Techniques for Rote Memorization

MIDS W241 Summer 2020 Final Project
Preliminary Evaluation

Tina Huang
Jeff Li
Gunnar Mein

Core Question

Experiment question

- Does rote vocalization + writing method improve memorization?

Motivation

- Validate techniques used in classroom
- Rote memorization important part of many fields

Experiment Design

Covariates: Age, gender, prior knowledge, reading level, practice

DiD - Part 1 - Baseline

1. Read 2 trivia topic paragraphs (less than 2 mins each) and questions
2. Watch 1 minute distraction video
3. Answer 4 questions about each topic (5 options multiple choice)

DiD - Part 2 - Control

- 2 new topics
- Rest of steps as in Part 1

DiD - Part 2 - Treatment

- 2 new topics
- **Vocalize**
- **Write**
- Rest of steps as in Part 1

Power calculation

- Online calculator:
  https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html
  - Sigma from binomial distribution
  - P = 0.2 o for right answer
  - 8 questions,
  - Sigma = 1.131.
  - 21 samples per group for 80% power

- Overestimated effect size from the outset
  - Performance not random
  - Performance skewed toward correct answers
  - Bug in pilot made pilot results iffy

- Experiment:
  - Final size: ~57 per group

- ○ Calculate Sample Size (for specified Power)
- ○ Calculate Power (for specified Sample Size)
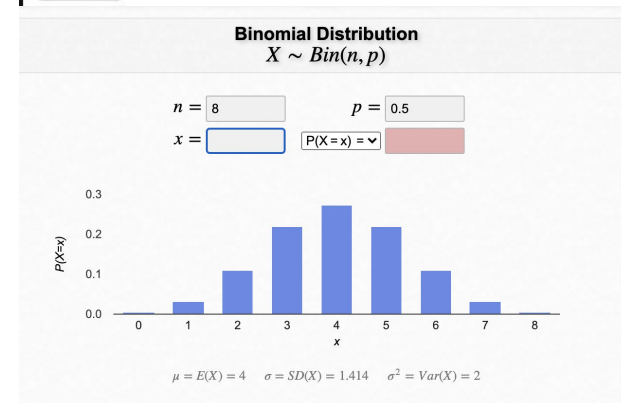
Enter a value for mu1: 6

Enter a value for mu2: 7

Enter a value for sigma: 1

- ○ 1 Sided Test
- ● 2 Sided Test

Enter a value for α (default is .05): .05

Enter a value for desired power (default is .80): .8

The sample size (for each sample separately) is: 16

Calculate

Binomial Distribution
$$X \sim Bin(n, p)$$

$n =$ 8 $p =$ 0.5

$x =$ ☐   P(X = x) = ▾

$\mu = E(X) = 4$ $\sigma = SD(X) = 1.414$ $\sigma^2 = Var(X) = 2$

BerkeleyMIDS

Code, Audience, Timeline

Website/Code

- Custom Java/HTML/JS, records to multiple CSVs
- On Microsoft Azure (small Linux box, $13 per month)
- Code available at https://github.com/gunnarmein-ucbischool/w241-rote

Audience

- Middle + high school students (emails from school, ask on Teams)
- Shared to connections through social media: LinkedIn, Facebook, etc.

