learning</td>
<td>What knowledge was learned? What skills were developed or improved? What standard of performance was set? What attitudes were changed?</td>
</tr>
<tr>
<td>Level 3: Behavior</td>
<td>Did students change their behavior back on the job? How much transfer of knowledge, skills, and attitudes occurred?</td>
</tr>
<tr>
<td>Level 4: Results</td>
<td>What organizational/patient benefits resulted? How did the quality of proficiency improve due to the system? Was the system cost-effective?</td>
</tr>
</tbody>
</table>
# The Kirkpatrick Model

<table>
<thead>
<tr>
<th>Level</th>
<th>Example Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4: Results</td>
<td>From hospital records: patient satisfaction, patient wait time, adequacy of patient history, time to perform procedure, time in hospital, appropriateness of lab tests and prescriptions, frequency of complications</td>
</tr>
<tr>
<td>Level 3: Behavior</td>
<td>Observer ratings of on-the-job performance using a checklist, completed self-assessment questionnaire, self reports, peer reports, supervisor reports: knowledge, skills, and attitudes</td>
</tr>
<tr>
<td>Level 2: Learning</td>
<td>Test scores, observer ratings of performance using a checklist, embedded measures: knowledge, skills, and attitudes performance was set? What attitudes were changed?</td>
</tr>
<tr>
<td>Level 1: Reaction</td>
<td>Reaction sheets, interviews, focus groups, informal comments: interesting, motivating, accurate, good teaching method, easy to use</td>
</tr>
</tbody>
</table>
Performance Measures

- Outcomes and process
- Checklists and automated scoring
- Must be aligned with intended objectives of the simulation
Performance Measures

• Checklists
  • Used with standardized patients, clinical simulations, and mannequins
  • Objective for clearly observable actions
  • Rater training required
  • Rubrics should appropriately reward different strategies, apply weights for actions of different importance
Performance Measures

- Automated scoring
  - Expert-based methods
  - Data-driven techniques
  - Domain modeling
Research Methods

- Random-assignment experiments
- Quasi-experiments
- Qualitative methods
- Heuristics for selecting methods
Selecting Methods

- What kind of investigation?
  - Project-linked
  - Identify promising practices
- What is the purpose of the evaluation?
  - Technology intervention improvement
  - Determine effectiveness
- Is causal inference desired?
  - Yes
    - Can students, classrooms, or schools be randomly assigned to groups?
  - No
    - Qualitative methods

Start with a quantitative approach to identify successful sites based on some measure, then use a qualitative approach to understand how successful sites are different and identify the practices apparently related to success.
Selecting Methods

Can students, classrooms, or schools be randomly assigned to groups?

- Yes
  - Can feasibility requirements be met?
    - Yes
      - Is information on conditions of applicability or the process producing outcomes required?
        - Yes
          - A random-assignment experiment combined with qualitative methods
        - No
          - Qualitative methods
    - No
      - A random-assignment experiment
- No
  - Feasibility Requirements
    -- The intervention is different from standard practice
    -- The intervention can be maintained over time
    -- Students are not denied access to an entitlement
    -- Human subjects protection requirements can be met
    -- The site will cooperate in implementing conditions
    -- Resources are available: funding, equipment, support
  - Can feasibility requirements be met?
    - Yes
      - Is information on conditions of applicability required?
        - Yes
          - A quasi-experiment combined with qualitative methods
        - No
          - Non-equivalent control groups
    - No
      - A quasi-experiment
  - Qualitative methods

Time series
The Take Away

• Technical quality

• Validity: the degree to which evidence supports the interpretations and uses of results

• A validity argument

• Begin with the objectives of the simulation

• Align measures and methods with the objectives
THANK YOU!

Bill Bewley
voice 310.825.7995
email bewley@cse.ucla.edu

Harry O’Neil
voice 213.740.2366
email honeil@usc.edu