Identifying misconceptions of introductory data science students using a think-aloud protocol

Philipp Burckhardt, Peter Elliott, Ciaran Evans, Sangwon Hyun, Kevin Lin, Amanda Luby, Christopher Peter Makris, Josue Orellana, Alex Reinhart, Jerzy Wieczorek, Ronald Yurko, Gordon Weinberg & Rebecca Nugent

Department of Statistics & Data Science
Carnegie Mellon University

Hi, my name is Alex Reinhart, and I'll be talking about our project to develop assessments and identify misconceptions in intro data science students by using think-aloud protocols.

Background and motivation

Goal: Better assess conceptual understanding for intro stats/data science students

Method: Think-aloud interviews

- Provide insight into *why* students answered the way they did
- Check whether question tests what we think it tests
- 33 interviews conducted during the Spring 2018 semester, in three rounds corresponding to topics covered in class
- Interviews proved invaluable to revise and validate questions
- Think-aloud protocol reviewed by CMU IRB

(For more on think-alouds, see Adams and Wieman, 2011)

Our department at CMU is redesigning its intro courses to be more modern and to reach a wider range of students, but to do this effectively, we need a way of testing if the new course is working, and diagnosing any problems that might occur.

So, at the same time, we have been working on a new conceptual assessment for the core statistics and data science concepts we think are most important for introductory students. To validate the assessment, we've been using think-aloud interviews, a method which has been previously used in other scientific fields to validate concept inventories.

A think-aloud interview asks a student to answer draft questions while thinking aloud. Because they are thinking aloud, the interview provides insight into *why* students answered the way they did, and checks that the question tests what we think it tests -- that students are not misreading questions or picking answers based on their wording or irrelevant details.

We conducted 33 interviews throughout Spring 2018, in waves corresponding to the topics covered in the course, so students were interviewed on each topic soon after learning it.

Our experience has shown that think-alouds are invaluable to understanding how students think about core statistical concepts, and it would be nearly impossible for us to develop a good assessment without them. We recommend think-alouds as a general tool for understanding student reasoning.

Prior concept inventories

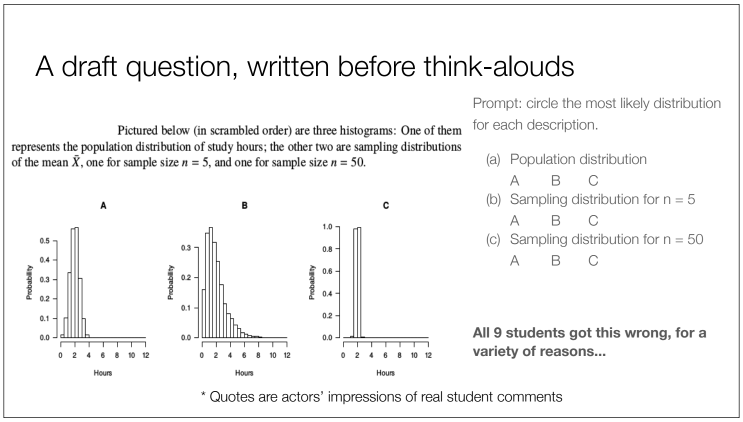
- There are several prior concept inventories in statistics, like the Comprehensive Assessment of Outcomes in Statistics (CAOS)
- CAOS was revised with instructor input and validated by a group of 18 expert raters
- All raters agreed that "CAOS measures basic outcomes in statistical literacy and reasoning that are appropriate for a first course in statistics"
- But think-aloud interviews were not used to ensure the questions really elicited the intended student reasoning

(delMas, R., Garfield, J., Ooms, A., & Chance, B., 2007)

There are, of course, previous statistical concept inventories, most prominently the Comprehensive Assessment of Outcomes in Statistics, or CAOS. CAOS was developed and revised with extensive input from statistics instructors, and validated by a group of 18 expert raters, who all agreed that it measures "basic outcomes in statistical literacy and reasoning that are appropriate for a first course in statistics."

