

# Understanding Student Use of AI Tools Surveys to Experiments

Rahul Divekar, Sophia Guerra, Lisette Gonzalez, and Natasha Boos

# Overview

- Many students are using ChatGPT and other AI tools for learning
- Our research question:
  - Do richer interactions with AI (LLMs) lead to better learning outcomes than search engines (Google)?

# How we designed the Experiment

- Survey asking students (68)
  - age and other demographics
  - which tools students actually use (chat, youtube, google, etc.)
  - what they use chatgpt for and how often
  - major and minor + GPA
  - how would they define their expertise level in using AI in their work



# Results of the survey

- Students used 3.64 tools on average
- Most common tools: Google, ChatGPT, Youtube, Wikipedia
- 100% of our sample had used ChatGPT
- Main uses of ChatGPT:
  - learning difficult concepts
  - writing and editing essays, emails, or resumes
  - brainstorming ideas or outlines
  - coding and data analysis
  - expertise to them = ability to prompt well, spot errors and apply the knowledge

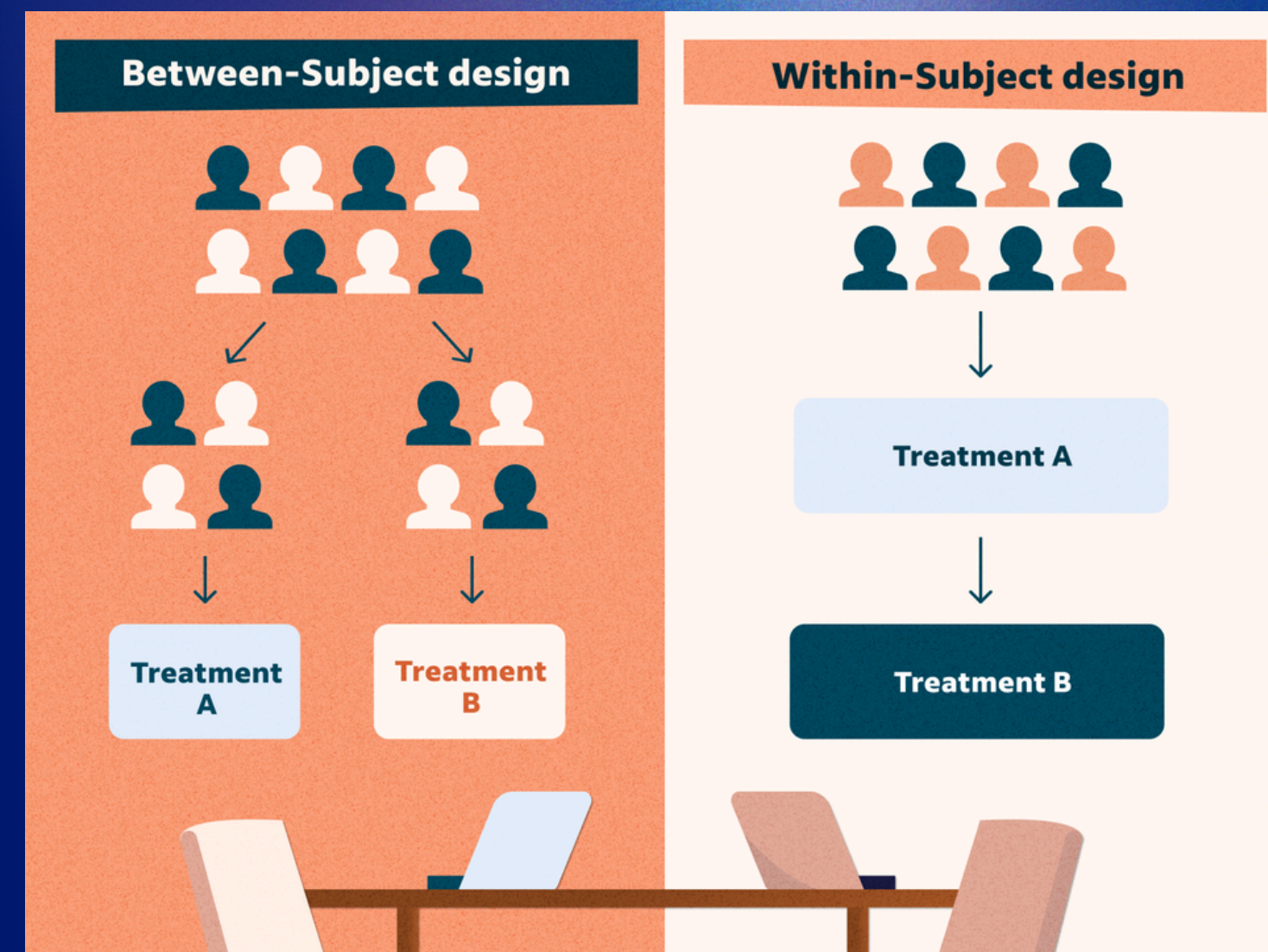
# Survey limitations

- Only 68 Bentley students → limited sample
- Self-report bias (students may misremember or exaggerate use)
- Observational → cannot show cause and effect
- Tool familiarity varies between students

# In order to be selected from the survey to be in the experiment

- Have experience with ChatGPT or an AI Tool
- Have experience utilizing a search engine
- Are technically proficient
