Final Meetup

## DATA 606 - Statistics & Probability for Data Analytics

Jason Bryer, Ph.D. and Angela Lui, Ph.D.

December 6, 2023

Final Exam

- Is now available on Blackboard.

- Due by end of day December 10th

- You may use your book and course materials.

- We expect you to complete the exam on your own (i.e. do not discuss with classmates, colleagues, significant others, etc.)

- There are two parts:

  1. Part one multiple choice questions and short answer questions.
  2. Part two has a small data set to analyze with R, then answer some interpretation questions.

- Put your answers in the Rmarkdown file and submit the PDF file. **Please do not post your answers online!**
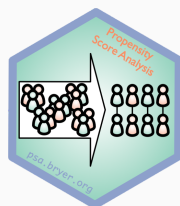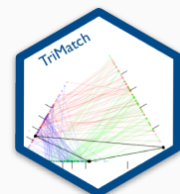
Jason's Work
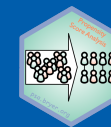
My statistical research interest is in propensity score methods. Propensity score analysis (PSA) is a quasi-experimental design used to estimate causality from observational studies.

Here are some resources for PSA:

- PSA Github repository includes slides slides and Shiny application: https://github.com/jbryer/psa
- Early version of an Intro to PSA book: https://psa.bryer.org
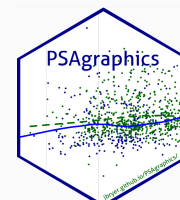- Recording of a talk given this semester for the NYC Meetup group here: https://www.youtube.com/watch?v=JLV4mtFhRMM

multilevelPSA
Multilevel PSA

PSAboot
Bootstrapping PSA

TriMatch
Matching with non-binary treatments

PSAgraphics
Graphical analysis of PSA

Jason's Work (cont.)

Here is list of some other R related projects I have worked on:

`likert` - Analysis and Visualization of Likert Based Items

`ShinyQDA` - R Package and Shiny Application for the Analysis of Qualitative Data

`brickset` - An R Package to Interface with the Brickset API for Getting Data About LEGO sets

`IRRsim` - An R Package for Simulating Inter-Rater Reliability

`mldash` - Machine Learning Dashboard

AmplifyApp, dashboard, and Future Mapping NYC

DAACS (Jason and Angela)

The Diagnostic Assessment and Achievement of College Skills (DAACS) is a suite of technological and social supports to optimize student learning. DAACS provides personalized feedback about students' strengths and weaknesses in terms of key academic and self-regulated learning skills, linking them to the resources to help them be successful students.

Applications of Data Science:

