Examining the Impact of an AI-powered Writing Platform in Upper-division Engineering Courses

Ruilin Wu
Advisors: Dr. Mark Warschauer, Jaeyoon Choi

UCI CP-LEADS

Digital Learning Lab

Introduction

Background

- Generative AI's surge in education demands a secure, scaffolded platform that preserves critical thinking, builds AI literacy, and supports instructors.
- PapyrusAI, built on GPT-4, serves as a Socratic tutor using curated and customizable prompts to guide students through topic development, outlining, and drafting.
- UCI upper division engineering writing courses used PapyrusAI to give students real-time, rubric-aligned feedback to refine their topics, outlines, and arguments.

Research Question

To what extent are students' final writing scores related to (a) the number of back-and-forth exchanges with PapyrusAI and (b) the quantity of their chat inputs?

Data

- Participants
  - **188** undergraduates enrolled in ENGR190W at UCI in Spring 2024
  - **144** valid student data for analysis
- Interactions
  - A total of **8,343** student chat messages with PapyrusAI

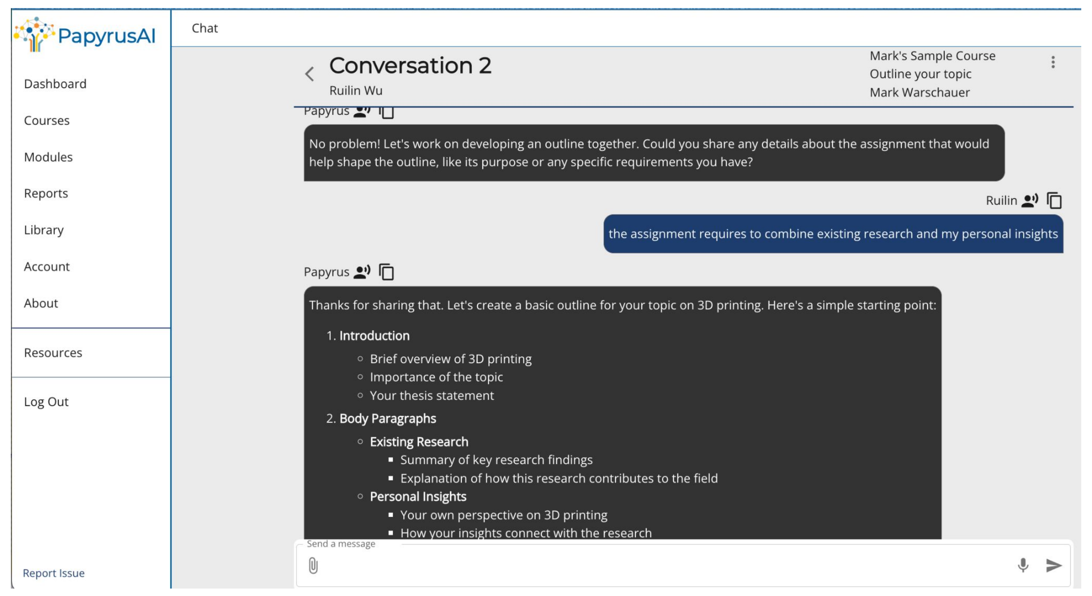

Figure 1. Example of one conversation at PapyrusAI

Research

Methodology

Variables & Measures

- Outcome
  - Final course score (0 - 100)
- Engagement metrics
  - Total conversations students exchanged
  - Mean number of sentences, words, characters per conversation
  - Mean response time (seconds)
  - Student's mean conversation per module

Multivariate Linear Regression

- Fit an OLS model:

$$\text{final\_score} = \beta_0 + \sum_{k=1}^{6} \beta_k X_k + \varepsilon$$

- β-coefficients with 95% confidence interval
- Multicollinearity

Relative Importance (LMG)

- Partition model R^2 to quantify each predictor's contribution

Result

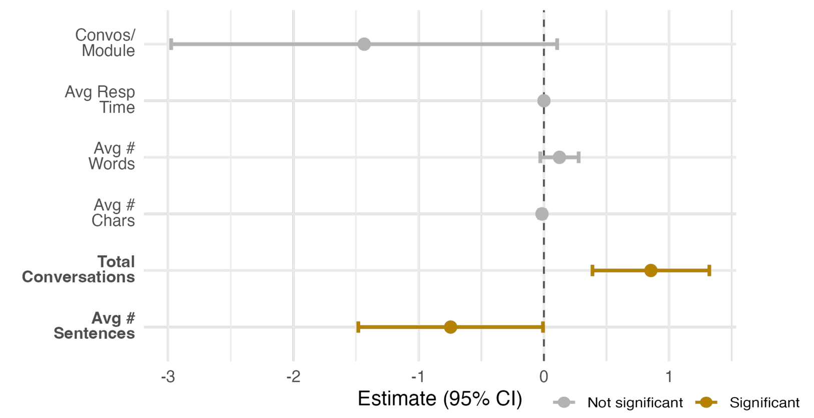

Figure 2. Forest plot of β estimates (95 % CIs) for six engagement metrics predicting final score

