‘Gold standard’ versus research in practice: Practical examples from educational research

Victoria Cockle
Team Leader Research: The Starpath Project
Researcher: The Evaluation of the Implementation of Manaiakalani Outreach

Faculty of Education, The University of Auckland
Considerations for research design (in educational context)

An example of what 'gold standard' research design might look like

Implications of that design in practice

Timing issues and other confounds

Implementation measures and limitations
  
  Self reporting measures
    
    Interviews
  
    Surveys
  
    Questionnaires

Objective(?!?) reporting measures
  
  Observations
  
  Artefact analysis

Analysis issues
General Considerations

- Matching research question(s) to research design
- Sample size and sampling bias
- Use of measures that are fit-for-purpose
  - Qualitative/Quantitative/Mixed methods
  - Balancing applicability of tested measures versus creating new ones
- Practical constraints -
  - Balancing your time, budget, organisational issues
    - Which questions are more important than others?
    - Which measures are more important than others?
  - Planning your analyses prior to data collection
What might ‘gold standard’ research look like in an educational setting?

- Most educational research concerned with evaluating school interventions/processes

- Ideal methods (from a ‘gold standard’ perspectives would involve:
  - Identify group of interest (who the intervention is aimed at targeting)
  - Take measurement of interest for all individuals (i.e., writing test)
  - Randomly assign individuals to 3 groups (one treatment, one ‘placebo’, one ‘control’)
  - Begin ‘treatments’ (i.e., groups 1 and 2) at same time
  - Carefully collect measurements throughout intervention ensuring these are standardised (i.e., comparable across time points and/or across schools)
    - Achievement data
    - Implementation measures
      - Qualitative
      - Quantitative
  - Assess differences between groups at end of treatment
What might this design look like in practice?

- Ethics
  - Issues to do with selecting students - if you have a method/intervention that works, is it ethical to restrict student’s access to it?

- Schools often choose a whole-school approach - i.e., school-wide intervention OR work with a target group only but include all members of the target group

- Implications - No “true” control group
  - Alternative: Use other schools with similar characteristics as matched comparison
    - Difficulties in matching, lack of available data - competitive nature of schools
  - Alternative: Use school’s own baseline comparisons
    - Assumption that cohorts have not changed significantly over time
    - Assumption that school systems/teaching has not changed significantly over time
    - Assumption that local/social/governmental issues have not changed significantly over time (e.g., housing crisis)
Timing issues and other confounds

- Issues of beginning ‘treatment’ at the same time
  - Possible within schools (i.e., if intervention is only in one school) but almost never happens across schools
  - If project and evaluation are being not run by same people (ideal)

- Issues of collecting data at the same time (especially if these rely on researcher data collection - e.g., observations/interviews)
  - How much does time matter? Does a 2-month lag matter?

- Other confounds
  - Different teachers and teaching styles
  - Different school structures/systems/foci

- Bias
  - Buy-in of participants => lag
  - People that agree to participate may have an agenda
Implementation Measures and limitations - SRMs

- Self reporting measures - might include interviews, surveys, questionnaires

- All SRMs - Good for finding out peoples’ perceptions (less useful for finding out what is actually happening)
  - Accessing participants and gaining consent is always an issue in practice
  - Volunteers - agenda of those that agree to participate a bigger issue in SRMs

- All SRMs - Wording of questions (open/closed; biased agenda vs blank agenda)

- Interviews:
  - Place and timing of interview (e.g. McDonalds), selection of interviewer

- Surveys/Questionnaires:
  - How to ‘give it’ to participants, ensure adequate number and representative responses? - Prioritising of measures? Being there?
  - Question complexity & length
  - Scales?
Implementation Measures and Limitations - Observations/Artefact Analysis

- Better at finding out ‘what’s actually happening’ (?)
- Well-designed tools allow for mixed qualitative/quantitative data collection
- Changing tools/methods on the fly? Time for pilot run?
- Observer bias - lots of moderation and training required
- Observed bias
  - Video recordings - technical constraints, cost, time
  - Peer observers - still have observer bias
- Artefact analysis
  - Consistency of documentation over time/schools/contexts
  - Analysis framework - open vs axial coding
Example of a good observation tool

<table>
<thead>
<tr>
<th><strong>Teacher group</strong></th>
<th><strong>Other group</strong></th>
<th># on devices:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Nature of text used</td>
<td>Learning Manage ment</td>
</tr>
<tr>
<td>Print – general</td>
<td>Teacher created</td>
<td>Totally digitally managed</td>
</tr>
<tr>
<td>Print – student</td>
<td>Student created</td>
<td>Online + verbal prompts</td>
</tr>
<tr>
<td>Digital – general</td>
<td>L/Model</td>
<td>Offline</td>
</tr>
<tr>
<td>Digital – student</td>
<td>Q&amp;A</td>
<td>Totally digitally managed</td>
</tr>
<tr>
<td>Q&amp;A</td>
<td>APK</td>
<td>Online</td>
</tr>
<tr>
<td>Online (Y/N)</td>
<td>Extended reading (m)</td>
<td>Extended writing</td>
</tr>
<tr>
<td></td>
<td>Information site</td>
<td>Email</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Task</td>
</tr>
<tr>
<td>Group activity</td>
<td>Practice</td>
<td>Order of tasks</td>
</tr>
<tr>
<td></td>
<td>Strategy</td>
<td>My work</td>
</tr>
<tr>
<td></td>
<td>Example</td>
<td>Tool</td>
</tr>
<tr>
<td></td>
<td>Critical</td>
<td>Other</td>
</tr>
</tbody>
</table>

- **Nature of task (tick):**
  - Game based
  - Game / Program
  - Constrained practice
  - Class site
  - Extended reading (single)
  - Extended reading (multiple)
  - Discussion (or one person on computer only) (FTF)

- **Nature of site (tick):**
  - Digital – student
  - Digital – general
  - Online (Y/N)
  - Information site
  - Google drive
  - Language
  - Other: e.g. off-task
  - Other: topic

- **Agency:**
  - Evidence of student decision

- **Working Together (tick):**
  - Both using computer (CMD)
  - Discussion (or one person on computer only) (FTF)

- **Learning Management:**
  - Online + verbal prompts
Analysis issues - So many assumptions!

- **Quantitative methods:**
  - Summary statistics/data visualisation always provides the biggest clue to changes in achievement
  - Options for no controls: Matched/Baseline Comparisons (next slides) allow researchers to determine likely shifts in achievement relative to expected
  - Hierarchical linear models/regressions - usually allow for only correlations

- **Qualitative methods**
  - Coding - open versus axial coding -> moderation and theoretical perspectives

- **All analyses**
  - What data you actually get
  - To use it or not to use it
  - ‘Incidental’ findings
Matched Comparison Example: Difference in Difference
Baseline Comparison Examples:

Average Success Rates Before and After Intervention

Average Roll-Based Success Rate (%)

No of years before and after implementation

Level 1
Level 2
Level 3
UE
Baseline Comparison Examples
Thanks for listening....

Victoria Cockle

v.cockle@auckland.ac.nz
021 023 10003

Team Leader Research: The Starpath Project
Researcher: The Evaluation of the Implementation of Manaiakalani Outreach
Faculty of Education, The University of Auckland