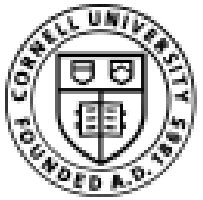


Cornell University

NSF DUE-IUSE: 1611482

Impact of lab curricula on students' critical thinking skills

Cole Walsh & N.G. Holmes, AAPT Summer Meeting 2019

Cornell University

NSF DUE-IUSE: 1611482

~~Impact of lab curricula on students' critical thinking skills~~

New perspectives on students' performance on physics concept inventories

Cole Walsh & N.G. Holmes, AAPT Summer Meeting 2019

Study rationale

- Previous research using concept inventories and attitudes surveys have found differences in performance between men and women

- The instruments are often assumed to provide objective measurements whose results are interpreted using a deficit model.

- Using a recently developed instrument that assesses critical thinking (CT) skills in physics labs with a unique testing format, we offer new perspectives.

Madsen et al. (2013); Henderson et al. (2017, 2019); Traxler et al. (2016)

Research Questions

- **How do different students perform on the PLIC?**
  - Prior preparation; dependent variable: Prescore
  - Student gains; dependent variable: Gain = Postscore – Prescore
  - Focus on gender here

- **Why do different students perform differently?**
  - Confidence on Survey
  - Attitudes towards labs
  - Self-efficacy towards labs
  - Test format
  - Test construct

Traxler et al. (2018); Henderson et al. (2018); Kalendar et al. (2018); Follette et al. (2017); Salehi et al. (2019)

How to assess critical thinking?

The Physics Lab Inventory of Critical thinking (PLIC)

- Context: case studies of hypothetical groups performing a mass on a spring experiment

What features were most important in comparing the two k values? Please select no more than 3 items.

- ☐ The difference between the two k-values
- ☐ The difference between the two k-values compared to the uncertainty
- ☐ The percent difference between the k-values (i.e. the fraction [k1-k2]/[k1+k2])
- ☐ The difference between the two periods
- ☐ The size of the uncertainty (or variability between data)
- ☐ How they accounted for human error
- ☐ Other (Please describe)

Walsh et al. 2019

How to assess critical thinking?

- What does it measure?: Students' critical thinking (CT) skills in the context of physics experimentation
  - evaluate models,
  - evaluate methods,
  - proposing follow-up investigations

See PERC poster (B58 Poster Session II Wed 8:15pm) for more details!

Data Sources

- Matched data from **2434 students** from:
  - **56 courses** (32 first-year [FY], 24 beyond-first-year [BFY])
  - **23 institutions** (9 four-year colleges, 2 master's-granting, 12 PhD granting)

Participation by self-declared gender

Research Questions

- **How do different students perform on the PLIC?**
  - Prior preparation; dependent variable: Prescore
  - Student gains; dependent variable: Gain = Postscore – Prescore
  - Focus on gender in this talk

