Developing A Career in Simulation Research: Strategies for Young Investigators

Aaron W. Calhoun, MD, FSSH
Associate Professor of Pediatrics
Director, SPARC Program
Chair, SSIH Research Committee
Disclosure

• Aaron W. Calhoun - Co-investigator on Health Resources and Services Administration (HRSA) targeted issues grant
Outline

• Getting Behind the Research Question-Addressing Assumptions
• Pitfalls of Research Design
• Putting it All Together-Developing a Long-term Plan for Investigation
Large Group Discussion-Past Research Question Successes and Barriers
Getting Behind the Question...
Consider the Following...

- Your Question: Does a course in sedation-related adverse events improve teamwork behavior in an MRI suite?
- Sounds like a reasonable place to start...
• In reality, this seemingly simple question rests on an unseen base of assumptions...

• Do we know the specific teamwork behaviors that lead to better outcomes in an MRI suite?

• Do we possess a means to reliably assess them?

• What types/level of evidence are we able to realistically gather?
A Schematic Approach

Do We Know the Ideal Teamwork Behaviors in the MRI Suite?

- Yes
- No

Do We Have a Means to Assess Those Behaviors?

- Yes
- No

Perform Observational and Qualitative Studies to Generate a Model

- Yes
- No

Create and Validate an Assessment Tool

- Yes
- No

Develop an Intervention and Perform a Quantitative Study to Address Question

- Yes
- No
Questions vs Hypotheses

General “Research Question”
  – Informed by experience
  – May or may not be specific or testable
  – No assessment may yet exist
  – Example: “What does unplanned mannequin death do to the learner?”

Hypothesis
  – Focused, definite and predictive
  – Capable of being quantitatively assessed
  – Example: “Can the negative impact of simulated death on learner emotion be mitigated by specifically addressing it during debriefing using an Advocacy-Inquiry approach?”
In the biomedical domain, this is relatively easy

- Relies on basic physiology or pathophysiology

In social sciences/education research it is not as straightforward

- No hard-and-fast “physiology” on which to rely
- In medicine, these types of studies often lack a cohesive approach
- Conceptual models can serve as helpful organizing frameworks

“The theoretical framework is the **structure that can hold or support** a theory of a research study.”

“The theoretical framework **introduces and describes** the theory that explains **why the research problem under study exists.**”

<table>
<thead>
<tr>
<th>Framework:</th>
<th>Hypothesis:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biomedical</strong></td>
<td>“Administration of Doxazosin (alpha 1 blocker) will lower systolic BP by 30 points.”</td>
</tr>
<tr>
<td><img src="image" alt="Diagram of α Receptor Physiology" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychological Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Science/Educational</strong></td>
<td>“Debriefing using the Advocacy/Inquiry method can reduce negative emotions associated with mannequin death.”</td>
</tr>
<tr>
<td><img src="image" alt="Diagram of Psychological Model" /></td>
<td></td>
</tr>
</tbody>
</table>
Conceptual Models

- General Research Question
- Conceptual Framework or Model
  - Qualitative Approach
  - Hypothesis #1
  - Quantitative Approach
  - Hypothesis #2
  - Quantitative Approach
  - Hypothesis #3
  - Quantitative Approach
Elements of a Qualitative Approach

Theoretical sampling
Based on concept
For representativeness & consistency

Data collection
- Observations
- Interview
- Visual material review

Open coding
- Conceptual labeling
- Categorizing

Axial coding
- Finding relationship between categories

Selective coding
- Selection of a core category
- Checking conceptual density

Memos Field notes
Acquire conceptual details

Constant comparison
Between incidents, data and theory
For precision & consistency
To avoid bias

Cho JY, Lee EH. Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. The Qualitative Report 2014 19(64): 1-20
• Validation is hypothesis driven
• Starting point - the decision
• Framework guide the collection of data to support a cohesive validity argument
• For a tool to be used in research such an argument is crucial
• Conceptual frameworks are crucial in building these arguments

Two validity frameworks are currently in wide use.

Each framework weaves together threads of data to generate a validity argument.

**Messick’s Framework:**

**Kane’s Framework:**
1. Examination of the themes/relationships within a specific model
2. Formulation of a specific hypothesis based on that theme or relationship
3. Decide on study design and outcome measures
   1. Kirkpatrick Hierarchy
4. Create protocol and conduct study
Kirkpatrick’s Model

- Evaluation of results (transfer or impact on society)
- Evaluation of behavior (transfer of learning to workplace)
- Evaluation of learning (knowledge or skills acquired)
- Evaluation of reaction (satisfaction or happiness)
Kirkpatrick’s Model

- Extremely difficult data to obtain
- Most successful quantitative educational studies focus on this evidence level
- Extremely difficult data to publish
1. Select a Theme/Relationship

2. Form a Hypothesis
- “Advocacy-Inquiry based debriefing will significantly improve learner emotional state after exposure to simulated death.”