  - Experience with word processing programs, internet, and computers
- Are university students actively enrolled in an undergraduate or graduate program
  - Are not enrolled in the MS Human Factors in Information Design
  - Are not enrolled in the BS Information Design and Corporate Communication
- Are above the age of 18

- Within-group design
  - each student tried both Chat and Google
- 20 Bentley students
- topics
  - how does the internet work
  - how does the power grid work
- random assignment
  - some students started with ChatGPT and others with Google



# Independent Variable

- Level 1: ChatGPT (LLM)
- Level 2: Google (Search engine)

# Dependent Variables

- Process measures (prompt length, number of queries/visits)
- Output measures (essay complexity, plagiarism rate)
- Outcome measures (graded performance / learning gain)

# Steps of the Experiment

## Pre-test Essay

- ▶ Each student wrote a short essay about the topic given without using any tools

## ChatGPT or Google

- ▶ They researched the topic with either chat or google and took notes which they had time afterwards to write an essay using the notes

## Switch conditions

- ▶ Students repeated the process on the second topic with the other tool

# Experiment limitations

- Small sample size (N=20) → low statistical power
- Short duration tasks → no measure of long-term learning
- Only 2 topics tested (Internet & Power grid)
- One version of ChatGPT → results has changed as tech evolves
- Plagiarism detection may bias against search-based essays
- Lab setting ≠ real-world student use

