ASCILITE 2024

Navigating the Terrain:

Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies

AI-Assisted Pedagogies: Enhancing Mathematical Literacy and Open-Ended Problem-Solving with ChatGPT

Paul Trotter, Michelle Vanderburg, Robert Vanderburg
Central Queensland University

This paper advocates using large language models (LLMs) to enhance students' open-ended mathematical problem-solving skills. The literature review reveals that using large language models (LLMs) in educational settings is an emerging field with limited research evidence. Much of the existing literature focuses on the perceptions of educators and other stakeholders and evaluations of LLM model accuracy rather than their potentially transformative impact on educational outcomes or personalised learning in mathematics. This paper reports on the ongoing development of a study to provide such evidence. It presents expected results derived from our preliminary development work, which suggests that teaching students to utilise LLMs like ChatGPT effectively can enhance their critical thinking and analytical skills in mathematical problem-solving. Initial observations and iterative refinements inform these anticipated outcomes during the pilot phase of our research.

Keywords: Large Language Models (LLMs), ChatGPT, Artificial Intelligence (AI) in Education Mathematical Problem-Solving, Open-Ended Problems, AI-Assisted Learning

Introduction

The advent of large language models (LLMs) such as ChatGPT heralds a potentially transformative era in education (Wardat et al., 2023), promising to revolutionise how students learn and interact with information. These advanced AI systems, capable of generating human-like text and understanding complex queries (Lo, 2023), offer unprecedented opportunities to personalise learning experiences, provide instant feedback, and support diverse educational needs. With the potential to act as virtual tutors (Chen & Leitch, 2024; Pardos & Bhandari, 2024), LLMs may provide educators with a powerful tool to augment traditional teaching methods. However, alongside these promising potential benefits, integrating LLMs into educational settings raises concerns (Kasneci et al., 2023). Issues such as academic integrity, the potential for misinformation, and other ethical implications are born from the promise of an AI-driven learning environment (Spennemann et al., 2024). If harnessing this technology's full potential is to be realised, it necessitates a careful and balanced approach to its introduction into our classrooms. While the potential for AI to transform learning is evident (Ansari et al., 2023; Guan et al., 2023; Lo, 2023), there is currently limited evidence of its measurable impact in the research literature. Given the emergent and embryonic nature of large language models (LLMs) in education (Lo, 2023), this article must draw on research that utilises LLM research from contexts beyond mathematics instruction to understand their potential applications and implications comprehensively. This is unsurprising given the rapid pace of technological change within the sphere of LLMs, and their recent emergence into the educational discourse. Ongoing research is essential to understand and fully harness the capabilities of LLMs such as ChatGPT.

The challenge of supporting problem-solving in mathematics

Reading comprehension is critical to students' ability to solve mathematical problems, particularly word problems (Cartwright et al., 2022). Mathematical word problems require students to read and understand the text, identify relevant information, and translate that information into mathematical equations or expressions (Whimbey et al., 2013). This process is inherently tied to reading comprehension skills. Research has shown that students' reading comprehension proficiency directly influences their performance in solving math word problems (Pongsakdi et al., 2020; Prabowo et al., 2023). Students who need help with reading comprehension are less likely to understand the problem's context and the specific mathematical tasks required, leading to

difficulties in mathematical problem-solving.

Furthermore, the ability to navigate lexical ambiguities, where words have different meanings in mathematical contexts compared to everyday usage, also plays a significant role (Cartwright et al., 2022). Students must be able to discern these differences to interpret and solve word problems accurately. Therefore, enhancing students' reading comprehension skills provides a potential route to significantly improving their mathematical problem-solving abilities (Pongsakdi et al., 2020; Prabowo et al., 2023).

Literacy support can significantly enhance mathematics performance, particularly in solving mathematical word problems (Cartwright et al., 2022; Pongsakdi et al., 2020; Prabowo et al., 2023). One approach found to be effective is computer-assisted instruction that integrates contextually enhanced reading interventions (Nair & Dubé, 2021). These interventions have led to greater mathematical word problem-solving gains than standard reading programs, as they help students better comprehend and process the language used in math problems. Additional evidence comes from Swanson et al. (2019), who developed a paraphrasing intervention that has proven beneficial, especially for English Language Learners (ELLs) with math difficulties. Teaching students to paraphrase relevant information improved their accuracy in solving mathematical word problems. This approach helped both monolingual and ELL students by making the problem-solving process more accessible and understandable. Techniques such as using keywords, visualisation, and graphic organisers can aid students in breaking down and comprehending complex word problems (Gates, 2018). Overall, integrating literacy support into math education would seem to aid in comprehension and empower students to tackle mathematical challenges more effectively, thereby improving their overall math performance.