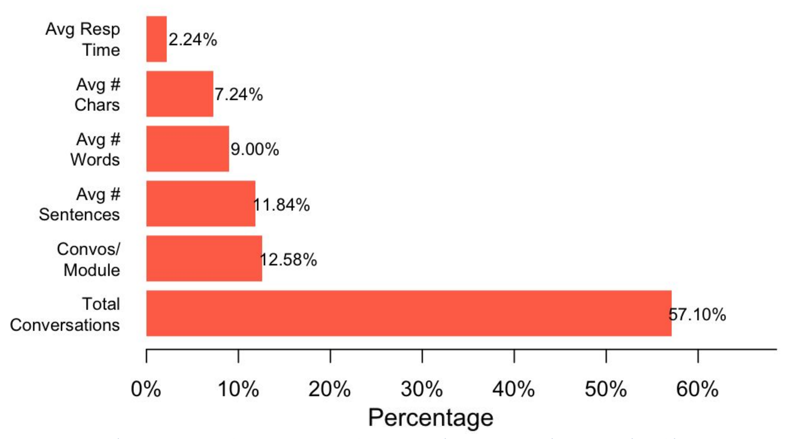

Figure 3. Percent of explained variance in final score contributed by each engagement metric

Significant Variables

- The total count of exchanges initiated by the student with the AI
- The average number of sentences students wrote in each message to the AI

	intercept	total_convos_count	avg_stu_char	avg_stu_word	avg_stu_sentence	avg_response _time	mean_convo_per_module
Estimate	84.40	0.854	-0.016	0.124	-0.74	0.000008	-1.435
P-value	< 2e-16***	0.0004***	0.172	0.111	0.048*	0.000013	0.068
VIF		2.693	792.339	1118.01	50.131	1.013	2.435
Observations 144; R^2 0.192; Adjusted R^2 0.157							

*Note: *p < 0.10; **p < 0.05; ***p < 0.01*

Table 1. OLS Regression Result of Student's Final Score

Analysis

- Although the R-squared is modest, values of 0.10 or higher are generally considered acceptable in social science research if some or most of the predictors or explanatory variables are statistically significant(Ozili, 2022).
- Greater AI interaction frequency is positively associated with higher grades.
- While the length of individual AI interactions did not correlate with improved grades, students with higher grades typically used fewer sentences on average.

