MERGA 47: Teacher Reads

How Knowledgeable is ChatGPT 4o? Assessing the Pedagogical Content Knowledge of a Generative Artificial Intelligence Tool

Scott Cameron
Faculty of Education
University of Melbourne
scameron@unimelb.edu.au

Carmel Mesiti
Faculty of Education
University of Melbourne
cmesiti@unimelb.edu.au

Paul Fijn
Faculty of Science
University of Melbourne
paul.fijn@unimelb.edu.au

The emergence of ChatGPT has transformed professional learning opportunities for teachers, offering innovative ways to develop their professional expertise through engaging in personalised dialogues with AI chatbots, exploring questions and topics directly related to their practice. However, the use of GenAI tools presents challenges, particularly regarding the trustworthiness and reliability of AI-generated outputs. These tools are known to include "hallucinations" or inaccurate information (e.g., Ming & Mansor, 2023), potentially exposing teachers to misleading or incorrect ideas. Given these considerations, it is crucial to critically examine the potential of GenAI tools to support the professional learning of teachers.

This paper reports a pilot study investigating the Pedagogical Content Knowledge (PCK) and Content Knowledge (CK) capabilities of ChatGPT 4o. Data included ChatGPT's responses to four items, each administered five times to identify response variation. Items required ChatGPT to generate explanations that would support student learning of directed number, geometry, decimals and area. Responses were coded as *fully correct, partially correct* (i.e., including an error that many teachers could identify and correct) or *incorrect* (i.e., an insufficient response or including errors that many teachers would not recognise). Overall, we analysed 115 responses and found that ChatGPT provided fully correct responses approximately 60% of the time, with an additional 10% being partially correct, and 25% incorrect. We also observed that the correctness of responses appears to be task-dependent, as indicated by the variation in the percentage of correct responses across the four items.

We were impressed by ChatGPT's ability to provide a wide range of unique responses to the items, noting the limitation of items that required ChatGPT to produce a diagram. ChatGPT's diverse responses (averaging 5 responses per item) demonstrate its potential as a valuable resource, offering a variety of approaches to enhance teaching strategies and support student learning. However, only about 60% of responses were fully correct, indicating that ChatGPT may not be sufficiently accurate to reliably develop mathematics teachers' PCK, especially for novices who may struggle to identify incorrect and partially correct responses. As these tools evolve and new tools are released to the market, monitoring of their PCK and CK is crucial. We posit that as the reliability of a GenAI tool on a PCK and CK test improves, they could become powerful resources for tailored professional learning, potentially leading to positive outcomes for teaching and learning.

References

Ming, G. K., & Mansor, M. (2023). Exploring the impact of ChatGPT on teacher professional development: Opportunities, challenges and implications. *Asian Journal of Research in Education and Social Sciences*, *5*(4).

For more information, please refer to the following paper presented at the 47th Annual Conference of MERGA in July 2025.
Cameron, S., Mesiti, C., & Fijn, P. (2025). How knowledgeable is ChatGPT 4o? Assessing the pedagogical content knowledge of a generative artificial intelligence tool. In S. M. Patahuddin, L. Gaunt, D. Harris & K. Tripet (Eds.), *Unlocking minds in mathematics education. Proceedings of the 47th annual conference of the Mathematics Education Research Group of Australasia* (pp. 93-100). Canberra: MERGA.