Timeline

- Small size test experiment: 7/21 - 7/23
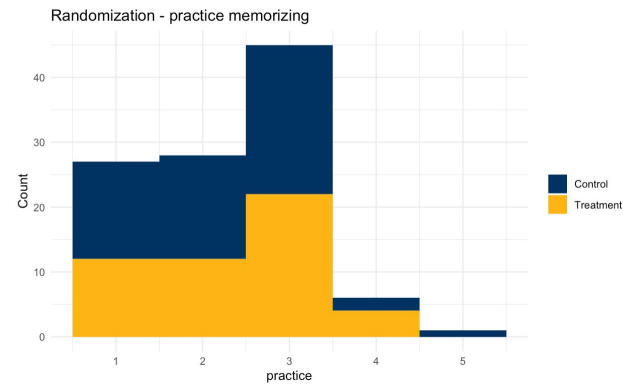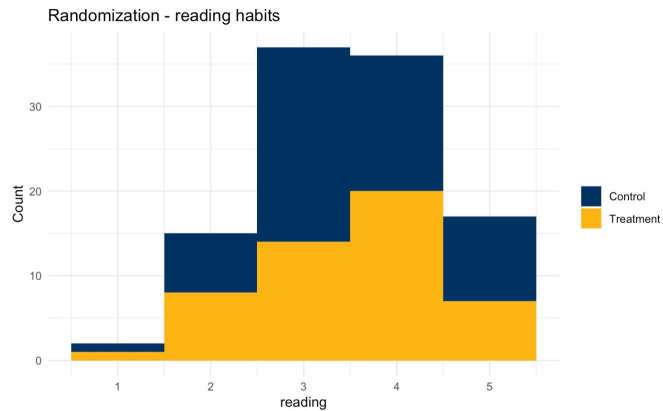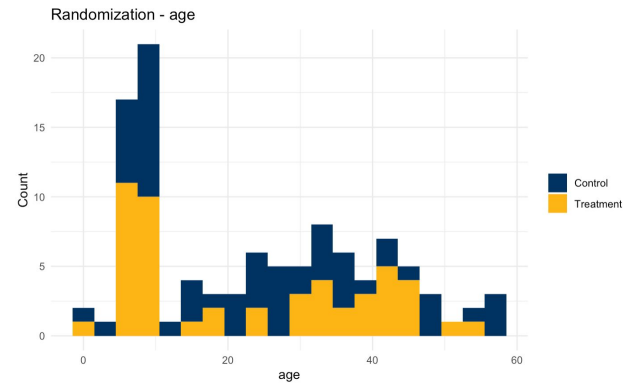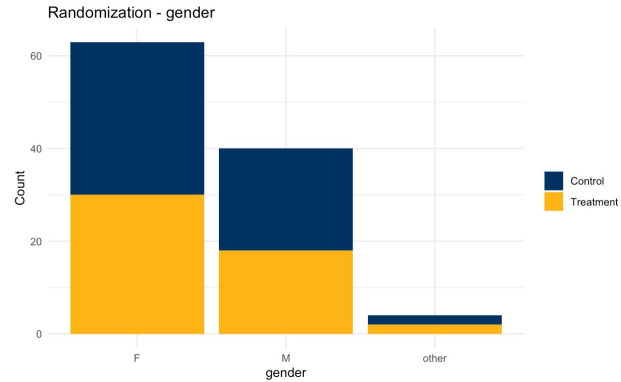- Actual experiment: 7/24 - 8/1

BerkeleyMIDS

Implementation problems

- Bug in code severely affected pilot study

- People be cheatin' (back button)