Ongoing research is essential to fully understand and harness the capabilities of technologies in higher education (Vanderburg et al., 2023), such as large language models (LLMs). This report investigates current literature on using LLMs, such as ChatGPT, as interactive literacy support tools for mathematics students tackling open-ended problems. LLMs may help students access and engage with problems more effectively by enabling them to ask questions and clarify meanings. Considering the nascent stage of large language models (LLMs) in educational settings (Lo, 2023), this article will need to reference research from various fields beyond mathematics instruction to explore their potential uses and impacts thoroughly. In addition, this report presents preliminary findings from an ongoing action research project that seeks to explore the potential of LLMs to enhance problem-solving skills in mathematics. Furthermore, we propose further research to evaluate the efficacy of this technology in positively impacting student outcomes.

Examining the potential of AI-assisted learning

A literature review by Chung Kwan Lo (2023) provided a comprehensive overview of ChatGPT's capabilities, applications, and challenges in the educational context. This overview reflected the diverse nature of this emergent field of study, discussing the accuracy and performance aspects of LLMs, their potential applications, and the ethical challenges to education. These diverse themes are similar to a review by Guan et al. (2023), who noted that ChatGPT tended to excel in critical and higher-order thinking, particularly in economics, but performs less satisfactorily in fields like law, medical education, and mathematics. However, LLM models exhibit varying performance across domains, which can fluctuate as models evolve (Giannos & Delardas, 2023). Consequently, assessing an LLM's accuracy at any given moment concerning specific performance criteria becomes challenging, especially with the rapid pace of technological evolution in this field.

These findings may call the efficacy of LLM’s as a tool to support mathematics learning into question. However, Zong & Krishnamachari (2023) focused on testing the accuracy of GPT-3, GPT-3.5, and GPT-4 models in solving math word problems (MWPs). The models were evaluated on their ability to classify word problems, extract equations from word problems, and generate new word problems. GPT-4 demonstrated the highest success rates, achieving 100% success in several categories. This study also highlighted several areas for improvement to achieve more reliable and accurate performance. Research on the performance of LLM’s underlines the imperative that students need to be taught to critically evaluate the output of LLMs, as these

models can occasionally produce inaccurate or biased information.

It is crucial to consider the nature of the limited empirical evidence regarding LLM technology implementation in educational settings (Ansari et al., 2023), as it becomes apparent that much of it centres on community perceptions of the technology and their views on its potential applications. This focus on subjective viewpoints highlights the need for more comprehensive and objective research to fully understand the impact and effectiveness of technological tools in enhancing personalised learning experiences. Zheng (2023) reports on the findings of a study utilising ChatGPT in teaching a data science course. Multiple benefits, including programming assistance, help with understanding new concepts, and data analysis, were cited as potentially benefiting the involved students. Benefits to educators and several potential challenges associated with the technology were also reported. However, caution needs to be exhibited as these findings were based on a survey of 28 graduate students and their instructors who filled out a questionnaire of 10 Likert scale questions.

An empirical study involving an extensive analysis of Twitter data (Fütterer et al., 2023) highlighted diverse global reactions to integrating AI in educational contexts. It revealed that while there is significant enthusiasm and optimism about the potential of AI tools like ChatGPT to revolutionise learning and teaching, there are also notable concerns regarding ethical implications, data privacy, and the potential for exacerbating educational inequalities. The analysis underscored the importance of addressing these concerns through transparent practices and inclusive policies to ensure that AI education innovations are effective and equitable. Similarly, a qualitative case study approach (Wardat et al., 2023) explored the perspectives of various stakeholders, such as students and educators, on using ChatGPT in teaching mathematics. While insights into stakeholders' perceptions are important and potentially illuminating, they also highlight the present limitations of the empirical research base, which seems focused on human reactions to LLM technology rather than its measurable efficacy. This underscores the need for research that rigorously examines the effectiveness of large language models (LLMs) in improving student attainment.

One study that examined the potential impact of integrating ChatGPT into the learning of tertiary students involved examining the quality of writing (Yesmin, 2023). The study was conducted in two phases within a single day: a traditional written exam and an assignment completed using ChatGPT. Both assessments consisted of 15 tasks related to basic information science, and a faculty member evaluated the results. The similarity index of ChatGPT-generated assignments was checked using Turnitin to assess academic integrity. The findings revealed significant student performance discrepancies between the traditional written exam and the ChatGPT-assisted assignment. While the highest score in the written exam was 25.5 out of 30, the lowest was 5. In contrast, the ChatGPT-generated assignments showed a much narrower range, with the highest score being 26 and the lowest 22. However, no statistical significance tests were reported, which may be a cause for concern. Moreover, the similarity index for ChatGPT-generated assignments was alarmingly high, reaching up to 89%, which is unacceptable under university regulations. None of the ChatGPT-generated assignments included in-text citations or references, raising concerns about academic ethics. The study concluded that while ChatGPT can enhance learning, its use must be regulated to maintain academic integrity and avoid plagiarism.

