

Optimizing Language-Focused Writing Feedback from Large Language Models through Prompt Engineering

Large Language Models

prompt engineering

Writing Feedback

Yu Tian, Andrew Potter, Jessica Early, Steve Graham, Parul Patil, Katerina Christhlf, Motahareh Darvishpour Ahandani, and Danielle McNamara

Arizona State University; {ytian126, ahpotter, jessica.early, steve.graham, ppatil57, kchristh, mdarvis2, dsmcnamas}@asu.edu

High-quality, individualized feedback is essential for supporting the development of students' academic writing skills, yet instructors face persistent challenges delivering such feedback at scale in writing-intensive courses. Large language models (LLMs) have the potential to provide timely and personalized feedback, but the pedagogical quality of this feedback is highly sensitive to the design of the prompt (Dai et al., 2024). Despite growing adoption of AI writing support tools, little empirical research has examined which prompting strategies most effectively elicit accurate and actionable feedback.

This project investigates how different prompt engineering strategies affect the quality and instructional effectiveness of language-focused writing feedback. Four prompting conditions were implemented and evaluated:

1. Baseline Prompt: providing a generic request for writing feedback.
2. Rubric-Augmented Prompt: embedding an analytic rubric targeting language features.
3. NLP-Feature-Informed Prompt: incorporating natural language processing (NLP) metrics such as lexical diversity and syntactic complexity.
4. Combined Prompt: integrating rubric criteria with NLP metrics.

Undergraduate students (N = 113) completed a timed narrative writing task and a revision session informed by LLM feedback assigned under one prompting condition. A panel of writing instructors independently evaluated the generated feedback and students' drafts and revisions using custom-developed rubrics.

Initial analyses suggest that augmented prompts (i.e., rubric-based, NLP-informed, and combined prompts) generate more development-oriented feedback focused on improving language sophistication and clarity, while the baseline prompt tends to produce predominantly accuracy-focused comments. We expect these augmented prompts to lead to more meaningful, development-related revisions and greater improvements in students' language quality.

This study advances learning engineering by providing empirical guidance on how prompt design shapes the pedagogical value of LLM-generated writing feedback. The findings of this study can inform the development of scalable writing support systems that adapt to learners' needs and prompt deeper engagement with language features that matter for academic writing. More broadly, the work demonstrates a systematic, evidence-driven approach in learning engineering (Baker et al., 2022) to optimizing human-AI interaction for complex learning tasks such as academic writing.

References

Baker, R. S., Boser, U., & Snow, E. L. (2022). Learning engineering: A view on where the field is at, where it's going, and the research needed. *Technology, Mind, and Behavior*.

Dai, W., Tsai, Y. S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, 7, 100299.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305T240035 to Arizona State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