- Email announcement was sent 3 hours before scheduled time

- Email announcement was buried under news

BerkeleyMIDS

Who completed this test?
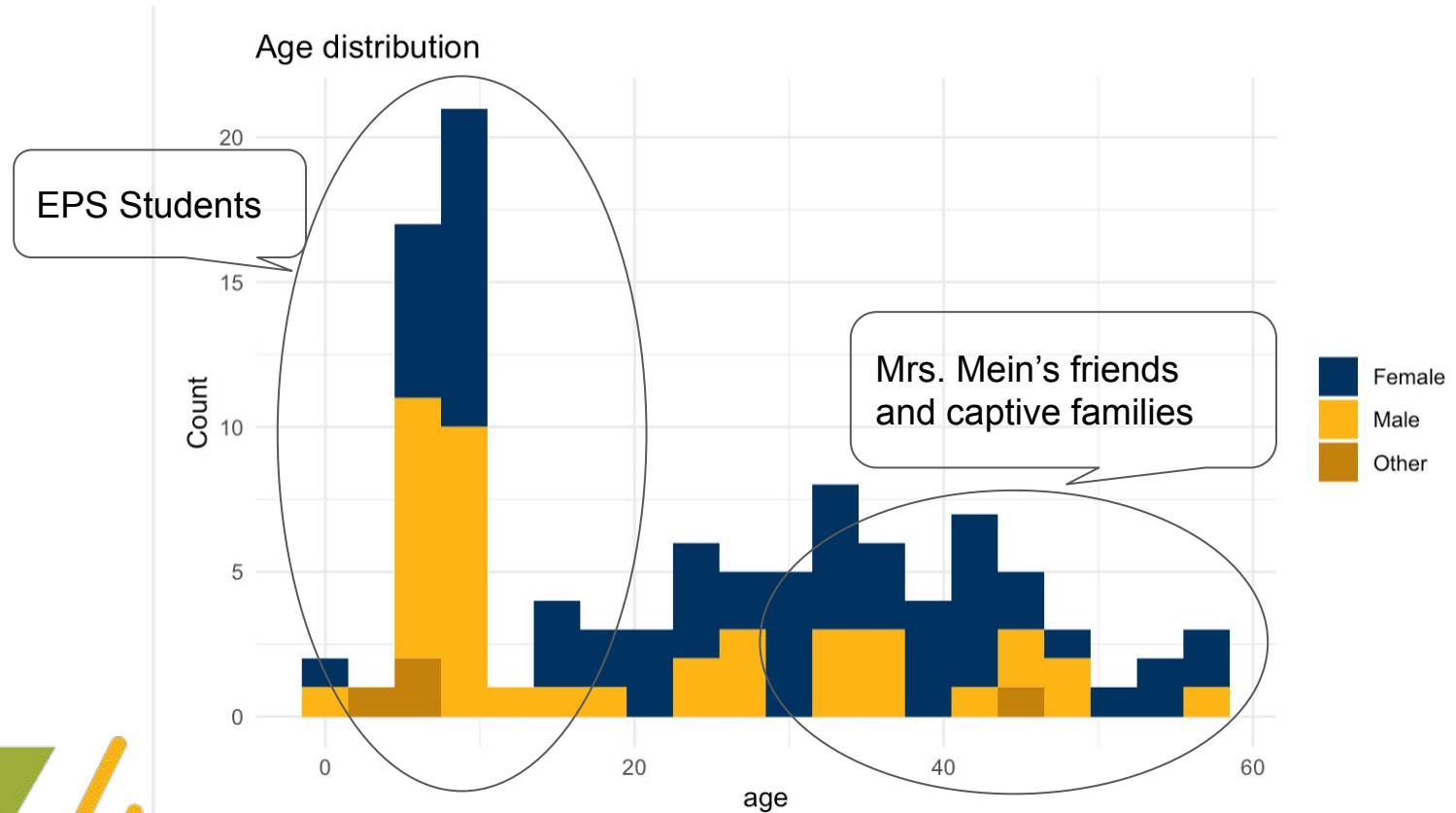
Randomization Check - Visual

Randomization check - Regression

Just after assignment:

```
Call:
glm(formula = treat ~ as.numeric(age) + gender + practice + reading +
    total_prior_knowledge, family = binomial(link = "logit"),
    data = df_cov)

Deviance Residuals:
    Min      1Q   Median      3Q     Max
-1.3466  -1.1913   0.9992   1.1545   1.3134

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           -0.597085   0.819745  -0.728    0.466
as.numeric(age)       -0.001914   0.010377  -0.184    0.854
genderM               -0.226372   0.366925  -0.617    0.537
genderother           -0.092063   1.430418  -0.064    0.949
practice               0.029549   0.182874   0.162    0.872
reading                0.121791   0.185597   0.656    0.512
total_prior_knowledge  0.038501   0.078087   0.493    0.622
```

Completed surveys only:

```
Call:
glm(formula = treat.x ~ as.numeric(age) + gender + practice +
    reading + total_prior_knowledge, family = binomial(link = "logit"),
    data = df_completed)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.304  -1.123  -1.006   1.221   1.479

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           -2.619e-01  9.426e-01  -0.278    0.781
as.numeric(age)       -8.774e-03  1.347e-02  -0.651    0.515
genderM               -3.255e-01  4.773e-01  -0.682    0.495
genderother            1.589e+01  1.455e+03   0.011    0.991
practice               7.319e-03  2.354e-01   0.031    0.975
reading               -1.316e-02  2.315e-01  -0.057    0.955
total_prior_knowledge  7.226e-02  9.910e-02   0.729    0.466
```

BerkeleyMIDS

Anticipated Interference

Potential Interference

- Not a real transferable treatment (interference would mean cheating)
- Potential reasons: Participants taking test together or sharing information
- Potential Impact: Underestimation of the ATE
- Solution:
  - Randomized test questions
  - Clustering by time slot:
    - New cluster randomly assigned every 5 minutes