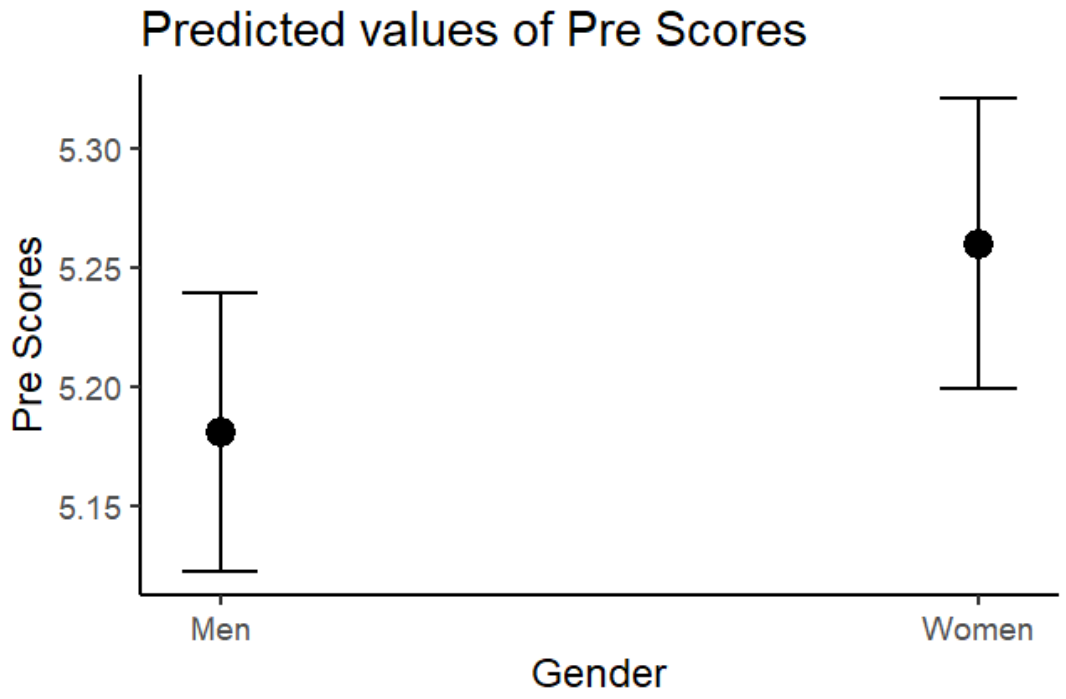
- **Why do different students perform differently?**
  - Confidence on Survey
  - Attitudes towards labs
  - Self-efficacy in labs
  - Test format
  - Test construct

Modeling Prior Preparation

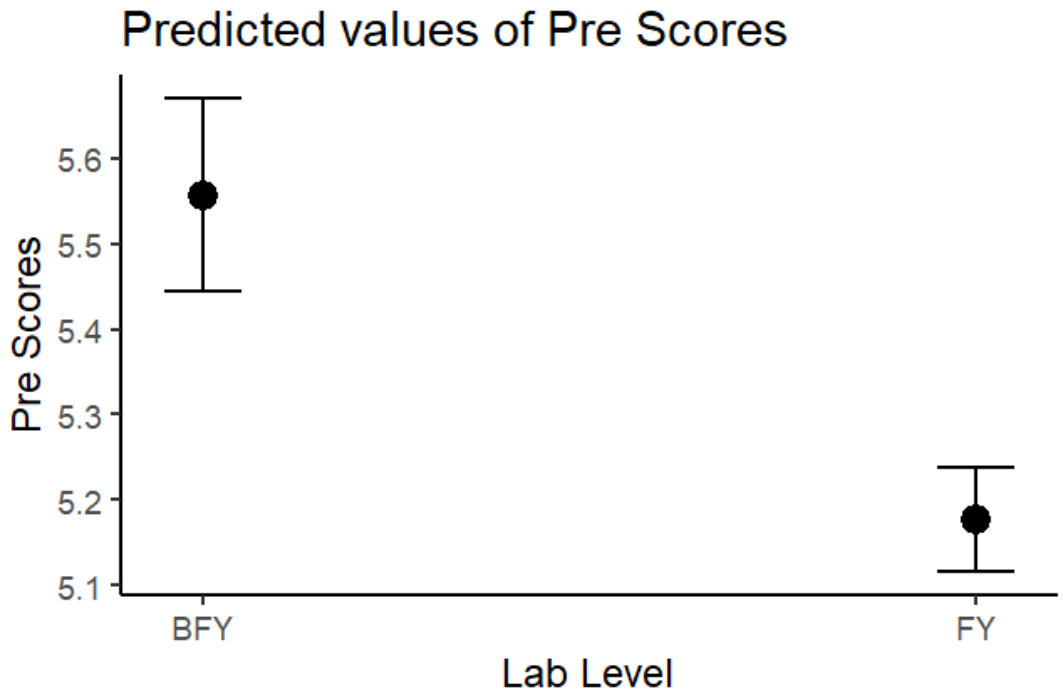
- Linear Mixed Model for **Prescores**
  - Random intercepts for courses
  - Fixed effects for:
    - Lab Level (FY or BFY)
    - Major (Physics, Engineering, Other)
    - **Gender (Men or Women)**
    - URM Status (URM or Majority)