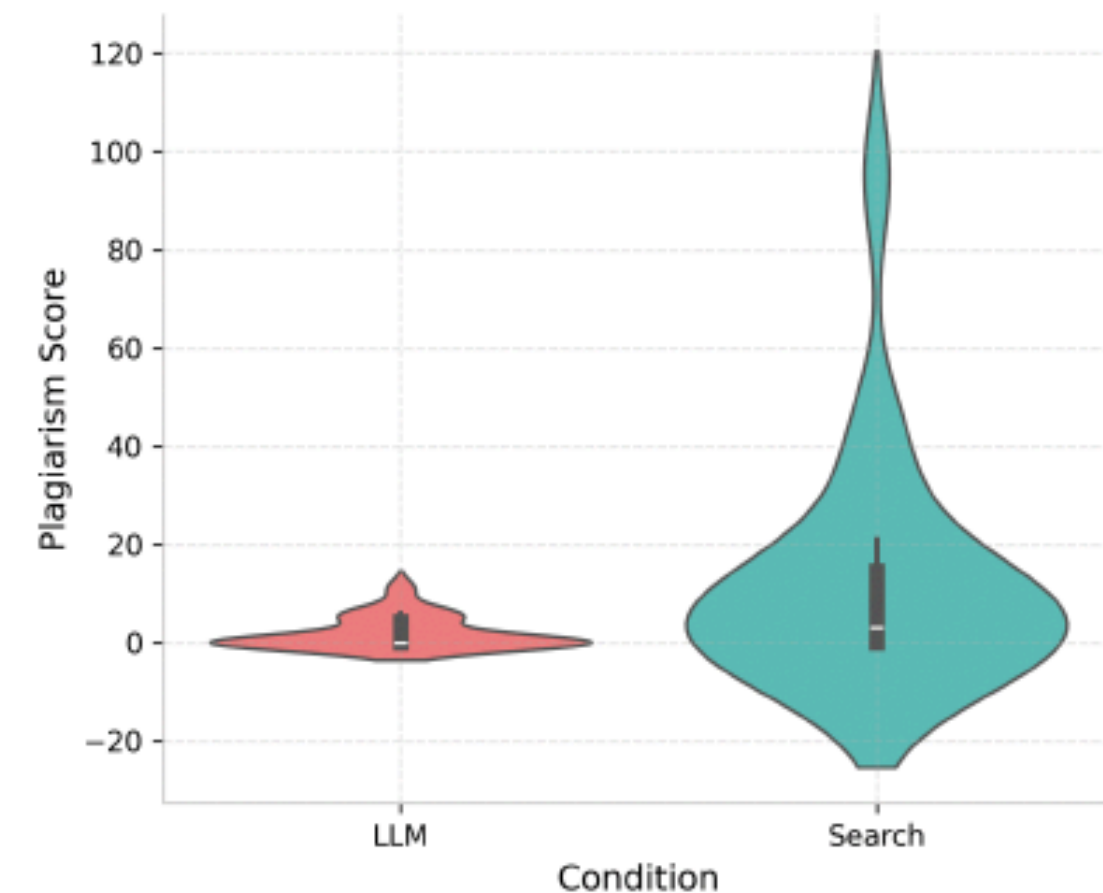


Fig. 4. Plagiarism Scores of Final Output across Two Conditions

# Interview

- All students were interviewed post completing all essays
- they were asked questions like:
  - Do you consider yourself an expert user of traditional search engines like Google
  - Do you perceive the AI-powered chat-like tool (e.g., ChatGPT) primarily as a search tool, an assistant, or something else?
  - How much of the information presented in the essay and comprehension quiz do you feel was your own contribution?