3. Decide on Outcome Measures
- Measure of emotional state on validated scale
- External assessment of emotional state by supervising attending

4. Design of Study
- Intervenotional trial in which learners exposed to simulator death are split into a group that is debriefed using Advocacy-Inquiry (intervention) and a group that is debriefed using another model (control)
- Assessments are conducted on learners before the case, after the case, and after the debriefing
- Supervising attending provides an independent assessment of learner emotional state before and after the session
Small Group Work-Shaping Our Questions
Pitfalls of Study Design
#1-Fishing

- Trawling through a large database testing for every possible association
- The Problem: If you search hard enough you will find something, but it may be meaningless
- Example: If you perform 20 statistical tests, and each is significant with a p-value of 0.05 then, by definition, at least one of those p-values is wrong, but you don’t know which
- Solution: Preselect a small number of planned associations for testing

#2-The Naked P-value

- Looking at p-values without effect size
- The Problem: P-values do not comment on how meaningful a difference is, only the likelihood that you will see it in the population
- Example: A drug that is shown to significantly (p<0.0001) alter life expectancy, but only by 5 days
- Solution: Report an effect size (r-value, odds ratio, etc) if possible

• Means for Likert Scale (1-5) items reported to 100-1000\textsuperscript{th} decimal
• The Problem: Likert Scale data is ordinal and thus averaging can be problematic (especially to this extent)
• Example: “Assessments of learners showed a change from 3.51 to 3.73 after the intervention
• Solution: If you decide to average surveys, do it to the first decimal place at most
• Medians are better for ordinal data
• Doctoring the Y-axis to make a difference between groups look more impressive than it really is
• The Problem: You can make a difference on a graph look as significant as you want it to with enough work
• Example:
• Solution: Use a realistic scale

#5-Power Calculations, Not Trends

- Claiming that a lack of significance is due to lack of power or reporting “trends” that do not achieve significance
- The Problem: This may well be true, but you cannot know this for sure without an accurate power calculation
- Solution:
  1. Perform a power calculation before the study to assess optimal sample size (they are not as complicated as you may think)
  2. Play by your own rules! Do not claim that a result you want to be significant, but does not meet criteria, has a “trend” toward significance. It either meets predefined significance criteria or it does not

#6-Disconnect Between Findings and Conclusions

- Conclusions dramatically overstate results
- Problem: Too much is being read into the data
- Example:
  - **Results**: In a sample of 50 residents, perceptions of personal musculoskeletal exam skills have declined over time
  - **Conclusion**: A national requirement to add a musculoskeletal rotation should be adopted by the ACGME
- Solution: Carefully consider whether a conclusion is justified by the evidence before stating it

#7- The Intervention Always Works

- The Discussion or Limitations sections are used to explain why an intervention that is “obviously” effective showed no actual benefit.
- Problem: The biases of the researchers are overriding the strength of the evidence.
- Solution: Be honest. If your data does not support your hypothesis then state this clearly.
- Recognize that negative studies can lead to interesting new areas of investigation.

• Completing the data gathering phase is only half of the process
• It is essential to submit and have your work accepted in a peer-reviewed journal
• Where do you start?
In Which Journal Should you Publish?

• This is extremely important because it directly affects format of the final article
• Simulation specific journal vs education journal vs specialty journal vs nursing journal
• Examples of journals accepting simulation and educational studies
  – Simulation in Healthcare-**simulation specific**
  – Advances in Simulation-**simulation specific**
  – BMJ Simulation and Technology Enhanced Learning-**simulation specific**
  – Clinical Simulation in Nursing-**simulation specific**
  – Journal of Graduate Medical Education-**resident/fellow specific**
  – Teaching and Learning in Medicine
  – Medical Teacher
How to maximize chances of acceptance

1. Look at example articles from a number of possible journals prior to writing
   - In which journal does your study seem most “at home?”

2. Once you find a possible journal, go to the website and find the “author instructions” page

3. Consider the article types available—how does your study fit best?
   - Empirical Research
   - Review
   - Curriculum Development
   - Conceptual
Types of Articles...

