Introducing think-aloud interviews as a tool to explore student statistical reasoning

Mikaela Meyer and Ciaran Evans
Carnegie Mellon University

http://www.stat.cmu.edu/teachstat/

Teaching Statistics Research Group @ CMU

Statistical Misconceptions

- Users of statistics often misunderstand or misuse statistical tools
  - Confusion about confidence intervals and error bars
  - p-hacking
  - Data manipulation
- Undergraduate classes are our first chance to correct misconceptions
- Some known misconceptions in undergrad stats:
  - Confusion about probability
  - Confusion about sampling distributions
  - Confusion about p-values
  - Etc…
- Want to understand how students think, not just what they get wrong

Investigating Misconceptions

How can we research what students do and don't understand? We started by talking to them, in over 40 hour-long think-aloud interviews.

- Interviews were conducted in a private room with student volunteers

Here's some student responses to a question about sampling distributions:

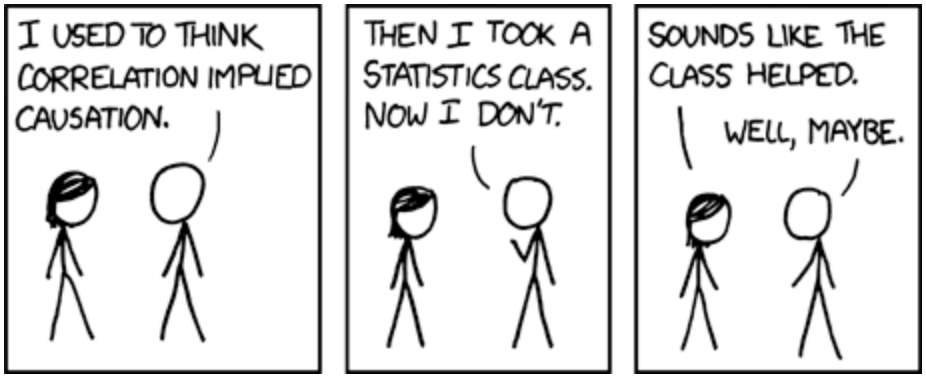
- "[The] population should be normally distributed"
- "Small *n* means few bars [in the histogram]"
- "I wasn't thinking about the average, more about the distribution of the data"
- "True sample size for [the] population should approach normal"

Why Think-Alouds?

- Interviewees read the question aloud and narrate their thinking
- No feedback from the interviewer (verbal reactions, facial expressions, ...)
- Student thought process differs when *doing* a problem vs. *explaining* their solution
- Developed in cognitive science by Ericsson and Simon
- Benefits:
  - Better understand *why* students get answer right or wrong - answer alone might not tell you what they think
  - Hear how confident they are in their responses
  - Better understand clarity of questions - we also use results to develop better questions

Misconception: Correlation vs. Causation

The difference between correlation and causation is important, and often stressed in introductory classes:



But what do students actually think about the difference between the two?

Accessed 8/15/19, https://xkcd.com/552/

Some Possibilities

- Ignorance that correlation may not imply causation
- Belief that correlation does not imply causation even in circumstances when causal conclusions actually can be drawn
- Knowledge of the phrase “correlation does not imply causation,” but inability to recognize causal language that *doesn't* use keywords like “causation”
- Incomplete understanding of why randomization is useful (or incomplete distinction between random sampling and random assignment)

We wrote seven questions on various aspects of correlation, relationships, and randomization

Example Questions

Q: A clinical trial randomly assigned subjects to either practice mindfulness meditation or a placebo relaxation exercise as a treatment for a cold. The trial found that subjects who practiced mindfulness meditation had a shorter time to recovery than students assigned to the relaxation exercise, and the result was statistically significant. Which conclusion does this support?

Q: A survey of Californians found a statistically significant positive correlation between number of books read and nearsightedness. Which of the following can we conclude about Californians?

Think-Aloud Results

- For questions in which causal conclusions cannot be drawn, almost all students got correct answer
- Many students were unwilling to draw causal conclusions even when they were justifiable
- Here’s what some of the students said:
  - “‘Correlation does not imply causation’ is a universal rule”
  - “When can we ever say something causes something else?”
  - “I think the word ‘causes’ is too strong... my friend who's a stats major always tells me you can't say this causes that---there's always other factors”
  - “Usually [you] can't assume causation”

Assessment Results

We also gave some of these questions in an assessment to hundreds of students in introductory statistics classes:

A survey of Californians found a statistically significant positive correlation between number of books read and nearsightedness. Which of the following can we conclude about Californians?

