Creating a standardized assessment to measure learning in introductory data science courses

Evan Dragich[1], Matt Beckman[2], Mine Çetinkaya-Rundel[1], Mine Dogucu[3,4], Chelsey Legacy[5], Maria Tackett[1], Andy Zieffler[5]

[1]Duke University, [2]Penn State University, [3]University College London, [4]University of California Irvine, [5]University of Minnesota

Motivation

As data science (DS) continues to grow in popularity among university course offerings, it is becoming crucial to successfully measure students' learning outcomes in introductory courses. To do this requires an assessment designed to which could additionally be used to evaluate pedagogical techniques or curriculum interventions.

Goals

By sharing the work, we hope that participants will become familiar with an assessment they may use for designing intro data science curriculum or researching classroom innovations. We also hope that this instrument can serve as an inspiration or a starting point to be tailored by future researchers more specifically to their courses or to another discipline (eg. by adding more programming concepts instead of data visualization to better serve a computing-focused introductory data science class, etc.)

Background

In the field of statistics, previous work on measuring students' reasoning skills led to the development of the Comprehensive Assessment of Outcomes in Statistics (CAOS). The revised CAOS 4, comprising 40 multiple-choice items on a variety of commonly-taught first-semester introductory concepts, was first administered in 2005 (Delmas et al., 2007).

In 2023, more high schools and universities continue to support the emerging field of data science via specific courses, concentrations, or even majors (Schanzer et al., 2022; Swanstrom, n.d.). Specifically, a review of five introductory data science courses found that, while choice of language varied, all curricula involved some amount of computing or pseudocode (Çetinkaya-Rundel & Ellison, 2021). The next highest frequency topics among curricula were inference and modeling, closely followed by data visualization and data wrangling, with most courses also having some smaller component of communication and ethics. The breadth of diversity captured in these DS curricula motivates the need for a language-agnostic, broad-scope data science assessment that can be tailored further to best meet the needs of specific programs.

In order to let items best measure students' thinking processes, think-aloud interviews with students are essential, not just to clarify potentially confusing wording, but also to ensure that students respond to each item via the thought process intended by the researchers (Reinhart et al., 2022). In addition to measuring students' overall data science mastery, the breadth of topics covered lends naturally to the development of *subscales*, or subsets of the overall scale's items from which a student's *subscore* can be calculated for a particular topic. Common methods include "subscale alphas, exploratory factor analysis, and confirmatory factor analysis" (Jorion et al., 2015).

Development Process

Initial Drafting

In late 2020, members of the research team conducted a search on various university-level introductory DS syllabi to identify common learning objectives. This list of topics was then used to generate **~60 initial items**, ranging in scope from simple graphical interpretations, sophisticated models, and "out-of-the-box" questions regarding topics such as ethics, data importing, and statistical communication.

Think-Aloud Interviews

Once compiled into a single, browsable HTML document, **six think-aloud interviews** were conducted with three professors in data science-adjacent fields, as well as three introductory DS teaching assistants (TAs).

Name	Department Affiliation	Institution Type	Field of Ph.D. Dissertation
Professor X	Statistics	Liberal Arts College	Biostatistics
Professor Y	Computer Science	R1 Research University	Computer Science Education
Professor Z	Computer and Information Science	Liberal Arts School within University	Statistics

Name	Degree Year and Level	Program
TA A	2nd Year Masters	Statistical Science
TA B	4th Year Undergraduate	Economics major, Statistical Science minor
TA C	1st Year Masters	Statistical Science

Interview Flow

- Big-picture questions (Faculty Only)
  - What topics **must** be in an introductory data science course?
  - What topics are **nice to have** in an introductory data science course?
- Item-by-Item Walkthrough
  - Initial reaction
  - Thought process and final answer
  - Comments and suggestions
- Big-picture questions
  - Faculty Only:
    - What are the **strengths** of the current assessment?
  - TA Only:
    - Are the pacing and length appropriate?
  - Both Faculty and TA:
    - What topics **are missing** from the current assessment?
    - What is in the current assessment, but **doesn't belong**?