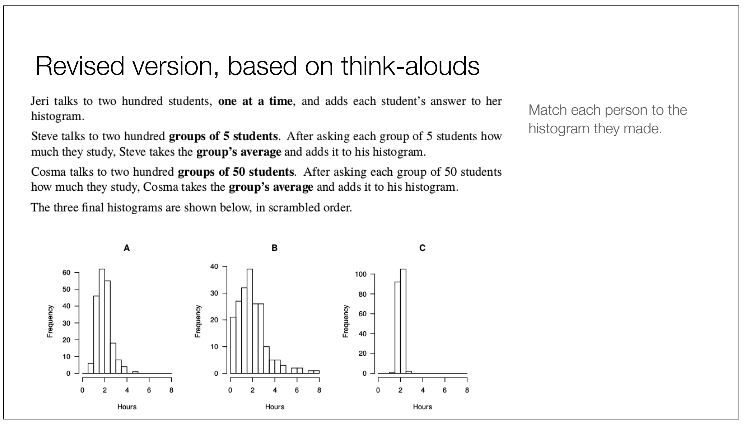
But think-aloud interviews were not used to ensure the questions really elicited the intended student reasoning, so we can't be sure CAOS tested the outcomes it was intended to test.

Let me show some examples to demonstrate how think-alouds helped us revise our questions and make them test the intended concepts, illustrated with real quotes from student interviews. Feel free to pause to read each question.



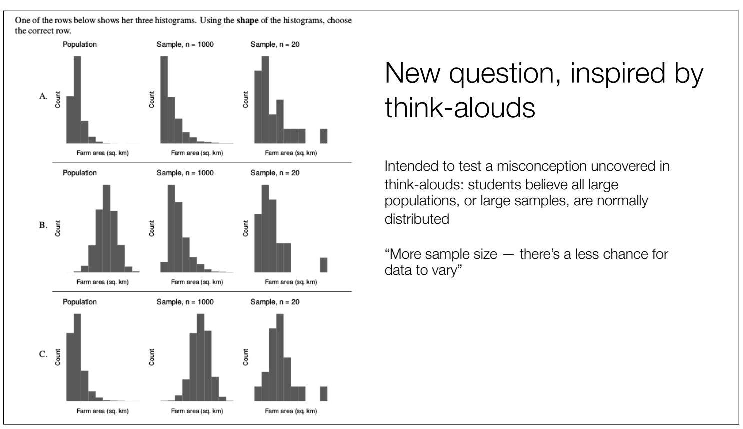
This question asks students to determine which histogram is the population and which are the sampling distributions of the mean, with either n = 5 or n = 50. To our surprise, all nine students interviewed got the question wrong -- they thought the sampling distribution with n = 5 should be graph C, since "small n means few bars". Two students said the population distribution should be graph A, since "the population should be normally distributed".

In follow-up questioning, students revealed that "I wasn't thinking about the average, more about the distribution of the data". This proved the question didn't work! None of the students expressed *any* reasoning about the decrease in variance as we average over more data, because they focused on irrelevant details of the question. We can't use this question to assess if students understand sampling distributions.



To fix this problem, we dramatically revised the question based on the interview results. In this version, the phrase "sampling distribution" never appears. Instead, we describe the situation mechanistically.

Students understood this version better and did correctly reason about averages, but the question took them quite a long time to work through – three and a half minutes on average.



We also developed a completely new question to address the second misconception revealed in think-alouds -- that students believe that all large populations should be normally distributed. In our latest round of think-alouds, half of students picked choice B and stated that populations should be normal, showing the question is identifying the misconception we wanted to test.

What's next?

- We recommend think-alouds as a way to understand student reasoning and develop assessments, interventions, and activities
- We plan to:
  - Develop a set of questions to assess understanding of introductory concepts
  - Administer test to a large sample of students
  - Compare pre-test and post-test scores to measure learning
- Look out for our survey!
- www.stat.cmu.edu/teachstat

In summary, think-aloud interviews are an essential tool for testing the validity of conceptual assessments, leading us to revise many of the questions we've drafted so far. Without think-alouds we could not be confident that our questions assess real student reasoning, or that answer choices correspond to specific misconceptions. We also would not have been able to identify the misconceptions that we did, and develop new questions to test them, just by looking at student answer choices.

In the coming semesters we plan to continue the process, drafting more questions, running more interviews, and administering the questions to a large sample of students to obtain pre- and post-test scores for our course, which will help us measure student learning and determine which concepts need extra work in the course.

Also, watch out -- in the coming months we will send out a survey containing many of our draft questions and requesting feedback on their appropriateness for general introductory courses.

For more information, check out our website, where we'll post updates as we get ready to share our questions and results.

Thanks!

References

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, 33(9), 1289–1312.

DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.

Many thanks to our numerous student participants.