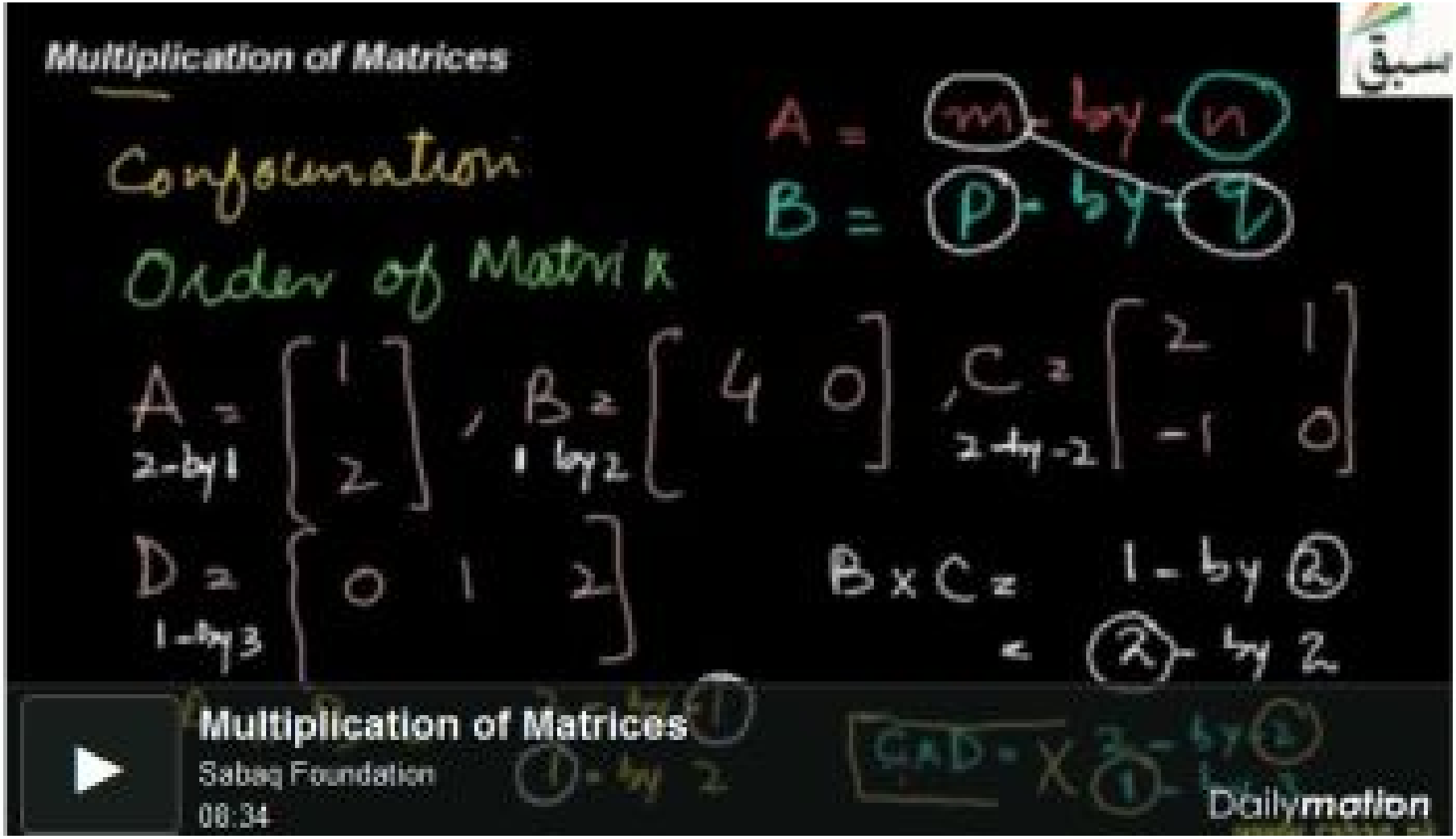
What's the Better Math Teacher: Video or Text? A Field Experiment in Online Education