Predicted Prescores

- $\beta_{Women} = \mathbf{0.08 \pm 0.04}, p = \mathbf{0.08}$
- If there is a difference, its small…

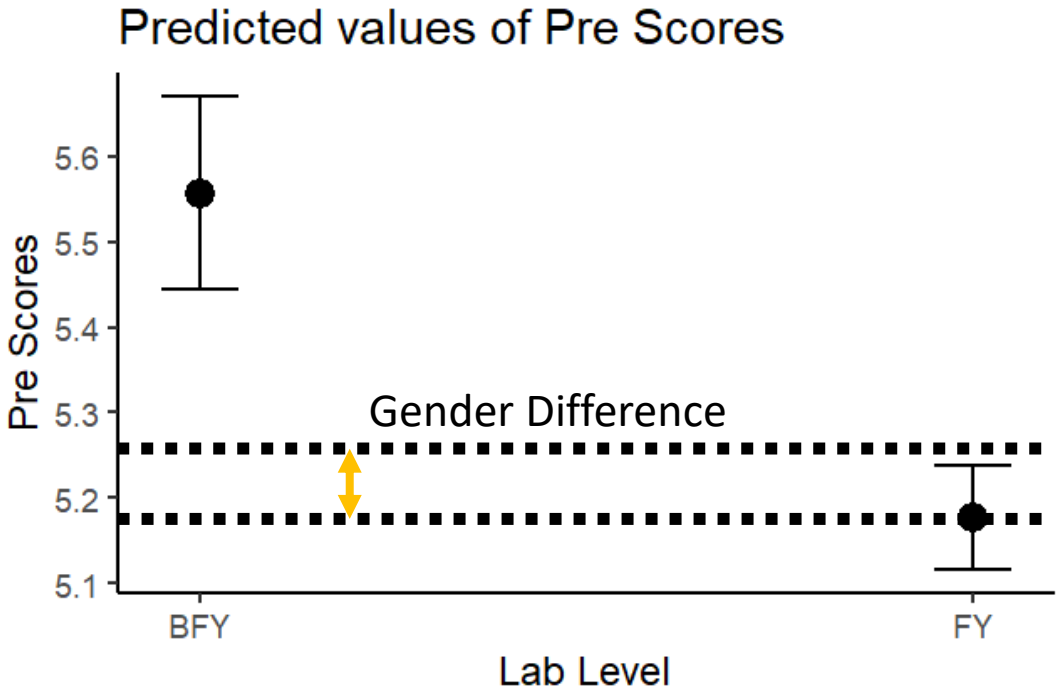


For an idea of scale…



- $\beta_{BFY} = \mathbf{0.38 \pm 0.13}, \boldsymbol{p} < \mathbf{0.01}$
- The difference between students in FY and BFY labs is almost **5X** the difference between men and women, on average.

For an idea of scale…



- $\beta_{BFY} = \mathbf{0.38 \pm 0.13}, \boldsymbol{p} < \mathbf{0.01}$
- The difference between students in FY and BFY labs is almost **5X** the difference between men and women, on average.