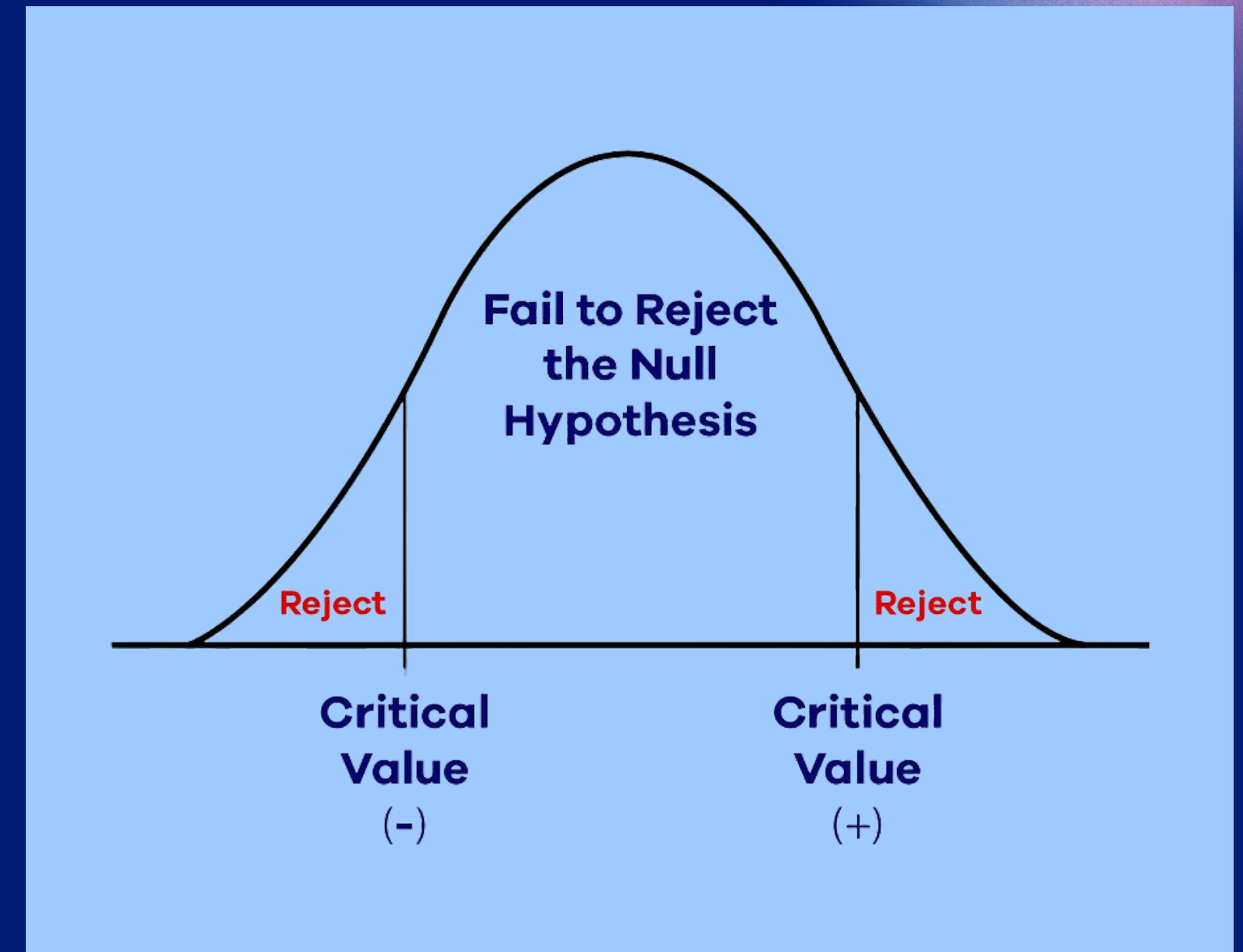
# Post Experiment

- All meetings with participants were recorded
  - both what they were saying and the screen
- Co-authors went through interviews and coded it throughout
  - some examples of codes:
    - golden quotes
    - learning easy google / learning easy gpt
    - preference google / preference gpt
    - ownership gpt 1-2 / 3-5 / 6-7 / other
    - satisfaction using gpt
      - satisfaction using gpt 1-2 / 3-5 / etc.

# Some Results



- Ran ANOVA (analysis of variance)
- Result: No main effect of tool type (LLM vs. Search) on grades
- $p\text{-value} > 0.05 \rightarrow$  Fail to reject null hypothesis



# Some Results



Process

Outputs

Outcomes

ChatGPT had longer, more conversational prompts

no difference in complexity  
Google → more plagiarism flags

Both tools improved learning compared to the pre-test

Google had short keyword inputs with more link clicking

ChatGPT → more unique wording but not better quality

no significant difference in final performance

Table 4. Categorization of Input Query Text and Across LLM and Search Tools

Category	Description	LLM Count	Search Count	Dominance Score
How/why Question	Text contains a how or why question. E.g., "how does the internet work"	41	25	0.24
Output Control	Text attempts to control output. E.g., "Make sure you start with...", "elaborate"	12	0	1
Reference to Previous Text	Text attempts to refer to previous output. E.g., "can you take all the information you just gave me and put it in one flowing, cohesive essay together"	11	0	1
What question	Text contains a what question. E.g., "what is the internet"	30	21	0.18
Difficulty Control	Text attempts to control difficulty of output. E.g. "In the most simplest terms."	15	7	0.36
Keywords	Text contains only keywords rather than full sentences. E.g., "wind city"	0	4	-1
Context	Text provides more information about the larger goal. E.g., "This is my essay topic...", "... your audience does not know anything about the topic"	2	0	1
Request Example	Text requests an example. E.g., "explain with an example how..."	2	1	0.33
Multimedia Request	Text requests multimedia. E.g., "... educational video"	0	1	-1

