3 Questions You Should **Never** Ask in Evaluation/ Assessment: …And a few you should!

**Presenters (in order of appearance)**

Chris Bourg, Stanford University, mchris@stanford.edu
Joshua Morrill, Morrill Solutions Research (MSR), joshua@morrillsolutions.com
Glenda Morgan, University of Illinois Urbana, gmorgan@illinois.edu
Question #1

The ambiguous question
Ambiguity in ...

- Wording
- Meaning
- Threshold
Double-barreled questions

• “Buy-back programs are a good idea for decreasing gun ownership.”

• “Within the next 5 years, the use of e-books will be so prevalent among faculty & students that it will not be necessary to maintain library collections of hard copy books”
Ambiguous meaning
Less ambiguous

Acknowledgements by Year, Stanford v. Penn State

Number of acknowledgements

Year


Stanford
Penn State
Ambiguous thresholds

- How much is enough?
- Fake data is your friend
Significance-Shminificance

...Is it meaningful?
Overpowered Tests

- Faculty use of digital resources (4400+ responses!)
  - First order tests, EVERYTHING IS SIGNIFICANT! (= Slow…)
  - Implications for Learning Analytics

If the phenomena we are trying to measure is the earthquake

The size of our sample determines if our seismograph is more or less sensitive.

Too FEW cases and we can only pick up on the BIG earthquakes.

Too MANY cases and every minor bump registers as an earthquake.

GOLD NUGGET

We are smarter than our tools. We can decide if we sound the alarm or not, but we need to make that decision mindfully!
False Precision

- False precision (The bane of 1-10 scales)

Is there REALLY a difference between 8 & 9s? 6s and 7s?

There is error in every measurement! **More scale precision does not necessarily mean more accuracy.**

**Good researchers are empathetic!** How are the people who are participating in your research seeing/understanding the process?
Lots of things influence the statistical significance / non-significance of a finding

**Chiefly…**
- *The Sample Size*
- *The Effect Size*

**In most evaluations** these are NOT optimum for significance testing…

**But --- EVEN WITHOUT SIGNIFICANCE ---** evaluation results can still be important and meaningful.
For Example…An Underpowered Test

“The library provides a vital service to campus.”

<table>
<thead>
<tr>
<th>Group</th>
<th>% Agree &amp; Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty (n=12)</td>
<td>83%</td>
</tr>
<tr>
<td>1st Year (n=25)</td>
<td>88%</td>
</tr>
<tr>
<td>2nd Year (n=22)</td>
<td>85%</td>
</tr>
<tr>
<td>3rd Year (n=4)</td>
<td>25%</td>
</tr>
<tr>
<td>4th Year (n=6)</td>
<td>66%</td>
</tr>
</tbody>
</table>

Something going on here? … maybe but not statistically…

At its best, research is a small flashlight in a dark room…

Data-driven decision making ≠ relinquishing personal judgment.
Approach --especially underpowered data-- with these questions

Is this data suggesting something important for my core business or service?

Do I need more information in order to take action?

What --if anything--do I need to know in order to take meaningful action?
Questions you SHOULDN’T ask?

...Probably about half of them
The Length vs. Empathy Relationship

Research Length / Time Investment for participants

- Short / Low
- Long

Empathy for / with Participants

- Low
- High
Do you **REALLY** need to know that?
Problems with too much data?
The (often) false allure of longitudinal data
A Few Suggested Readings

--From Josh--
How to Measure Anything by D. Hubbard

--From Chris--
Evaluation: A systematic Approach by Rossi, Lipsey & Freeman

--From Glenda--
Improving Survey Questions by F. Fowler -&- Research Methods in Anthropology by R. Bernard
Questions for us?

Thank you!

Presenters  (in order of appearance)

Chris Bourg, Stanford University, mchris@stanford.edu
Joshua Morrill, Morrill Solutions Research (MSR), joshua@morrillsolutions.com
Glenda Morgan, University of Illinois Urbana, gmorgan@illinois.edu