Research Questions

- **How do different students perform on the PLIC?**
  - Prior preparation; dependent variable: Prescore
  - Student gains; dependent variable: Gain = Postscore – Prescore
  - Focus on gender here

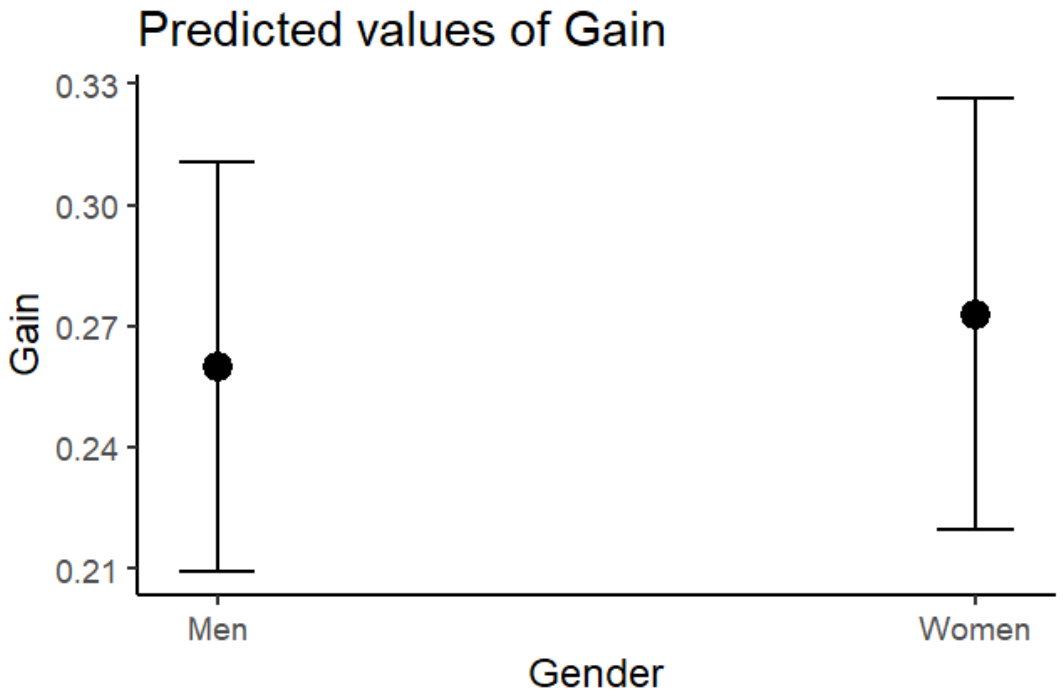
- **Why do different students perform differently?**
  - Confidence on Survey
  - Attitudes towards labs
  - Self-efficacy in labs
  - Test format
  - Test construct

Modeling Student Gains

- Linear Mixed Model for **Gains**
  - Random intercepts for courses
  - Fixed effects for:
    - **Prescores**
    - Lab Level (FY or BFY)
    - Major (Physics, Engineering, Other)
    - **Gender (Men or Women)**
    - URM Status (URM or Majority)

Predicted Gains

- $\beta_{Women} = \mathbf{0.01 \pm 0.05}, \boldsymbol{p} = \mathbf{0.782}$
- There's really no difference here even after controlling for prior preparation.