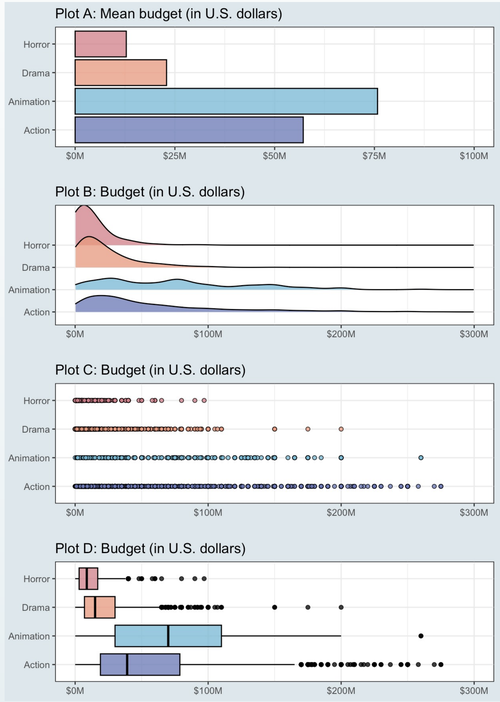
Interview Refinements:

- Recontextualize items for cultural awareness and to combine into fewer passages
- Change question order to avoid priming effects
- Clarify scope as being less technically statistical
- Arrive at final prototype with **26 items** spread over **15 passages**

Sample Item: Movie Budgets I

A data scientist at IMDb has been given a dataset comprised of the revenues and budgets for 2,349 movies made between 1986 and 2016.

Suppose they want to compare several distributional features of the budgets among four different genres---Horror, Drama, Action, and Animation. To do this, they create the following plots.



Fill in the following table by placing a checkmark in the cells corresponding to the attributes of the data that can be determined by examining each of the plots.

	Plot A	Plot B	Plot C	Plot D
Mean	☐	☐	☐	☐
Median	☐	☐	☐	☐
IQR	☐	☐	☐	☐
Shape	☐	☐	☐	☐

References

Çetinkaya-Rundel, M., & Ellison, V. (2021). A Fresh Look at Introductory Data Science. *Journal of Statistics and Data Science Education, 29*(sup1), S16–S26.

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

Jorion, N., Gane, B., DiBello, L., & Pellegrino, J. (2015). Developing and Validating a Concept Inventory. *2015 ASEE Annual Conference and Exposition Proceedings*, 26.497.1-26.497.12.

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczorek, J., Elliott, P., Burckhardt, P., & Nugent, R. (2022). Think-Aloud Interviews: A Tool for Exploring Student Statistical Reasoning. *Journal of Statistics and Data Science Education, 30*(2), 100–113.

Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J. G., Lerner, B. S., Fisler, K., & Krishnamurthi, S. (2022). Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum. *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, 22–28.

Swanstrom, R. (n.d.). *Data Science Colleges and Universities*. Ryan Swanstrom. Retrieved September 26, 2022, from https://ryanswanstrom.com/colleges/

Sample Item: Data Confidentiality

A newspaper reports on the results of a survey from a small (<2000 students) university. The university agrees to have the data released to the public so long as the students' identities and academic standing information are kept confidential. Select the safe combinations of variables that are unlikely to identify any individual students. Explain.

a. Class year and sports played
b. Student ID and dorm ZIP code
c. GPA and major
d. Birth date and phone number
e. None of the above

Next Steps

We will continue to pilot the assessment in large-scale classroom settings and make ultimate refinements to question wording, inclusion, and order. We then hope to begin exploring validity and subscales. In parallel, we hope to convert the format of the assessment from a Qualtrics survey to a robust Javascript framework like the current CAOS.

Interested in helping us pilot our assessment?

We are looking for introductory data science instructors to pilot our assessment in their classes! The 26 items can be used as part of a formative or summative assessment, optional study material for a final exam, or just as an in-class activity.

The full assessment can be found at the QR code on the left. Scan the QR code on the right to leave us your information in a Google form. We will reach out with more information about piloting the assessment in your classroom. Thanks so much!

bit.ly/ds-assess

bit.ly/ds-pilot