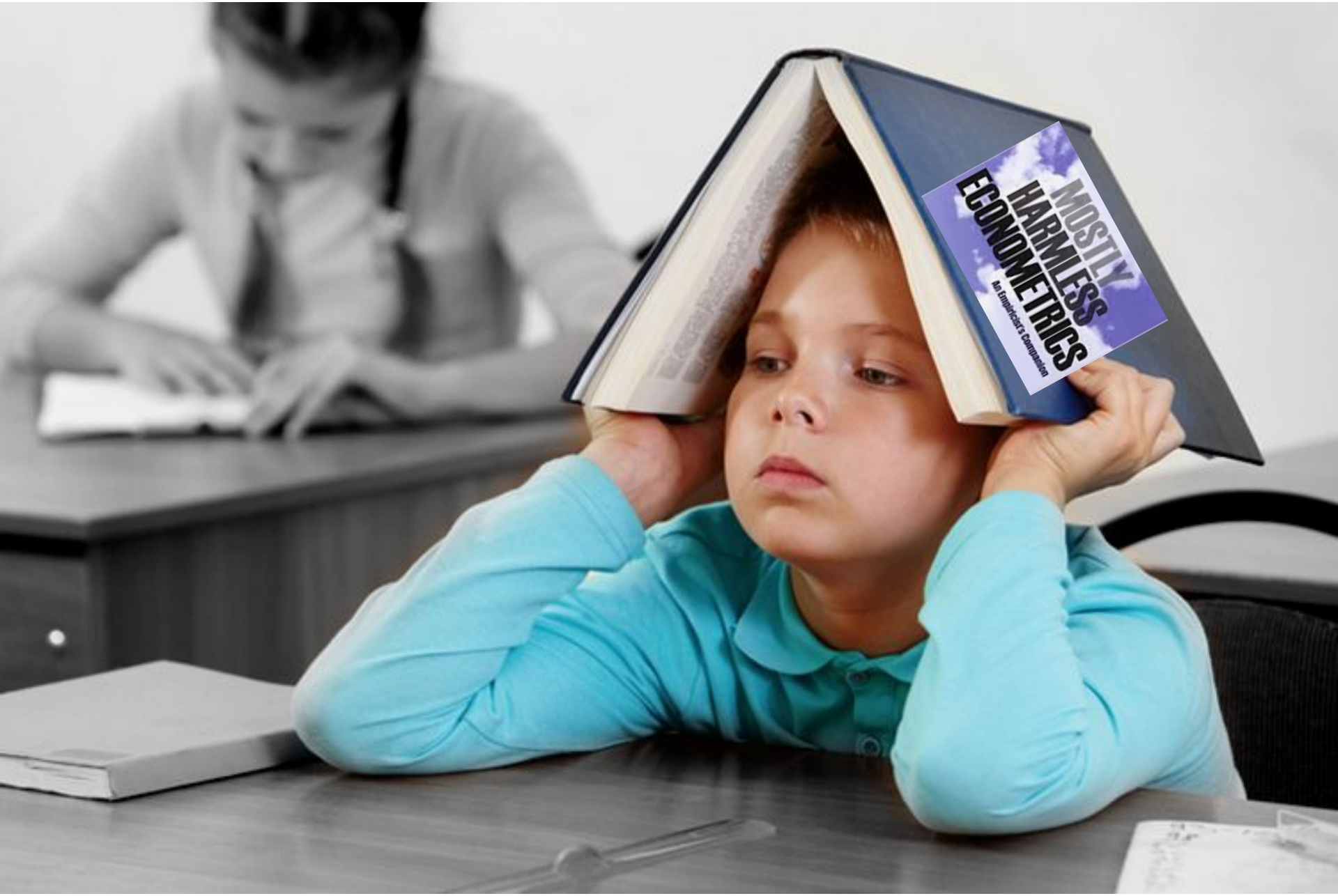
Karthik Chepudira, Brad McMinn, Ross Katz, Brandon Shurick, and Matt Hayes

Online Education is the Future!



Or is it?

- **Accessibility benefits are obvious**, but learning productivity benefits are hard to quantify
- **Comparisons are mismatched**: dense, poorly written text books vs. clear, concise, and colloquial videos
- Online methods utilize **“old media” Andragogy & Pedagogy** on “new media” platforms
- **Anything new** would seem better than the alternative…