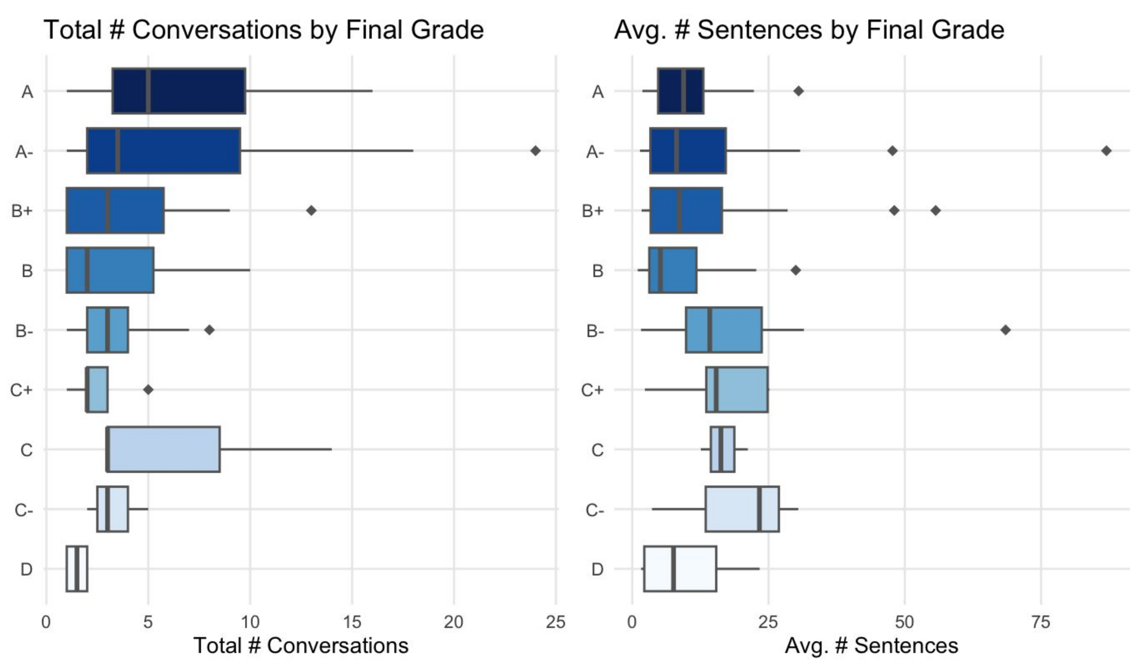

Figure 4. Engagement metrics across letter grade

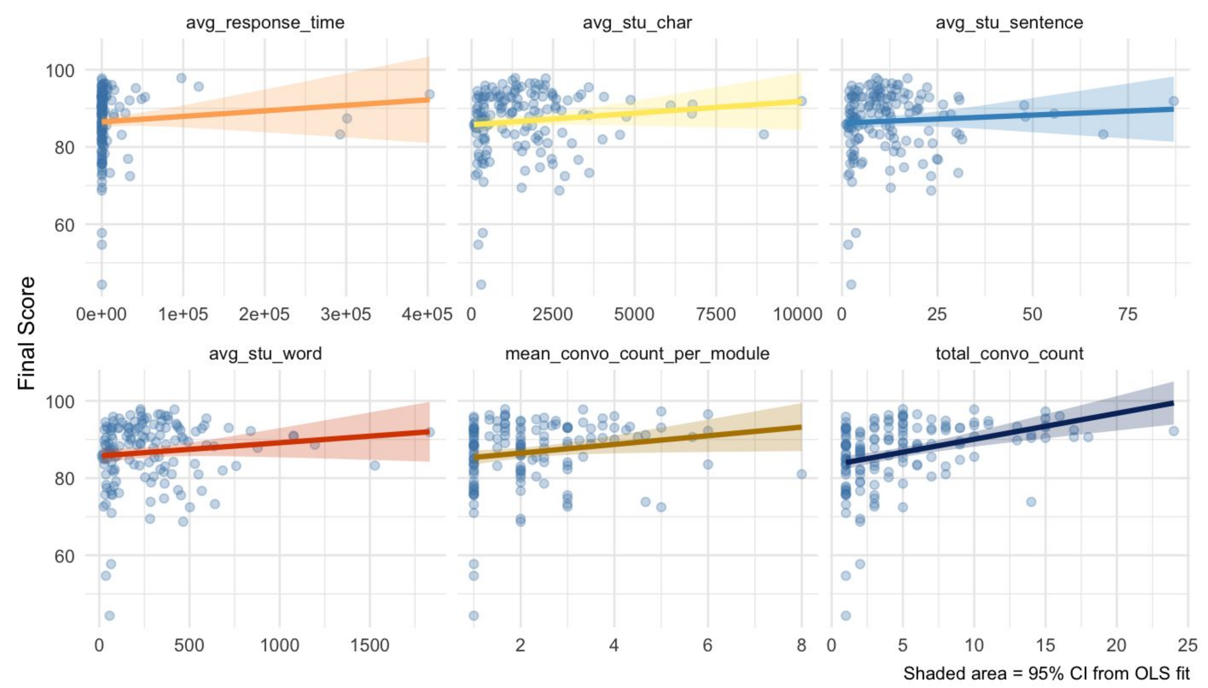

Figure 5. Association between predictors and final scores

Conclusion

Discussion

- Interacting with AI can be conceptualized as a form of retrieval practice (Karpicke & Blunt, 2011):
  - High-performing students engaged with the AI more frequently and used concise language.
  - Lower-performing students interacted less often but supplied longer, more elaborate inputs per conversation on average.

Implications

- Equip instructors with scaffolded, targeted prompts to help underperforming students formulate more effective AI queries.
- The significant effect of avg_stu sentence indicates that the way students structure and segment their thoughts may reflect a level of engagement and coherence that is more conducive to learning.

Limitations

- Potential confounding variables were not captured, which may affect model efficacy.
- A larger, more diverse dataset is required to validate these findings.
- Inclusion of a randomized control group would strengthen causal claims.

Future Direction

- Quantitative coding of chat logs to categorize AI-interaction patterns and uncover student support needs.
- Stage-wise analysis of AI engagement's relationship with writing performance across ideation, drafting, revision, and finalization.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No.23152984. Thanks the Digital Learning Lab for providing the data, and Dr. Mark Warschauer and Jaeyoon Choi for their invaluable guidance.

Reference

[1] Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

[2] Ozili, P. K. (2022, June 5). The Acceptable R-Square in Empirical Modelling for Social Science Research. Retrieved from papers.ssrn.com website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4128165