

LONGITUDINAL INSIGHTS FROM STUDENTS' SCIENCE REPORTS THROUGH AI-DRIVEN DATA ANALYSIS

James D Cleaver^a, Laura K. McKemmish^a, Sara H. Kyne^a

Presenting Author: James D Cleaver (j.cleaver@student.unsw.edu.au)

^aSchool of Chemistry, Faculty of Science, University of New South Wales, Sydney NSW 2052, Australia

KEYWORDS: Authentic Science Inquiry, Education Data Mining, Natural Language Processing, STEM Education Research, Student Research

SUBTHEME: Equity, Diversity and Inclusion

Student authentic science inquiry (ASI) reports are written investigations produced by secondary students as part of open-ended research projects completed as part of the science curriculum aligned with classroom activities. These reports are a rich source of data, offering insight into students' interests, their framing of scientific problems, and the way they communicate ideas. This makes reports especially valuable for exploring student engagement with science in secondary school.

This research presents a methodological approach for analysing student-ASI reports. The research draws on a multi-year corpus of de-identified reports submitted by Australian secondary students to state science competitions between 2015 and 2024. These reports contain rich evidence of students' interests, discipline choice, problem-framing, and scientific communication as they undertake ASI. The corpus reflects a wide diversity of school contexts and student voices, including projects submitted by students from regional areas and schools serving low socio-economic communities, making it a valuable source of data for equity-aware analysis.

Using natural language processing (NLP) and generative AI (GenAI) techniques, the research applies topic modelling, semantic clustering, and sentiment analysis (Chew et al., 2023) to examine the types of questions students pose, the ways they present evidence, instances of uncertainty in language, and how their scientific reasoning is conveyed through text. These techniques allow researchers to identify recurring themes, and capture patterns in students' engagement with scientific inquiry. These techniques support the analysis of report data to identify longitudinal trends/changes and enable comparisons across demographic and geographic contexts.

This research responds to growing evidence that ASI fosters student engagement, agency, and science identity in both secondary and tertiary settings (Marley et al., 2022; Sadler et al., 2010). The research leverages emerging capabilities in education data mining and AI-supported qualitative data analysis (Chew et al., 2023; Christou, 2024). By centring student reports as a primary data source, the research contributes a scalable method for investigating how students describe their ASI experience longitudinally. The approach offers insights for curriculum developers, educators, and program designers seeking to understand and support diverse student participation in STEM.

REFERENCES

- Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). *LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding* (No. arXiv:2306.14924). arXiv. <https://doi.org/10.48550/arXiv.2306.14924>
- Christou, P. (2024). Thematic Analysis through Artificial Intelligence (AI). *The Qualitative Report*, 29(2), 560–576. <https://doi.org/10.46743/2160-3715/2024.7046>
- Marley, S. A., Siani, A., & Sims, S. (2022). Real-life research projects improve student engagement and provide reliable data for academics. *Ecology and Evolution*, 12(12), e9593. <https://doi.org/10.1002/ece3.9593>
- Sadler, T. D., Burgin, S., McKinney, L., & Ponjuan, L. (2010). Learning science through research apprenticeships: A critical review of the literature. *Journal of Research in Science Teaching*, 47(3), 235–256. <https://doi.org/10.1002/tea.20326>