BerkeleyMIDS

Results - at a glance

Score difference between baseline and experiment

BerkeleyMIDS
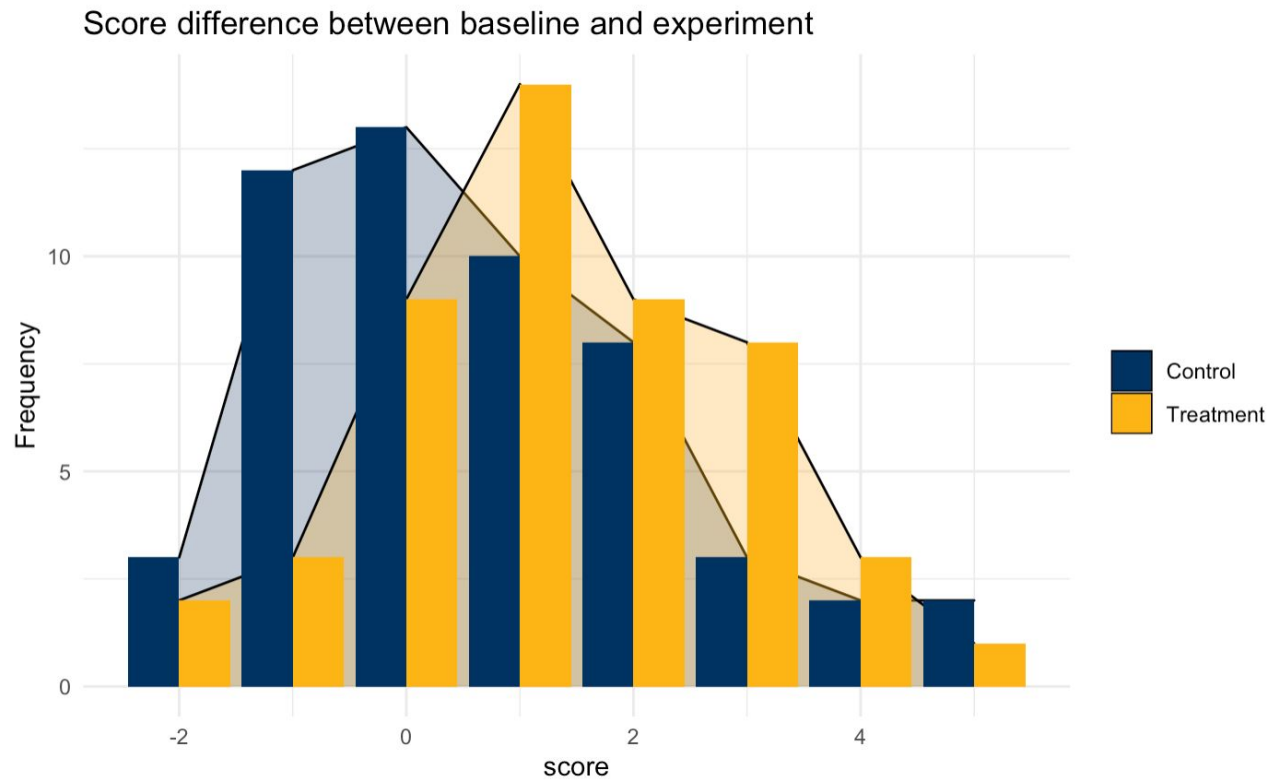
Results - full regression tables

Table 2: Regression Results with Clustered Standard Errors

| | Dependent variable: | |
| | score | |
	(1)	(2)
treat	0.663*	0.659*
	(0.327)	(0.319)
age	0.006	
	(0.009)	
prior_knowledge	−0.027	
	(0.077)	
reading	−0.025	
	(0.184)	
gender2	0.712	
	(0.399)	
practice	0.153	
	(0.176)	
Constant	0.126	0.660**
	(0.731)	(0.245)
Observations	94	97
Adjusted R^2	0.007	0.029
Residual Std. Error	1.679 (df = 87)	1.646 (df = 95)
F Statistic	1.112 (df = 6; 87)	3.884 (df = 1; 95)

ATET - Average Treatment Effect on Treated

- ATE, ITT not measurable in the face of significant attrition

- ATET = 0.66 (SE 0.33)

- Statistically significant

- Worth the extra effort? You decide.