- Choosing the correct article type impacts acceptance
- Questions to consider
  1. How novel is my curriculum/intervention?
  2. How strong is my data?
     - Kirkpatrick level 1 data better suited to conference abstract format
  3. How well does what I have done fit in with more theoretical considerations (educational models, etc)?
Small Group Work- Refining the Plan
Putting It All Together: Developing a Long-Term Plan
Charting a Course

- The value of careful planning
- The value of serendipity
• Often your next research question is directly dependent on the answer to the last

• A “program of research” starts with a series of linked questions
Developing New Research

- Were my results expected or unexpected?
- How do these results impact the overall field of research I am engaged in?
- What new questions within that field arise naturally from these results?
- Do these results point to a different field of interest?
- What question should I investigate next?
Developing The Overall Program

- What **financial resources** (if any) will I need to continue to investigate in this area?
- What **mentoring** (if any) will I need to continue to investigate this area?
- What **opportunities for research dissemination** exist in this area?
- How can this area of research **contribute to advancement** in my career?
Developing The Overall Program

Original Research Domain

- Initial Research Question
- Study Development and Implementation
- Data Analysis
- Next Research Question

New Research Domain

Initial Research Question in New Domain

Etc.
"Well, what the? ... I thought I smelled something."
An Example: Research in Simulation Ethics

• Began with a brief case report of a simulation involving deception and death
• Received concerning reviewer comments regarding the ethics of our approach
• Called a mentor/medical ethicist from my training program for advice
• Indirect source of the review-concerned about many aspects of the work
Given the controversial nature of the case, Editor-in-Chief suggested expanding the piece into an ethical discussion of death and deception in simulation.

Left us with a choice as to how to respond.

We took up the challenge...

“If we pull this off, we’ll eat like kings.”
Deception

Using Simulation to Address Hierarchy-related Errors...

IMSH 2014 Debate: Death and Deception in Simulation

Deception and Simulation Education: Issues...

Phenomenology of Deception Study (under design)

The Importance of Deception in Simulation

Deception in Simulation: A Response...

Death

Initial Case Report

Case and Commentary: using Simulation to Address Hierarchy...

When The Mannequin Dies: Creation...

Confirmatory Study at Northwestern (under review)
The Cost of Innovation

Galileo Galilei  VS  Large Hadron Collider
Bowling Balls vs Large Hadrons

$160 (for a good one)

VS

$9,000,000,000 (for the only one)
The Value of Serendipity

• Sometimes a negative review can be a positive thing
• Often, controversy simply means that no-one has actually studied it yet
• It may not only be a promising field, but there could be a lot of low-hanging, low-cost fruit
• Planning is good, but some of the best things that happen are completely unexpected
• Take advantage of those opportunities
Question and Answer
<table>
<thead>
<tr>
<th>Content Evidence</th>
<th>Does the content of the tool match the construct being measured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Process</td>
<td>What are the specific connections between the questions and the resulting scores? How does this affect the quality of the data?</td>
</tr>
<tr>
<td>Internal Structure</td>
<td>How reproducible are the scores (also known as Reliability)? Are all the items assessing the same construct?</td>
</tr>
<tr>
<td>Relationship to Other Variables</td>
<td>How well do scores derived from the tool match other accepted measures of the same construct (if they exist)?</td>
</tr>
<tr>
<td>Consequences</td>
<td>What will the scores be used for in the “real world” and can that use be supported?</td>
</tr>
</tbody>
</table>
An Alternative Approach: Kane’s Framework

<table>
<thead>
<tr>
<th>Inference Type</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td>What is the specific decision that the tool is intended to facilitate?</td>
</tr>
<tr>
<td>Scoring Inference</td>
<td>What is the strength of the connection between rater observations and assessment tool scores?</td>
</tr>
<tr>
<td>Generalization Inference</td>
<td>What do those scores imply about learner performance in the specific setting of the test?</td>
</tr>
<tr>
<td>Extrapolation Inference</td>
<td>How easily does performance in the test setting translate to the environment of actual practice (i.e. the “real world”)?</td>
</tr>
<tr>
<td>Implication Inference</td>
<td>Given the prior three inferences, what do the scores then imply about the learner’s ability with regard to the decision in question?</td>
</tr>
</tbody>
</table>

Making the Validity Argument Using Messick’s Framework

Weaker Argument

Stronger Argument

Content

Response Process

Internal Structure

Rel. to other Variables

Consequence

Tool #1

Tool #2
Making the Validity Argument Using Kane’s Framework

State the proposed decision to be made with the tool

State the hypothesis/argument to be made regarding the tool

Test the weakest assumptions in the argument using the four inferences
  • Scoring
  • Generalization
  • Extrapolation
  • Implications

Evaluate Evidence

Evidence unfavorable, tool requires revision

Reject Argument

Evidence favorable, but more is needed...

Accept Argument