- 87% of students correctly selected “We cannot determine which factor causes the other, because correlation does not imply causation”

Assessment Results

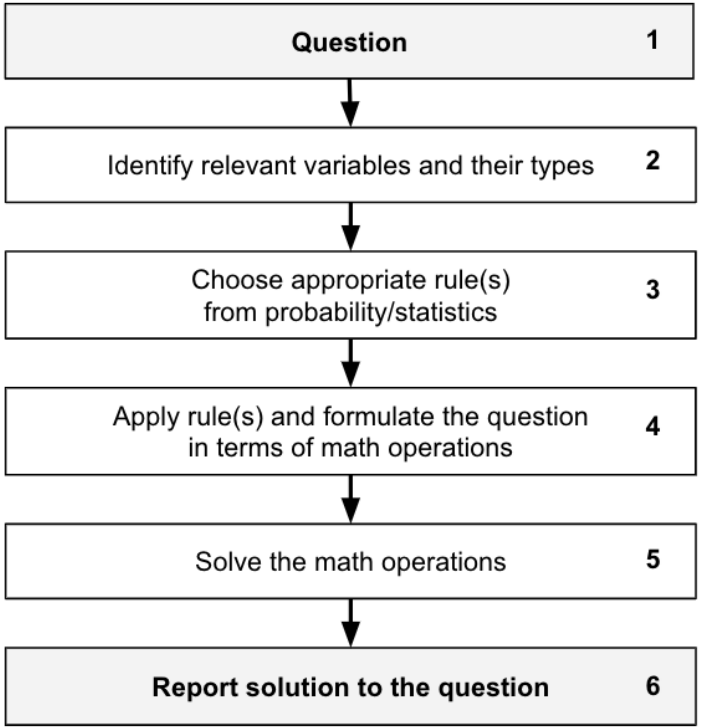
A clinical trial randomly assigned subjects to either practice mindfulness meditation or a placebo relaxation exercise as a treatment for a cold. The trial found that subjects who practiced mindfulness meditation had a shorter time to recovery than students assigned to the relaxation exercise, and the result was statistically significant. Which conclusion does this support?

- Only 35% of students got correct answer
- Over 50% selected “We cannot draw any conclusions because correlation does not imply causation”

Think-aloud interviews can be combined with cognitive task analysis (CTA)

- Method from cognitive science research used to detail the steps used to solve a problem

Cognitive Task Analysis (CTA) for introductory statistical inference problems



Misconception: Variables, Parameters, and Math

When asked to find MLE of the mean of a univariate normal distribution:

What experts said…	**What novices said…**
● “We use the log likelihood because it is a, it is a one-to-one function” ● “So it’s gonna be the mean, but let’s prove it.” ● “And just to check that is a maximum, you take the second derivative and check that it is... hmm, check that it is negative, so that it is a maximum”	● “I always get weirded out when I have to do the derivative of a sum, like I don't really know if there's rules…” ● “So we just take the derivative of this with respect to... what do you call it, sigma, right? Yeah, yeah, so sigma. Or is [it] with respect to sigma, or with respect to mu?”