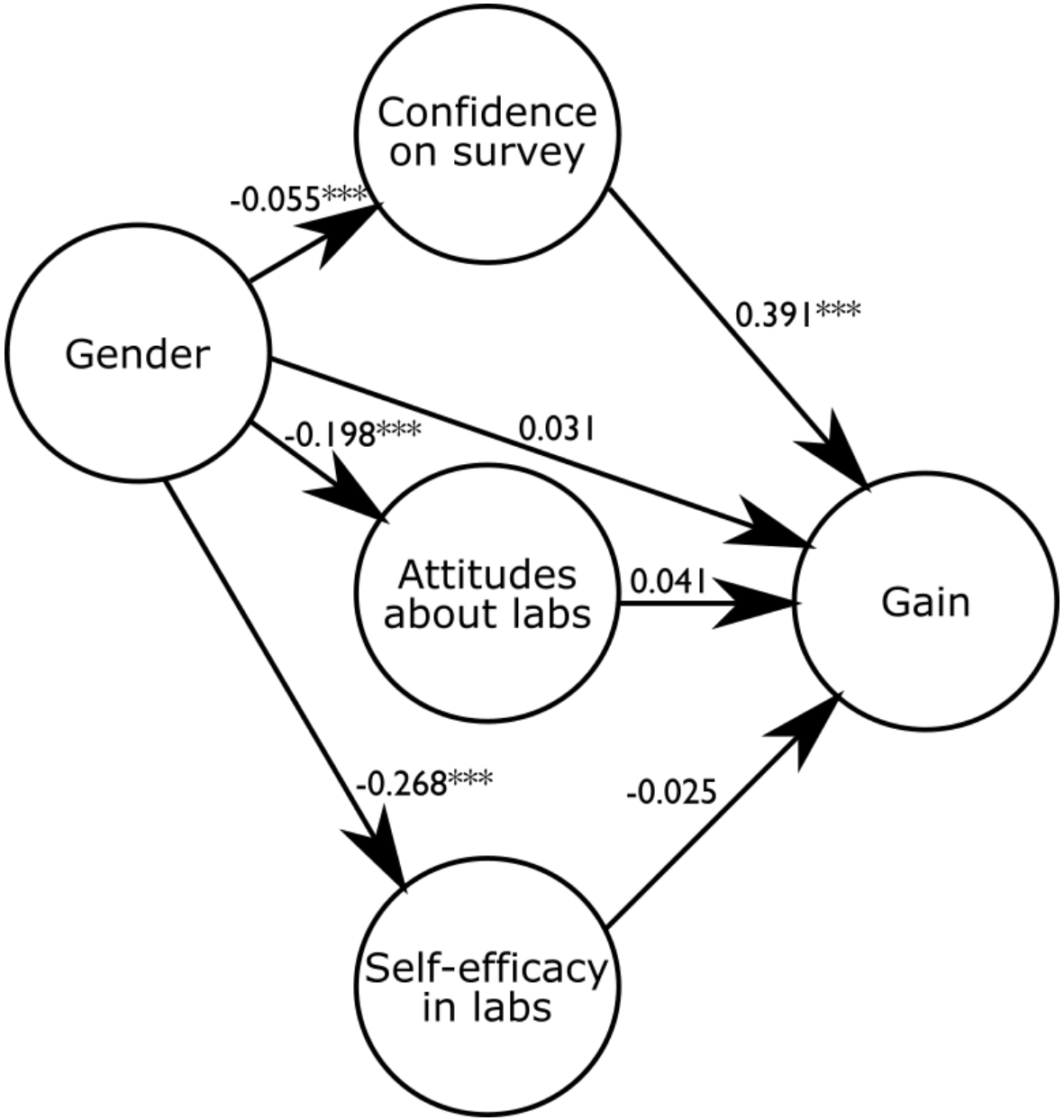


Research Questions

- **How do different students perform on the PLIC?**
  - Prior preparation; dependent variable: Prescore
  - Student gains; dependent variable: Gain = Postscore – Prescore
  - Focus on gender here

- **Why do different students (not) perform differently?**
  - Confidence on Survey
  - Attitudes towards labs
  - Self-efficacy in labs
  - Test format
  - Test construct

Multilevel structural equation modeling (SEM)



Follette et al. (2017); Kalendar et al. (2018); Nokes-Malach et al. (2018)

Research Questions

- **How do different students perform on the PLIC?**
  - Prior preparation; dependent variable: Prescore
  - Student gains; dependent variable: Gain = Postscore – Prescore
  - Focus on gender here

- **Why do different students (not) perform differently?**
  - Confidence on Survey
  - ~~Attitudes towards labs~~
  - ~~Self-efficacy in labs~~
  - Test format
  - Test construct

Conclusions

- Practically no difference in prescores for men and women on the PLIC and both men and women improve equally, on average.
  - In future work we will explore the intersectionality of students' identities

- Why are our results different from those collected using other instruments in PER?
  - Our findings about students' confidence, attitudes, and self-efficacy agree with prior literature, but don't explain the discrepancy in performance results
  - Is it because of the measurement tool (i.e., the multiple-response format)?