Whilst it seems clear that further research is urgently required to ascertain the potential efficacy of LLMs in education (Guan et al., 2023; Lo, 2023), ChatGPT-generated help is reported to produce learning gains equivalent to human tutor-authored help on mathematics skills (Pardos & Bhandari, 2024). This study involved 274 randomly assigned participants to one of three conditions: control (who received no help), human tutor-authored help, or ChatGPT-generated help. The study used a pre-test and post-test design to measure learning gains. Statistically significant results were obtained. Participants in all conditions were even at pre-test, but the group that received ChatGPT support produced statistically significant learning gains of 17% compared to the no-help control (1.85%) with $p \ll 0.001$. Human tutor-authored hints produced a learning gain of 11.62%, also statistically significant at $p=0.001$. No statistically significant differences in learning gains were observed between the ChatGPT and human tutor conditions ($p=0.416$). The findings of this research paper would indicate that ChatGPT-generated help produced learning gains equivalent to human tutor-authored help in

mathematics skills development.

Applying AI-assisted learning to enhance open-ended mathematical problem-solving

Our proposed study will examine using a large language model (LLM) such as ChatGPT to improve students' ability to solve open-ended mathematical problems. Our methodology involves a structured approach to evaluate the impact of ChatGPT on students' mathematical problem-solving skills. Initially, students will be presented with an open-ended mathematical problem, and their responses will serve as a pretest for comparative purposes. Following this, students will receive instruction on how to use ChatGPT to interact with the problem effectively. This training will include demonstrations on how to ask ChatGPT to rephrase questions, define terms, and clarify meanings. Additionally, students will learn how to use follow-up questions to explore the effects of different observations and assumptions. They will then be encouraged to use ChatGPT as a literacy coach to help them provide an improved response to a second mathematical problem, which will serve as the first post-test (post-test 1). Finally, students will be asked to solve a third mathematical problem without the assistance of ChatGPT (post-test 2). This second post-test will be used to assess any potential improvement in the students' problem-solving skills, indicating the lasting impact of the initial ChatGPT intervention.

Significance and expected outcomes

Development of the intervention and piloting of the study has commenced, with early findings tentatively emerging. However, caution needs to be exercised, given the preliminary nature of this work. Rather than label these findings as preliminary, it is prudent for us to discuss expected outcomes informed by evidence from the pilot. The expected outcomes from our research are anticipated to demonstrate the impact of integrating ChatGPT into the problem-solving process on students' mathematical understanding and problem-solving skills. Informed by findings from our pilot study, we foresee that initially, students will exhibit wide variability in the quality of their responses to an open-ended mathematical problem. By teaching students how to use ChatGPT to pose the initial problem and subsequently ask further questions to refine their answers, we expect them to gain deeper insights into the problem, enabling them to explore different observations and assumptions without the burden of manual recalculations. This intervention is likely to result in improved quality of responses. However, a narrower range of quality is also expected, indicating that the initial post-test could measure the quality of ChatGPT's assistance rather than the students' understanding. However, we predict that a follow-up post-test without the use of ChatGPT will show a residual effect, with the quality of students' responses improving from the initial pretest. This would suggest that the skills and insights gained using ChatGPT have a lasting positive impact on their problem-solving abilities.

Conclusion

Large Language Models (LLMs) promise a multitude of applications in the field of education (Guan et al., 2023; Lo, 2023). They may serve as intelligent tutoring systems, providing personalised feedback and tailored learning paths based on individual student needs and progress. LLMs may also assist in automating administrative tasks such as grading, scheduling, and even generating lesson plans, freeing educators to focus more on direct student interaction. However, as Ansari et al. (2023) explain, most studies in this field are non-empirical, with only 27% providing empirical research evidence. This underlines the urgent need for further empirical studies in diverse educational contexts. Preliminary findings from the development of our study show that explicitly teaching students how to use LLMs, such as ChatGPT, can develop critical thinking skills and enable students to improve their mathematical literacy and problem-solving skills. Nonetheless, this is a rapidly evolving field of study that urgently requires empirical research evidence, and the ongoing development of our approach is designed to contribute to this need by providing such evidence.