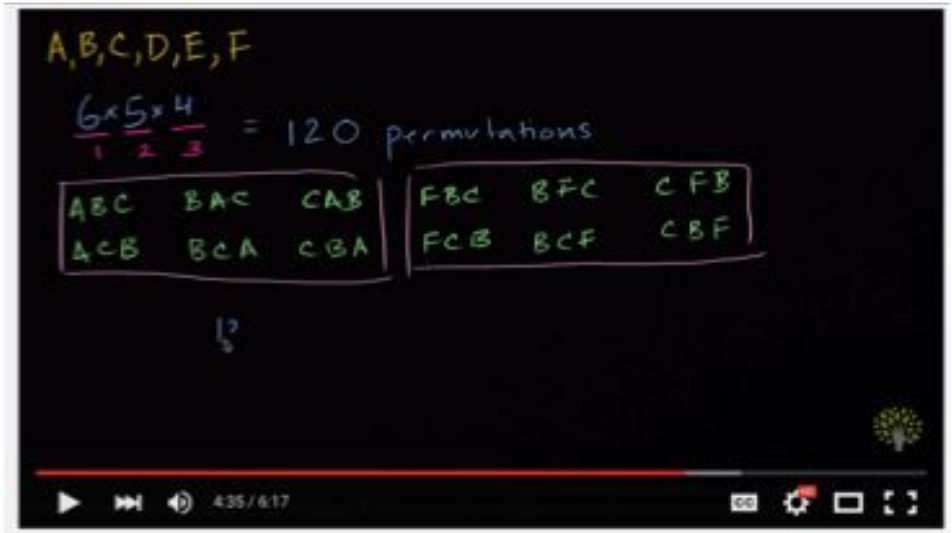

MOSTLY
HARMLESS
ECONOMETRICS
An Empiricist's Companion

Experiment:
Video vs. Text vs. No Math Instruction

- **Pre/Post Test – Difference in Differences**
- **Topic Selected**: Combinatorics
  - Standalone topic
  - Can be taught in 5-6 minutes
  - Hard to remember rules from school
- **Treatments**
  - **Video**: Khan Academy
  - **Text**: Transcription of Khan Academy Video with language simplification and visual aids embedded in text
- **Subjects**: Mechanical Turk, 151 total