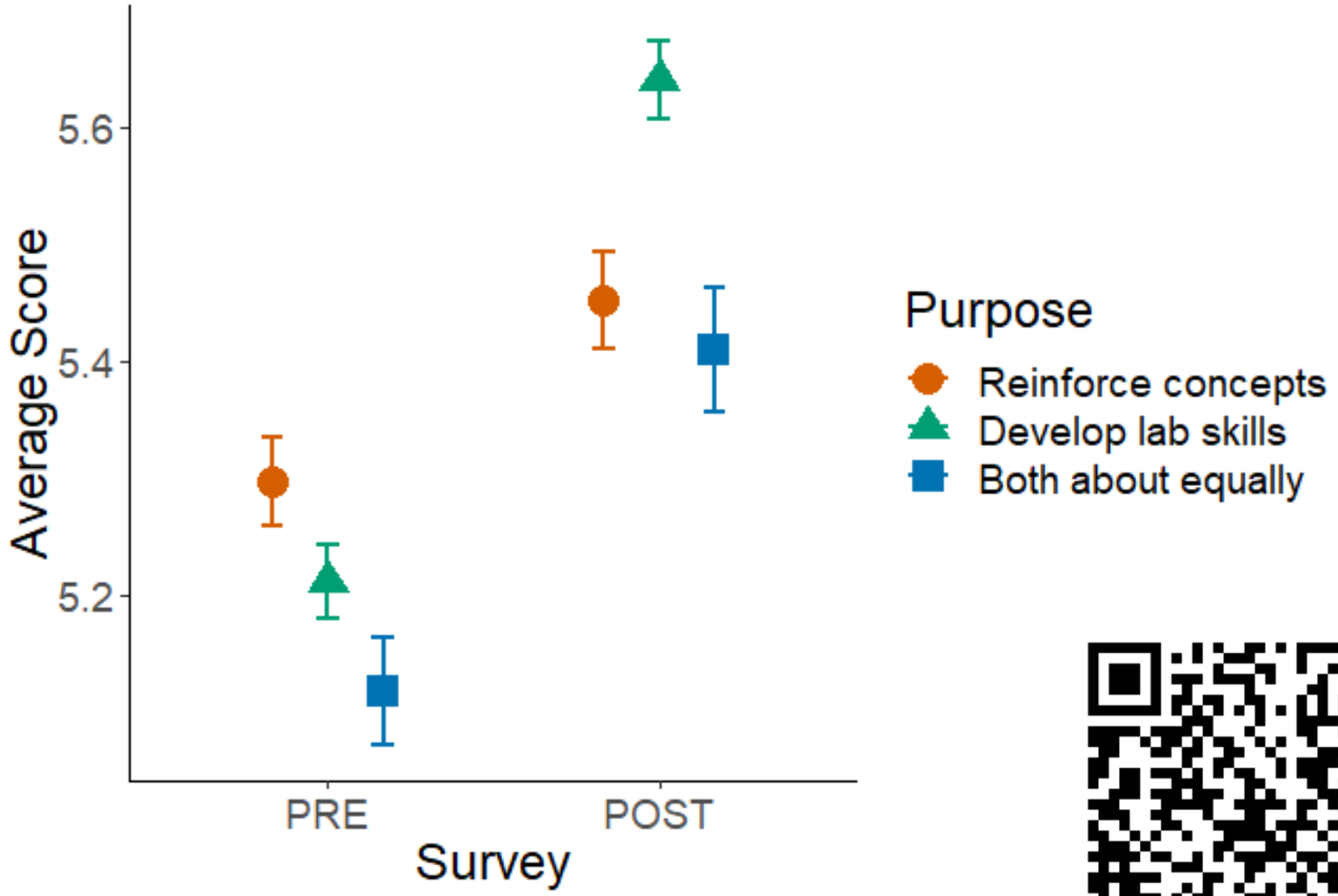
Or is there something that distinguishes CT skills from conceptual knowledge and attitudes?