References

Ansari, A. N., Ahmad, S., & Bhutta, S. M. (2023). Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review. *Education and Information Technologies*, 1-41. https://doi.org/https://doi.org/10.1007/s10639-023-12223-4

Cartwright, K. B., Taboada Barber, A., & Archer, C. J. (2022). What's the Difference? Contributions of Lexical Ambiguity, Reading Comprehension, and Executive Functions to Math Word Problem Solving in Linguistically Diverse 3rd to 5th Graders. *Scientific studies of reading*, *26*(6), 565-584. https://doi.org/10.1080/10888438.2022.2084399

Chen, C., & Leitch, A. (2024). LLMs as Academic Reading Companions: Extending HCI Through Synthetic Personae. *arXiv preprint arXiv:2403.19506*. https://doi.org/https://doi.org/10.48550/arXiv.2403.19506

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: global reactions to AI innovations. *Scientific Reports*, *13*(1), 15310. https://doi.org/10.1038/s41598-023-42227-6

Gates, P. (2018). The importance of diagrams, graphics and other visual representations in STEM teaching. *STEM education in the junior secondary: The state of play*, 169-196. https://doi.org/https://doi.org/10.1007/978-981-10-5448-8_9

Giannos, P., & Delardas, O. (2023). Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. *JMIR Medical Education*, *9*(1), e47737. https://doi.org/https://doi.org/10.2196/47737

Guan, C., Ding, D., Gupta, P., Hung, Y.-C., & Jiang, Z. (2023). A Systematic Review of Research on ChatGPT: The User Perspective. *Exploring Cyber Criminals and Data Privacy Measures*, 124-150. https://doi.org/https://doi.org/10.4018/978-1-6684-8422-7.ch007

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education sciences*, *13*(4), 410. https://doi.org/https://doi.org/10.3390/educsci13040410

Nair, A., & Dubé, A. K. (2021). Overcoming Barriers to Remote Learning: Computer Assisted Instruction to Enhance Mathematical Word Problem Solving.

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PloS one*, *19*(5), e0304013. https://doi.org/https://doi.org/10.1371/journal.pone.0304013

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *Zdm*, *52*(1), 33-44. https://doi.org/10.1007/s11858-019-01118-9

Prabowo, A., Suparman, S., Li, C. S., Janan, D., & Damayanti, T. D. (2023). The effect of reading literacy to mathematics comprehension of elementary school students in Indonesia and Malaysia. *Int J Eval & Res Educ ISSN*, *2252*(8822), 8822. https://doi.org/https://doi.org/10.11591/ijere.v12i1.25714

Spennemann, D. H., Biles, J., Brown, L., Ireland, M. F., Longmore, L., Singh, C. L., Wallis, A., & Ward, C. (2024). ChatGPT giving advice on how to cheat in university assignments: how workable are its suggestions? *Interactive Technology and Smart Education*(ahead-of-print). https://doi.org/https://doi.org/10.1108/ITSE-10-2023-0195

Swanson, H. L., Kong, J. E., Moran, A. S., & Orosco, M. J. (2019). Paraphrasing Interventions and Problem–Solving Accuracy: Do Generative Procedures Help English Language Learners with Math Difficulties? *Learning Disabilities Research & Practice*, *34*(2), 68-84. https://doi.org/https://doi.org/10.1111/ldrp.12196

Vanderburg, R., Harris, L., Dargusch, J., & Richardson, S. (2023). Low socioeconomic status students transitioning from Vocational Education and Training (VET) to university: Examining definitions of success. *Higher Education Research & Development*, *42*(3), 742-756. https://doi.org/10.1080/07294360.2022.2089098

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286.

Whimbey, A., Lochhead, J., & Narode, R. (2013). *Problem solving and comprehension: 7. ed* (7th ed.). Routledge. https://doi.org/10.4324/9780203130810

Yesmin, S. (2023). Redesigning tertiary educational evaluation with AI: A task-based analysis of LIS students' assessment on written tests and utilizing ChatGPT at NSTU. *Science & Technology Libraries*, 1-13. https://doi.org/https://doi.org/10.1080/0194262X.2023.2269230

Zheng, Y. (2023). ChatGPT for Teaching and Learning: An Experience from Data Science Education. *arXiv.org*. https://doi.org/10.48550/arxiv.2307.16650

Zong, M., & Krishnamachari, B. (2023). Solving math word problems concerning systems of equations with GPT models. *Machine Learning with Applications*, *14*, 100506. https://doi.org/10.1016/j.mlwa.2023.100506

ASCILITE 2024

Navigating the Terrain:

Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies

Trotter, P., Vanderburg, M., & Vanderburg, R. (2024). AI-Assisted Pedagogies: Enhancing Mathematical Literacy and Open-Ended Problem-Solving with ChatGPT. In Cochrane, T., Narayan, V., Bone, E., Deneen, C., Saligari, M., Tregloan, K., Vanderburg, R. (Eds.), *Navigating the Terrain: Emerging frontiers in learning spaces, pedagogies, and technologies*. Proceedings ASCILITE 2024. Melbourne (pp. 635-640). https//doi.org/10.14742/apubs.2024.1443