- We use natural language processing and predictive models to machine score the essays.
- We use DAACS data to estimate "risk scores" for students failing so we can target them with resources to help them be successful.
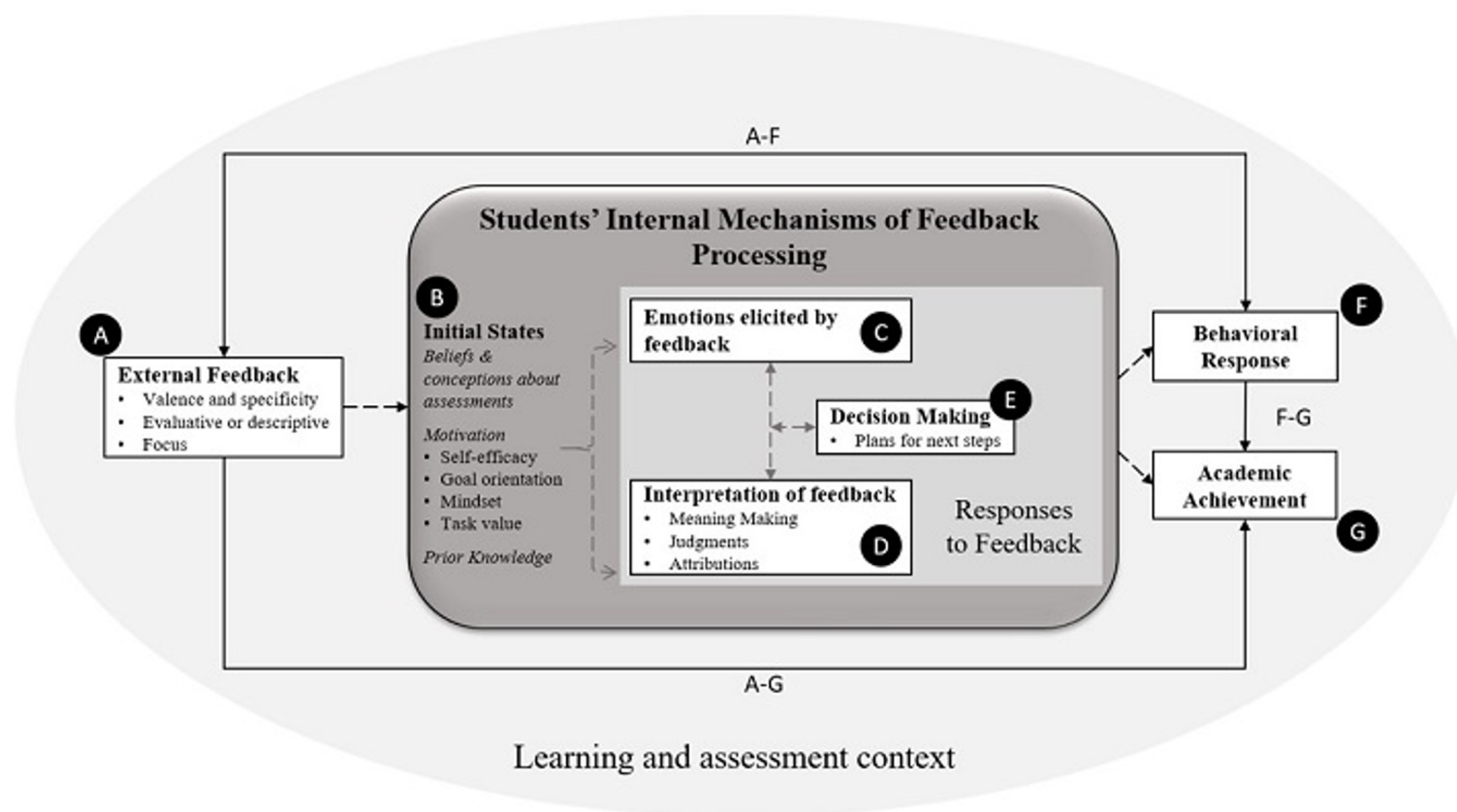
Received a $3.8 million grant this year from the Institute of Education Sciences to test the efficacy at three institutions.

Figure 1

Proposed internal mechanisms involved in students' processing of feedback

Note: solid lines = empirically supported relationships; dotted lines = proposed interplay of internal mechanisms in feedback processing

Exploring Responses to Feedback in DAACS

Some Research Questions

1. What are the sentiments conveyed in students' essays?
2. What judgments about SRL feedback emerge in students' essays?
3. What meanings do students make of the SRL feedback? (a focus on the content criterion?)
4. What attributions do students make if any?
5. To what strategies do students tend to commit?
6. Are these (above five bullets) related to 1) the number of dots they received per domain? 2) the number of feedback pages they viewed? 3) the level of feedback specificity they viewed?

Data Sources

- Students' essays on the DAACS Writing Assessment
- Students' SRL Assessment Results and Feedback

Analyses

- Sentiment Analysis
- Content Analysis

Thank You

This has been a great semester. Please don't hesitate to reach out:

✉ Email: jason.bryer@cuny.edu

 Github: https://github.com/jbryer

🔗 Personal Website: https://bryer.org

in LinkedIn: jasonbryer

 Mastodon: @jbryer@vis.social

✉ Email: angela.lui@cuny.edu

in LinkedIn: angela-m-lui-8b051928

You can download all course materials on Github. Click the clone or download link to download zip file.