Nature of science and general CT assessments have observed similar results (VASS, VNOS, CLA+, CWRA+, CAT, CCTST)

Halloun (1996); Khalick (2000); Council for Aid to Education (2014, 2015, 2016, 2017); Stein et al. (2007); Facione (1990)

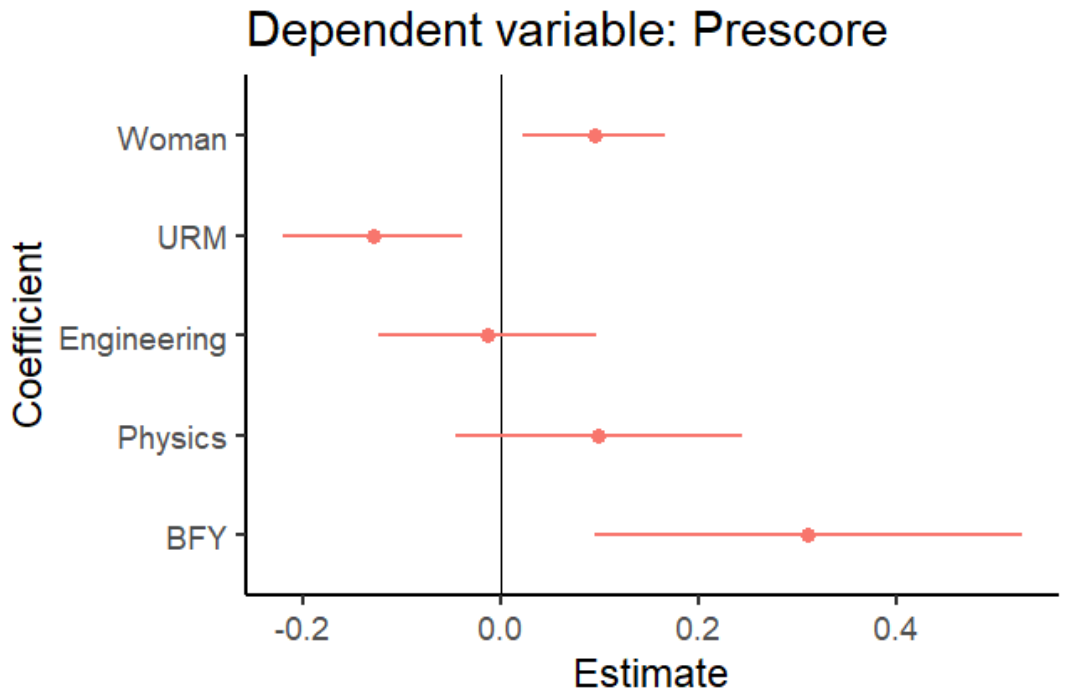
Thank You! Questions?

Students trained in labs designed to develop lab skills see larger gains, on average.

