Developing an assessment for concepts in introductory statistics and data science

P Burckhardt, P W Elliott, C Evans, S Hyun, K Lin, A Luby, C P Makris, M Meyer, J Orellana,* R Yurko, G Weinberg, J Wieczorek,† R Nugent & A Reinhart

Dept. of Statistics & Data Science, *Center for the Neural Basis of Cognition, Carnegie Mellon University, and †Dept. of Mathematics and Statistics, Colby College

Overview

- We are developing an assessment of introductory statistics concepts
- Think-aloud interviews with students helped us uncover new misconceptions and improve assessment questions
- We administered the assessment to ≈ 200 students at two different institutions
- **Goal**: Provide ways to measure student learning, so we can conduct more pedagogical research

Assessing Learning in Intro Stats

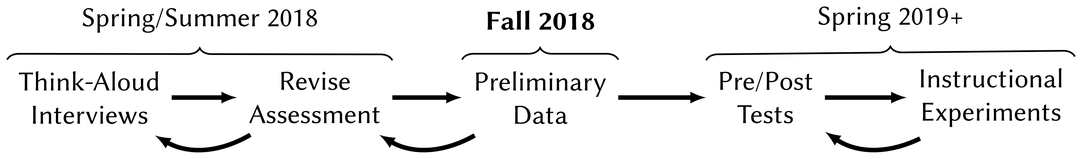
- To improve teaching, need to assess what the students are learning
- Must avoid ambiguous questions and confirm that questions actually measure student learning, not just test-taking skills
- Used think-alouds: have students think aloud while answering draft assessment questions (see Adams and Wieman 2011, Burckhardt et al. 2017)
- Think-alouds elicit misconceptions and misreadings, and help us revise and write new questions

Think-Aloud Results

- Conducted 36 interviews, each ≈ 1 hour long, in rounds timed to topics introduced in 36-200
- Interviews tested roughly 50 draft questions
- Used student feedback to revise questions, then re-tested in later think-aloud rounds

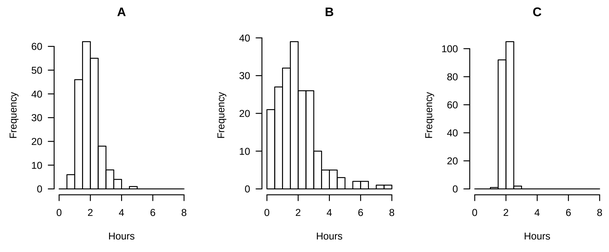
Data Collection

- After think-alouds, built revised assessment
- Students also asked to rate their confidence in each answer
- Used ISLE to administer the assessment to 95 students in 36-202 in Fall 2018; results presented here
- Administered paper version to 117 introductory students at Colby College, in 5 course sections taught by 3 instructors



A Think-Aloud Surprise

Original Q: "Pictured (in scrambled order) are three histograms: One of them represents the population distribution of study hours; the other two are sampling distributions of the mean $\bar{X}$, one for sample size $n = 5$, and one for sample size for $n = 50$."



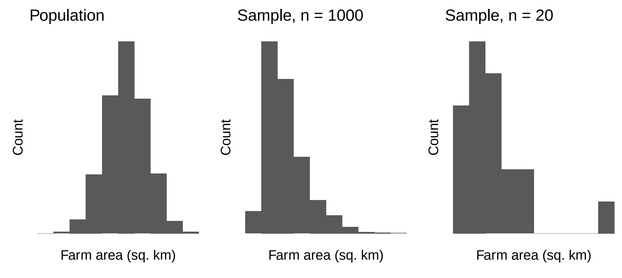
- All 9 interviewees wrongly thought the sampling distribution with $n = 5$ was C
- "Small n means few bars"—didn't think about sampling distributions at all
- Some also thought "the population should be normally distributed"
- Didn't expect this misconception—so we split the question in two

Better Targeting with Revised Questions

Revised Q: "Steve talks to two hundred **groups of 5 students**. After asking each group of 5 students how much they study, Steve takes the **group's average** and adds it to his histogram. *[...]*"

Students got this version right. Also added a new question about the distribution of samples taken from populations:

New Q: "Farmer Brown collects data on the land area of farms in the US. *[...]* She then takes two random samples from the population, of sizes $n = 1000$ and $n = 20$, and plots histograms of the values in each sample. One of the rows shows three histograms. Using the **shape** of the histograms, choose the correct row."

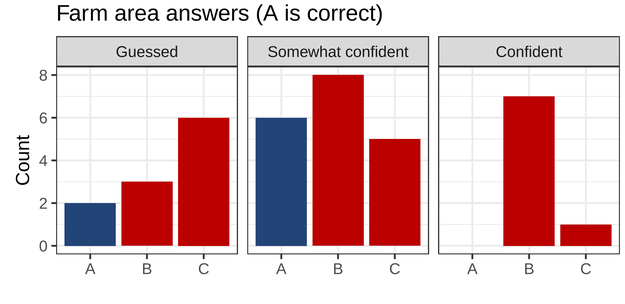


In think-alouds, about half of students wrongly chose B (shown) and stated that populations should be normally distributed.

A Misconception Revealed

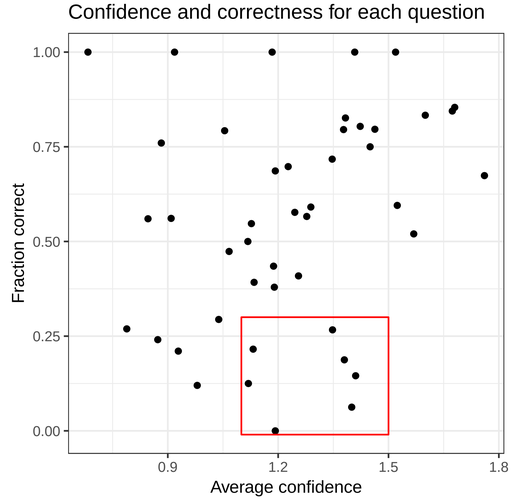
Self-reported confidence can be revealing:

- Wrong answers aren't always guesses
- Confident mistakes are likely misconceptions
- Data from 36-202 shows that confident students were more likely to pick the *incorrect* answer



Spotting False Confidence

Misconceptions can be identified by finding questions students get wrong while confident (in red):



Next Steps

- Student beliefs are surprising: to build a good assessment, think-alouds are invaluable
- We recommend this process for anyone writing assessments in *any* field
- Will collect pre/post data in 36-200 in Spring 2019 to assess student learning
- Will survey instructors to get their input on the assessment topics and questions
- Results will guide new pedagogical experiments
- The validated test can be used to assess learning and aid redesign for new Dietrich General Education curriculum

References

Adams and Wieman (2011), *Int. J. Sci. Educ.* **33** 1289–1312.

Burckhardt et al. (2017), *Teaching & Learning Summit.*

Hyun et al. (2018), *eCOTS.*

Acknowledgments

Thanks to the Eberly Center for advice, to the Department of Statistics & Data Science for financial support, and to all the students who participated in interviews and assessments.