Examples of Treatments

Video



Text

Combinatorics

This topic deals with figuring out the number of combinations that are possible given the following:

- How many options are there?
- How many objects or people need to choose those options?

For example, we could ask the question:

How many different ways are there to sit a 6 people in 3 chairs?

The 6 people are named "A", "B", "C", "D", "E", and "F".

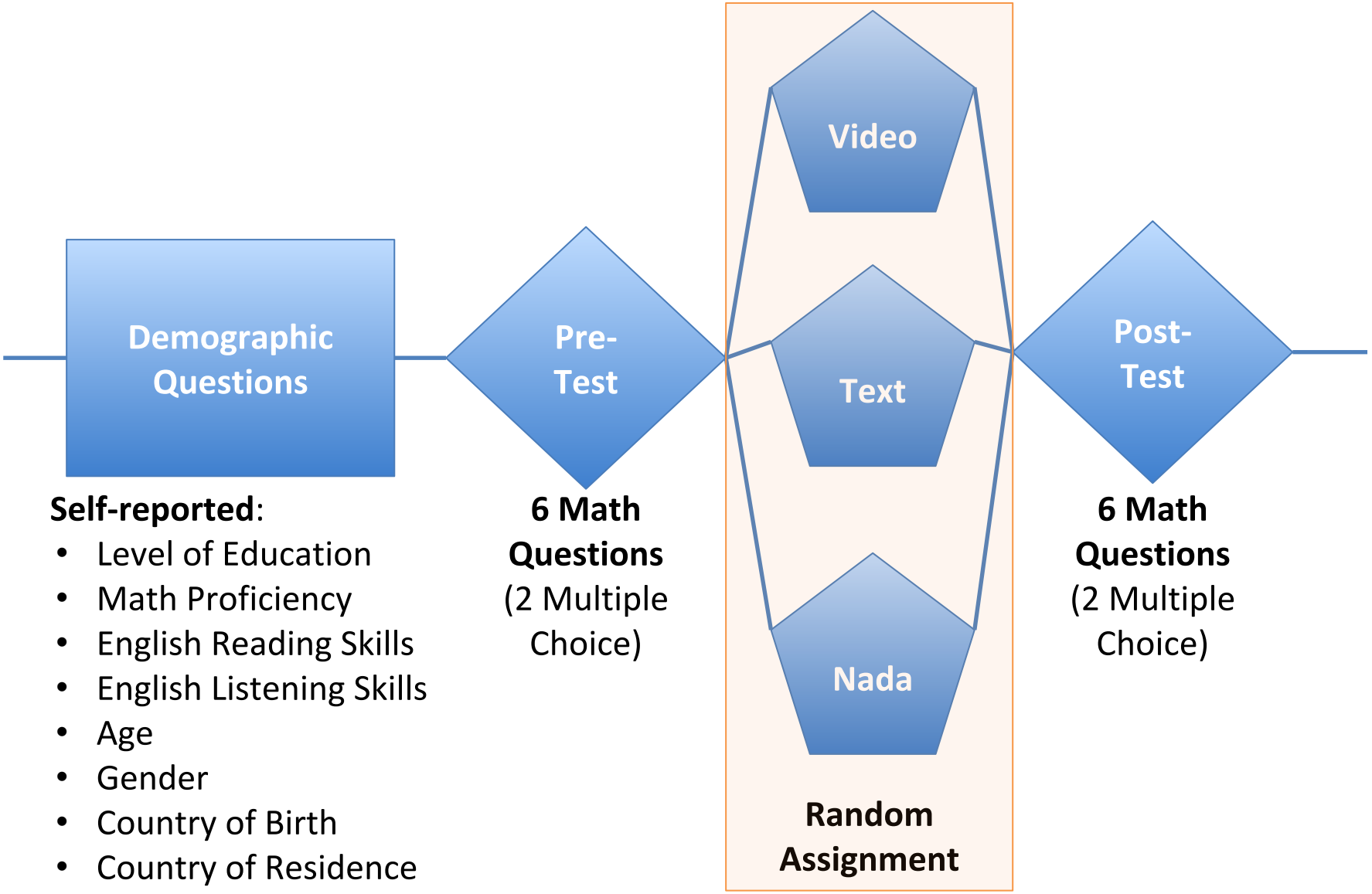
The chairs are numbered 1, 2, and 3.