BerkeleyMIDS

Heterogeneous Treatment Effects

- Tested for age, gender, reading, practice

- No significant effects were found

Discussion

Things that went wrong::

- Test calibration
- Implementation details
- Attrition
- Initial power calculation

Things that went right:

- **Treatment effect!**
- Clustering
- Randomization
- Recording
- Cost-effectiveness

Attrition

- ● 28% drop between submission of covariates and first test
  - ○ 22% had already dropped off after pressing "start"
  - ○ 37% drop from submitted covariates by the end
  - ○ Potential reasons: test is too long, lack of interest, technical difficulties

Attrition by stage

BerkeleyMIDS
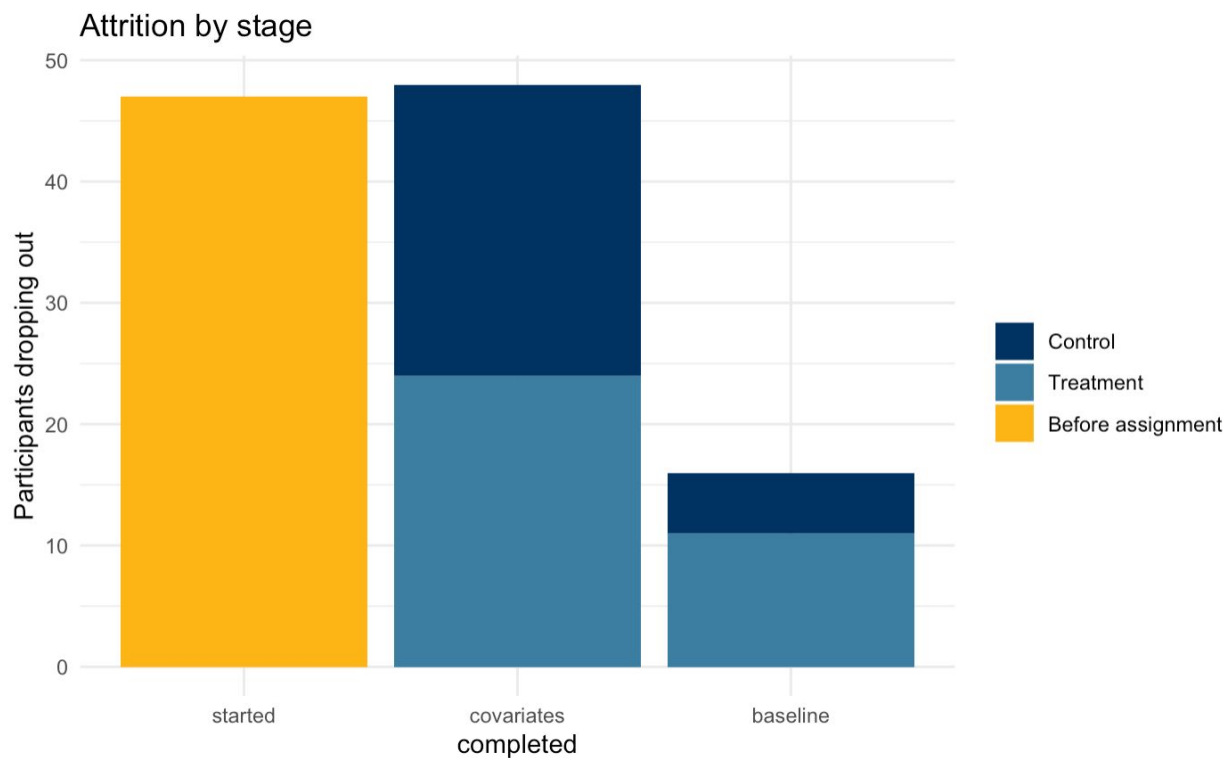
Improvements for Future Research

- Longer pre-study (2-3 weeks)

- Harder questions

- Shorter content

- Hard-code page flow

- Captive audience

- Random order of baseline/treatment?

BerkeleyMIDS

Thank you for listening!

Questions?