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

Ruilin Wu
Advisor: Dr. Mark Warschauer, Jaeyoon Choi

Digital Learning Lab

Introduction

- Generative AI's surge in higher 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.
- A pilot study in UCI engineering writing courses used PapyrusAI to give students real-time, rubric-aligned feedback to refine their topics, outlines, and arguments.

Research Question

How does the use of PapyrusAI vary across instructors, and what impact does this have on student engagement and final writing scores?

Methodology

Participants:

- **317** undergraduates enrolled in ENGR190W at UCI across four quarters between two instructors.
  - Spring 24, Fall 24, Winter 25, Spring 25.
- Students generated a total of **22,914** chat messages with PapyrusAI.

Analytical approach:

- Multivariate Linear Regression.
- Correlation coefficients to examine relationships between PapyrusAI use and course performance.
- Focused on two instructors who implemented PapyrusAI differently across quarters.

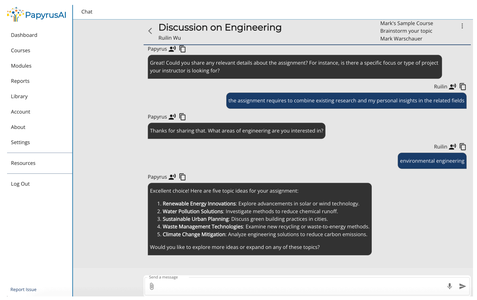


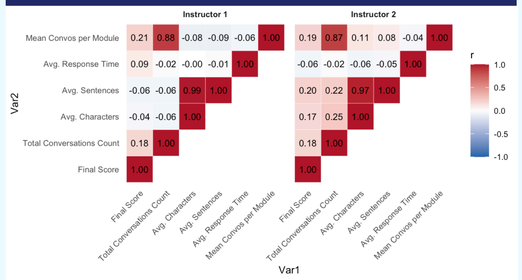
Figure 1. Example of one conversation at PapyrusAI

Research

Linear Regression Models Predicting Final Score from PapyrusAI Usage

Predictor	Whole Set	Instructor 1	Instructor 2
Total Conversations Count	-0.0011	-0.0007	-0.0089
Avg. Student Characters	0.00002	0.00066	-0.00083
Avg. Student Sentences	0.091	-0.091	0.149
Avg. Response Time	0.000005	0.000006	-0.000005
Mean Convos per Module	0.475	0.471	0.505
Adjusted R^2	0.026	0.030	0.046
N	287	148	139
F-statistic (p-value)	2.50 (0.031)	1.92 (0.095)	2.31 (0.047)

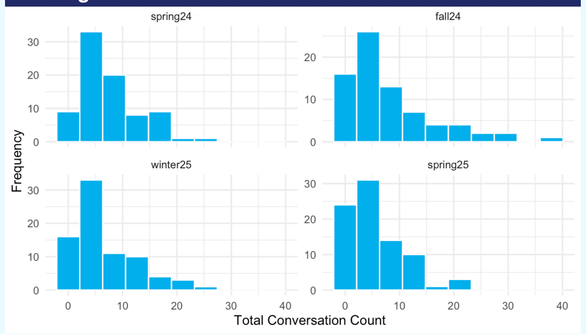
Correlation Matrices by Instructor



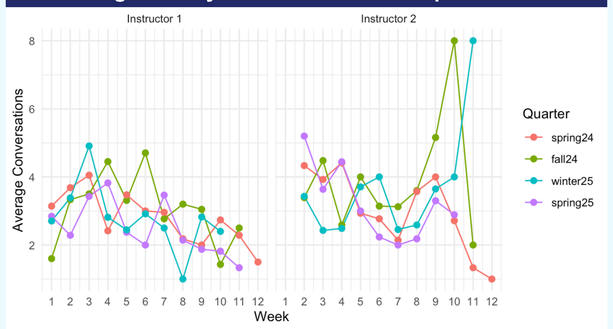
Analysis

- Both correlation and regression analyses show that AI usage metrics have weak but noticeable links to final scores. Interaction quantity (e.g., conversation counts) shows mild positive correlations ($r \approx 0.18$–0.21) and small regression effects, indicating limited predictive strength.
- Across instructors, Mean Conversations per Module consistently emerges as the strongest positive predictor ($\beta \approx 0.47$–0.51). This suggests that students who engage consistently across modules, rather than simply more often overall, tend to achieve slightly higher scores.
- Metrics such as Average Characters, Average Sentences, and Response Time show minimal correlations ($r \approx 0$ to 0.09) and weak or inconsistent regression coefficients — implying that longer or quicker messages do not necessarily lead to better outcomes.
- Both instructors show similar correlation patterns, but Instructor 2's model explains more variance ($R^2 = 0.046$), suggesting that course design or integration approach may influence how AI engagement relates to learning outcomes.

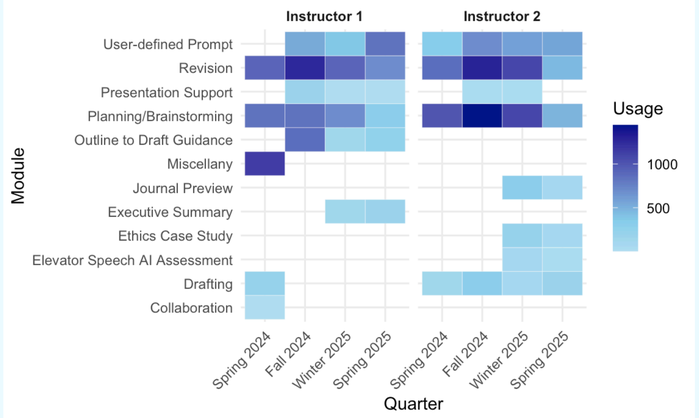
Histogram of Total Conversations a Student Produced



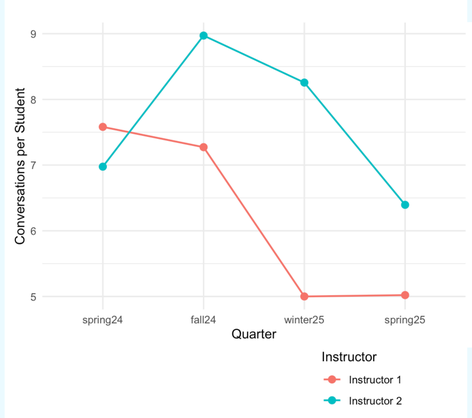
Average Weekly Conversation Count per Student



Module Usage Heatmap



Average Conversations per Student by Quarter



Discussion

- Quantity of AI interactions = modest predictor
  - Frequency of chats alone explains little variance in final scores.
- Instructional context and module design matter more
  - How instructors integrated PapyrusAI and which modules were emphasized strongly shaped engagement.
  - Instructor 2's focus on drafting linked to stronger correlations with performance.
- Planning and revision = consistent student needs
  - Repeated across quarters and instructors, highlighting persistent challenges in engineering writing.
  - Aligns with writing research showing students struggle most at early- and mid-stage drafting.

Limitation

- No control group for direct comparison.
- Enrollment numbers varied substantially across quarters.
- Possibility that students also used external AI tools outside of PapyrusAI, which was not measured.

Future Direction

Research should investigate:

- How instructors can strategically integrate AI tools into writing pedagogy.
- The balance between student autonomy and structured guidance.
- How module-level design choices affect engagement and outcomes.

Important to address equity and ethics, including:

- Unequal access to AI tools across student populations.
- Potential disparities in engagement and benefits.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No.23152984.