What are all the **permutations** of putting six different people into three chairs?

Let's start with the first chair. If we haven't seated anyone yet, how many different people could we put in chair number one?

There are 6 different people who could be in chair number one. You could also say there are six different scenarios for who sits in chair number one.

Experimental Structure



Self-reported:

- Level of Education
- Math Proficiency
- English Reading Skills
- English Listening Skills
- Age
- Gender
- Country of Birth
- Country of Residence

6 Math Questions (2 Multiple Choice)

6 Math Questions (2 Multiple Choice)

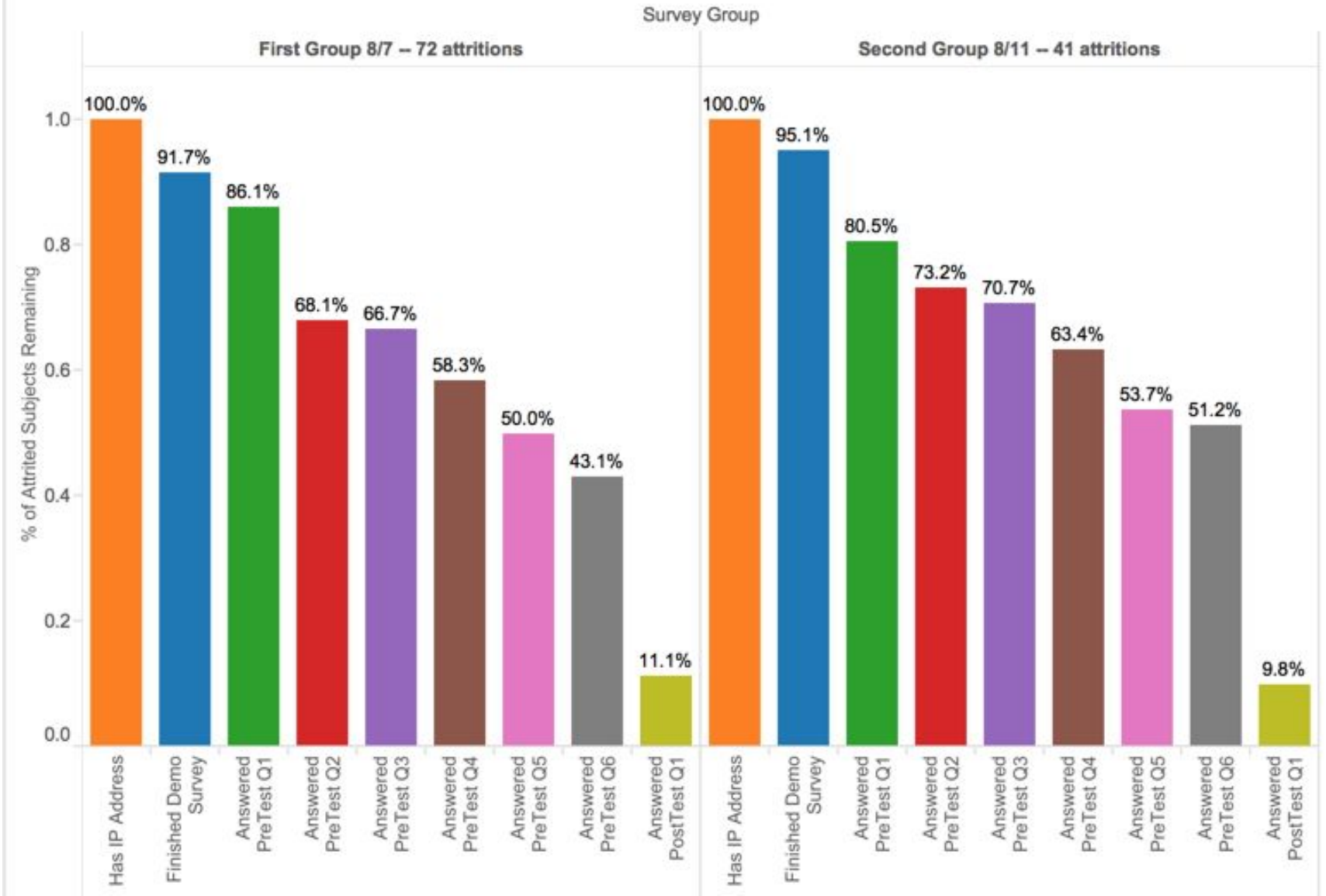
Additional Elements of Experiment

- Subjects Gathered through **Mechanical Turk**
- **Pay for Performance** (5 cents per correct answer on both Pre- and Post-Test + 1.50 for Taking the Survey)
- **Pilot Test** Conducted to Compare Sequences and Combinatorics Topics – Combinatorics Selected
- **Questions Matched** for Difficulty Between Pre- and Post-Tests
- **Deterred Random Guessing** by Timing Each Question/Treatment and Alert of Monitoring in Introduction
- Assumption that Googlers would get 100% on both tests

Problems Faced in Executing Experiment

1. Selection of the “Right” Topic and Content
2. Ensuring Fair Comparison of Video and Text
3. Ensuring Compliance (Not Random Guessing, Actually Consuming Treatment Content, Not Googling Answer)
4. Ensuring Similar Difficulty in Pre- and Post-Tests
5. What Subject Pool?
6. To Pay or Not to Pay for Performance?
7. Issues of Generalizability
8. Considering Appropriate Price Point for MTurk
9. Subject Browser Issues with Qualtrics Timer
10. Cleaning Data from Qualtrics for Analysis

Attrition


Survey Group
First Group 8/7 -- 72 attritions
Second Group 8/11 -- 41 attritions
% of Attrited Subjects Remaining
1.0
0.8
0.6
0.4
0.2
0.0
100.0%
91.7%
86.1%
68.1%
66.7%
58.3%
50.0%
43.1%
11.1%
100.0%
95.1%
80.5%
73.2%
70.7%
63.4%
53.7%
51.2%
9.8%
Has IP Address
Finished Demo Survey
Answered PreTest Q1
Answered PreTest Q2
Answered PreTest Q3
Answered PreTest Q4
Answered PreTest Q5
Answered PreTest Q6
Answered PostTest Q1

