

Final Meetup

DATA 606 - Statistics & Probability for Data Analytics

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

May 11, 2022

Final Exam

- Is now available on Blackboard.
- Due by end of day May 15th.
- 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!**

Think Like a Data Scientist

May 24th 7:00pm to 8:30pm

This event will feature five data science projects which grew from a semester-long independent studies course (DATA661) at CUNY-SPS. These projects contribute significantly to the FutureMapping and DAACS projects. Join us to learn about the range of knowledge, skills, and contributions these budding data scientists have to share with us!

- "Introducing DataDash" by Bikram Barua
- "Leveraging Shiny Modules to Make Multi-Tabbed Leaflet Maps" by Rachel Greenlee & Zach Palmore
- "Learning MongoDB and MongoLite: A Discussion of Data" by Gabriel Abreu
- "Use of Structural Topic Models in R to Interpret Students' Responses to Automated Feedback on Self-Regulated Learning" by Sean Connin
- "DAACS Writing Assessment Sentiment Analysis" by Santiago Torres

Register here: <https://www.meetup.com/Albany-R-Users-Group/events/285821889/>



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. It is generally conducted in two phases:

1. Estimate propensity scores (i.e. probability of being in the treatment) using the observed covariates. a. Check balance b. Re-estimate propensity scores
2. Estimate effect sizes using typical group differences (e.g. t-tests)

See my [Github repository](#) or [Intro to PSA slides](#). Also the PSA Shiny application:

```
psa::psa_shiny()
```

Areas I have worked on:

- Multilevel PSA (see [multilevelPSA](#) R package)
- Matching with non-binary treatments (see [TriMatch](#) R package)
- Bootstrapping PSA (see [PSAboot](#) R package)

DAACS

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:

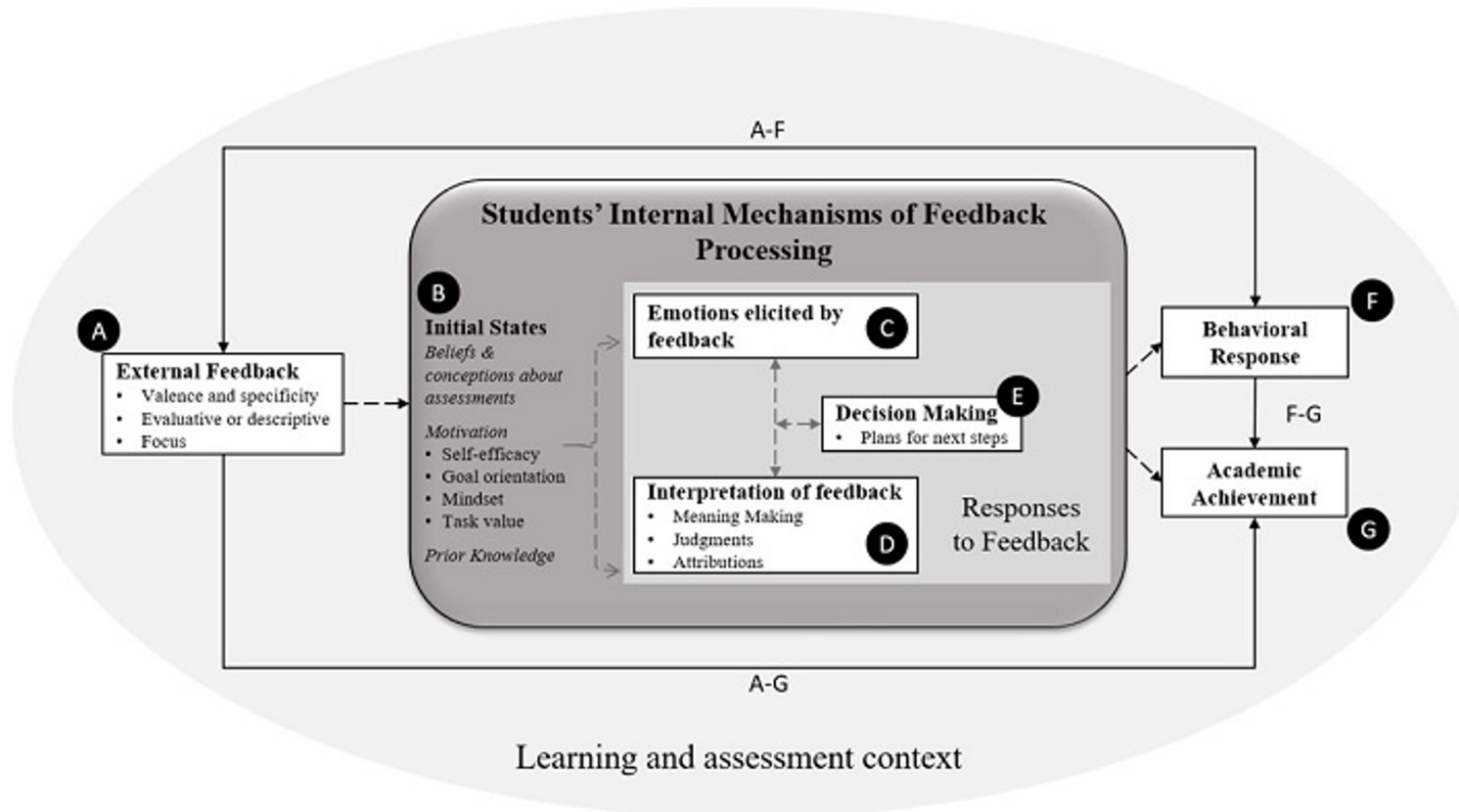
- We use natural language processing and predictive models to machine score the essays.
- We use DAACS data to estimate "risk scores" for students failing so we can target them with resources to help them be successful.

Received a \$3.8 million grant this year from the Institute of Education Sciences to test the efficacy at three institutions.

Students' responses to feedback

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>

🌐 LinkedIn: [jasonbryer](#)

🐦 Twitter: [jbryer](#)

✉ Email: angela.lui@cuny.edu

🌐 LinkedIn: [angela-m-lui-8b051928](#)

You can download all course materials on [Github](#). Click the [clone or download](#) link to download zip file.