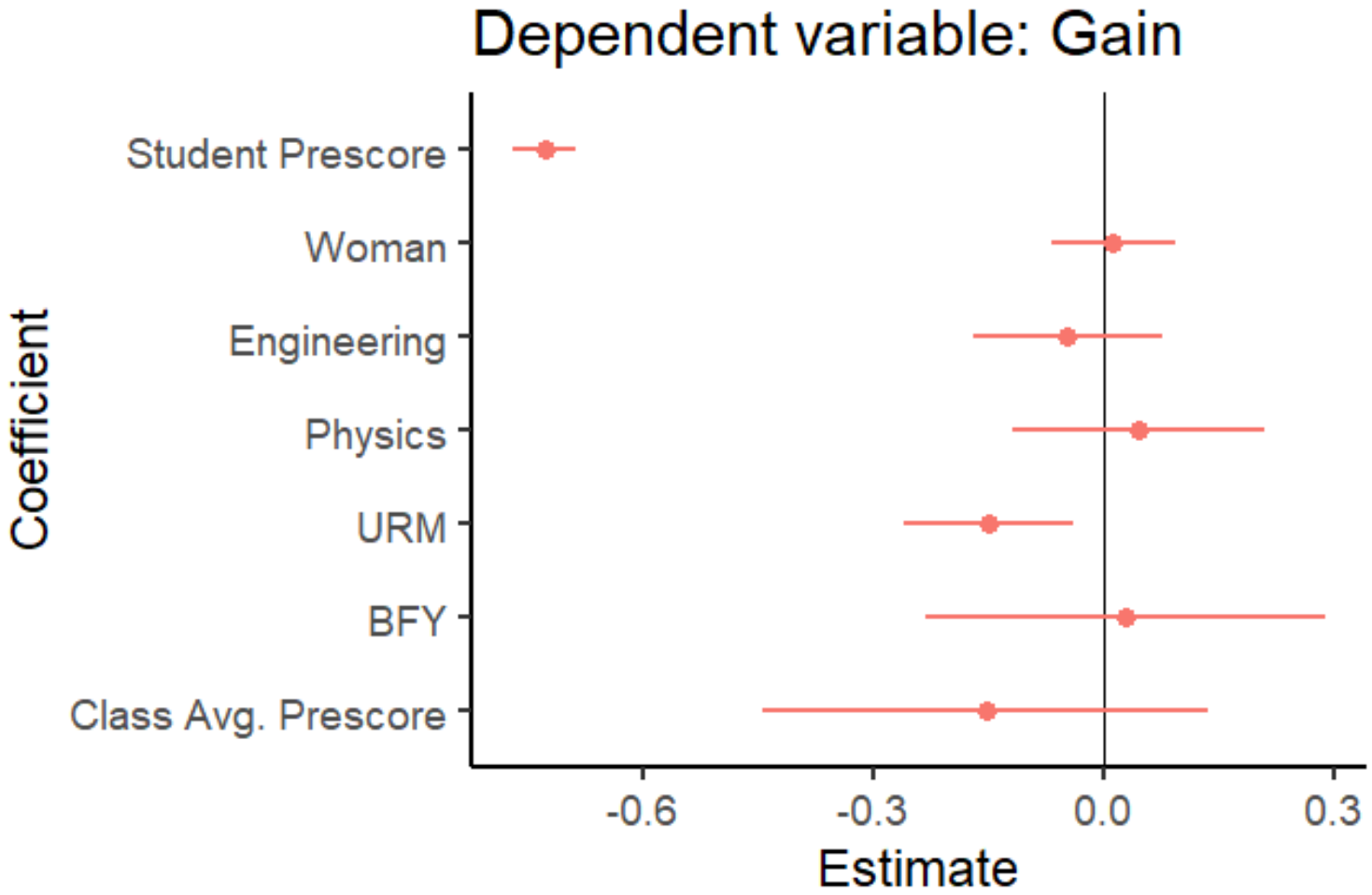


HMI results; Prescore

- After survey, student, and course filters, 4211 students remained in our dataset
- We imputed data for students who were missing either a pre or postsurvey using hierarchical multiple imputation (HMI)
- Results agreed with that from matched sample with improved precision



HMI results; Gain



Confidence, attitudes, and self-efficacy questions from the PLIC

Confidence about responses to survey

- How difficult were the questions in this survey?
- How confident do you feel in your responses to this survey?
- How much effort did you put into this survey?

Attitudes about labs

- Lab Experiments are:
  - Interesting -> Boring
  - Useful -> Useless
  - Easy -> Hard
  - Fun -> Scary

Self-efficacy in labs

- I feel confident analyzing data
- I feel confident doing experiments in labs

Thank you!

Questions?