Demographic Comparisons Across Treatment Groups

Means for Numerical Covariates Across All Treatment Groups

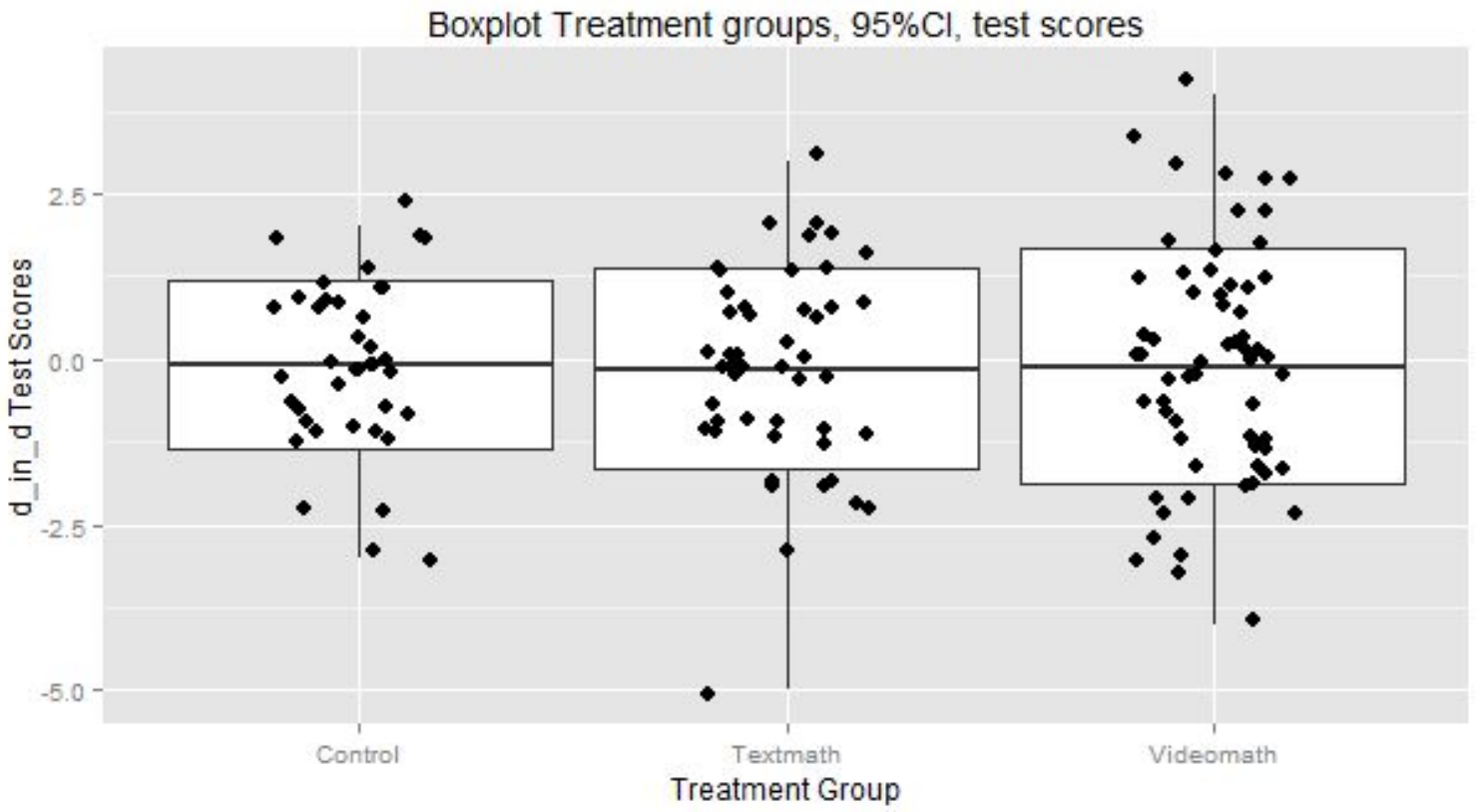
Covariate	Mean for Text Math group	Mean for Video Math group	Mean for Control group
Level of Education (1-7)	4.61	4.31	4.08
Math Skill (1-5)	2.80	2.94	2.48
English Reading (1-5)	4.61	4.63	4.53
English Listening (1-5)	4.65	4.68	4.48
Age in Years	36.04	35.65	36.30

Statistical Analysis of Difference in Differences

Regression Results

	Dependent variable:
	d_in_d
factor(treatmentgroup)textmath	-0.068
	(0.336)
factor(treatmentgroup)Videomath	-0.038
	(0.319)
Constant	-0.075
	(0.249)
Observations	151
R2	0.0003
Adjusted R2	-0.013
Residual Std. Error	1.575 (df = 148)
F Statistic	0.020 (df = 2; 148)
Note:	*p<0.1; **p<0.05; ***p<0.01

Statistical Analysis - Boxplot



Discussion of Results

1. No statistically significant difference between the treatment groups and the control group
2. No statistically significant difference between treatment groups
3. No Heterogeneous treatment effects
4. Higher variance among video subjects, but not statistically significant

Problems of Generalizability

Were the Mechanical Turks trying?

If some were not, adds noise

Hidden non-compliers may not be randomly distributed across treatment groups.

Even if they were trying, is the group representative?

Easier sample to gather, but different than a group of students whose grades rely on combinatorics

Areas for Further Research

- Similar test on **students** rather than Turkers
- Can any math subject be taught in 6.5 minutes? Perhaps need **higher dosage**.
- Include **similarly timed control** to make the surveys of equal length
- Consider **alternate topics** for video vs text vs control comparison