Writing Questions and Assessments

Problem: Think-aloud interviews are hard to scale

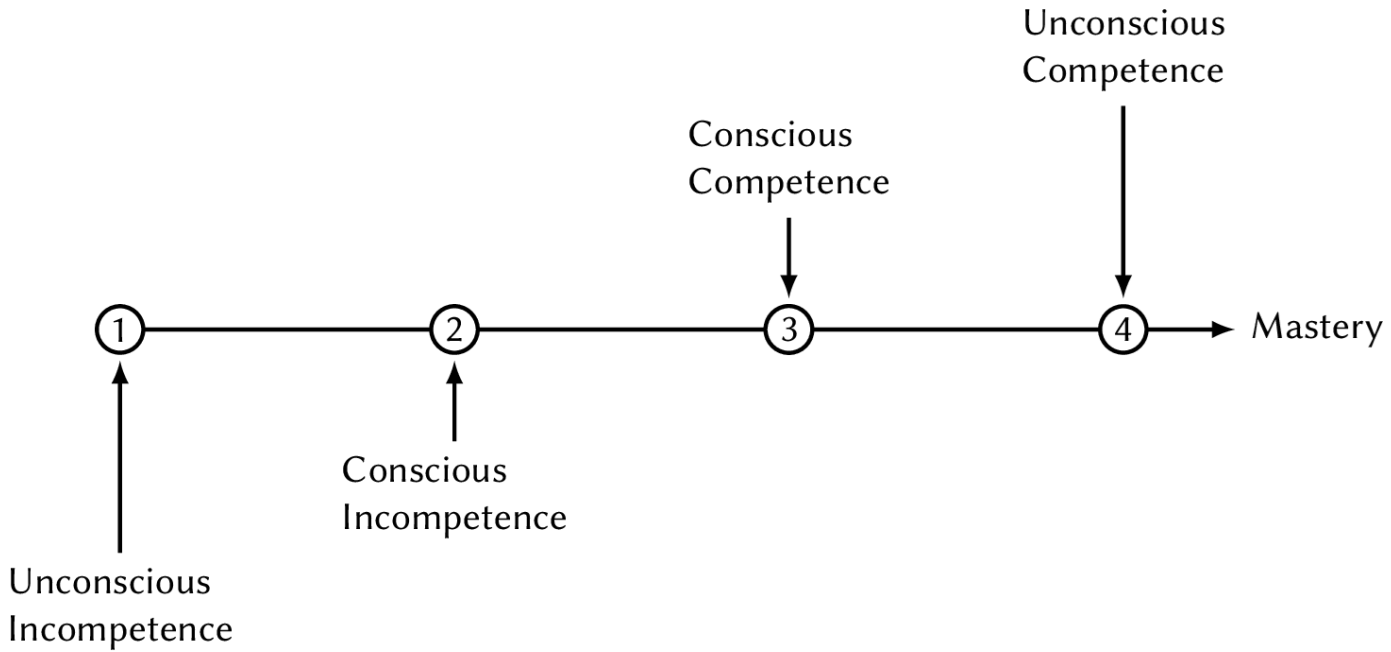
Solution: Make assessment from think-aloud questions, give to many students

Problem: Good assessment questions can be hard to write

Solution: Use feedback from think-alouds to revise and improve questions

Why is it hard to write questions?

Experts assess student learning, but they don't think like students



(Sprague & Stuart, *The Speaker's Handbook*, 2002)
(Ambrose et al., *How Learning Works: 7 Research-Based Principles for Smart Teaching*, 2010)

An unexpectedly confusing question:

Two draws are made at random from the box containing

1 2 3 4

After taking out the first draw, you lose it, and nobody knows what was written on it. You draw a second time. Are the two draws independent?

A. The draws are independent
B. The draws are dependent
C. Not enough information to tell

(Freedman, Pisani, and Purves, *Statistics*, 1978)

A revision:

Two draws are made at random from the box containing

1 2 3 4

After taking out the first draw, a **duck eats it**, and nobody knows what was written on it. You draw a second time. Are the two draws independent?

A. The draws are independent
B. The draws are dependent
C. Not enough information to tell

Revising Questions

1. Some questions have small issues that cause irrelevant misunderstandings

   Solution: Simple edits

2. Students have multiple misconceptions about a single question

   Solution: Split into multiple parts/write new questions

3. Students sometimes get right answer for wrong reason

   Solution: depends on situation

Think-alouds are challenging to implement, but worth the trouble

- Time constraints, interviewer training, getting students to participate, etc.
- Think-alouds are useful tools to learn about learning
  - They provide direct access to student thinking
  - Think-aloud results can help guide instruction
- Questions will be available to anyone who asks
  - Looking for research partners to collect data and participate in future studies
- Future directions: think-aloud interviews for data analysis, other courses, etc.

Thank you!

If you would like to read more about our work, see our recent paper on arXiv:

Reinhart et al. “Think-aloud interviews: A tool for exploring student statistical reasoning”. Preprint on arXiv https://arxiv.org/abs/1911.00535

If you want to read more about our other projects, check out our website:

http://stat.cmu.edu/teachstat/

If you want to contact us, reach out via email:

mrmeyer@andrew.cmu.edu, clevans@andrew.cmu.edu, and/or areinhar@stat.cmu.edu