Assessment of Student Learning and Misconception Identification in Intro Statistics

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

Department of Statistics & Data Science, Carnegie Mellon University

Overview

- We piloted think-aloud protocols to understand students' thought processes while answering assessment questions
- Preparation for larger interview and assessment study in Spring 2018
- Goal: Improve quality of assessment and instruction in introductory statistics classes
- One piece of Department's intro stat redesign to align with updated Dietrich General Education curriculum

Assessing Learning in Intro Stats

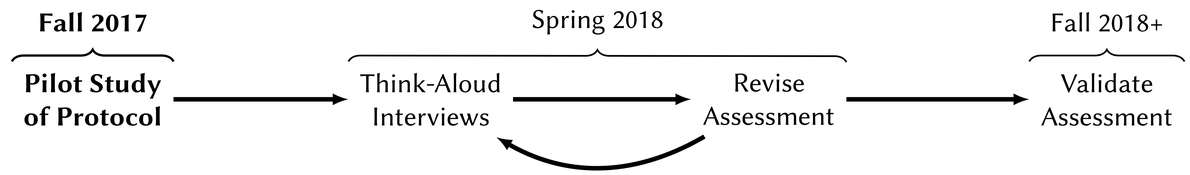
- To improve teaching, need to be able to assess what the students are and are not learning
- Must know whether assessment actually measures student learning (not just test-taking skills, ambiguous questions, etc.)
- We need to know what misconceptions students have to build good assessments

Concerns about Existing Assessments

- Often focus on memorization or equations rather than concepts
- Tested for psychometric reliability, but not for assurance that answers reflect understanding
- DelMas et al. (2007), Jaccobe et al. (2014) report detailed feedback from experts such as statistics course instructors, but not from students

Think-Aloud Interviews

- Students think aloud while answering draft assessment questions
- Interviewers interpret answers: do students choose the right answer for the right reason? Are there common misconceptions?
- Method widely used in science education literature; discussed in Adams & Wieman (2010), Bowen (1994)
- Previous research shows students often interpret questions very differently from instructor expectations



Our Pilot Think-Aloud Protocol

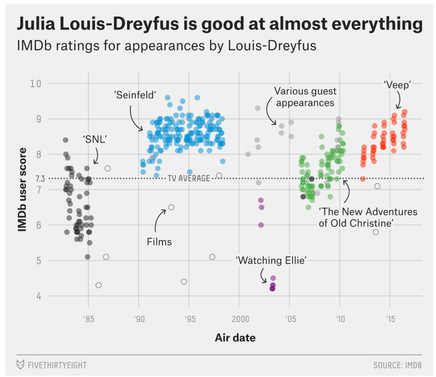
- Students volunteer to be interviewed after solicitation in class, receive $10 gift card incentive
- One interviewer and one note-taker meet the student for 45 minutes to an hour
- Student is asked to think out loud while reading and solving 2 warmups and 10 statistics questions
- "This isn't a test of you... You are helping us test the assignments... You won't hurt our feelings"
- Interviewer says nothing but "Please remember to think out loud" until all questions are complete
- Interviewers ask follow-up questions at the end of the protocol, then solicit student feedback
- Script based on Nodder (2015)

Warmup Questions

~~"I bought a cell phone and a case for a total of $710. The phone cost $700 more than the case. How much did the case cost?"~~

Read the graph and then answer the following questions, explaining how you answered the questions using the graph.

1. Which Julia Louis-Dreyfus TV show had the highest average ratings on IMDb?
2. Did her TV shows or her films get higher ratings?
3. Did Louis-Dreyfus get better average ratings in the 1980s or in the 2000s?



Example Assessment Question, with *Proposed Revisions*

A student participates in a blind Coke versus Pepsi taste test. She correctly identifies which soda is which four times out of six tries. She claims that this proves that she can reliably tell the difference between the two soft drinks.

You want to determine the probability of anyone getting at least four right out six tries just by ~~chance alone~~ *guessing randomly*.

Which of the following would provide an accurate estimate of that probability? Explain why.

(a) Have the *same* student repeat this experiment many times and calculate the percent of the time she correctly distinguishes the brands.

(b) Simulate this on a computer, with a 50% chance of guessing the correct soft drink on each try, and calculate the percentage of times there are *at least* four correct guesses ~~in~~ *out of* six ~~trials~~ *tries*.

(c) Repeat this experiment with a large sample of people and calculate the percentage of people who make *at least* four correct guesses out of six tries.

(d) ~~All of the above methods would provide an accurate estimate of the probability.~~

Open Questions About the Protocol

- How do we motivate students to volunteer?
- Should we take audio or video recordings?
- How do we consistently train interviewers?
- How do we avoid giving students feedback while they answer?
- Students want to *explain* instead of thinking aloud; how do we change this?
- Students may take varying amounts of time—should question order be randomized?
- How to balance need for many revisions with need for sufficient testing of each?
- When during the semester should we run interviews? Students may not have seen all concepts when we interview them

Next Steps

- We recommend this process for anyone writing assessments
- Pilot study has helped us think more carefully about student misconceptions and assessment design
- Full study plan: interview many students, writing and iteratively revising question bank, before validating test as a whole
- Validated test can be used to assess learning, aid redesign for new Dietrich General Education curriculum

References

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

Bowen (1994), *J. Chem. Educ.* **71** 184–190.

DelMas, Garfield, Ooms, and Chance (2007), *Stat. Educ. Res. J.* **6** 28–58.

Jacobbe, Case, Whitaker, and Foti (2014), *ICOTS* **9**.

Nodder (2015). Exercise files for the Lynda.com course on "Foundations of UX: Usability Testing."

Acknowledgments

Thanks to the Eberly Center for advice and feedback. Thanks to the Department of Statistics & Data Science for financial support, and to student volunteers in pilot interviews.