

2 - 4 July 2026

Zurich , Switzerland

A Preliminary Evaluation of Artificial Intelligence-Facilitated Marking Processes Across Seven Dimensions of Student Hand-Written Work

Dr. Fai-hang Lo

The Chinese University of Hong Kong

Abstract

This preliminary study attempted to evaluate the performance of some common artificial intelligence service platform, such as Copilot, as a teaching assistant specialized in marking student coursework in seven dimensions: reading handwriting, assessing hand written short/long answers, rubric-based grading, marking according to structured marking schemes, checking hand written calculations, interpreting hand drafted graphs, and recognizing scientific patterns in hand drawn diagrams. Copilot demonstrated clear strengths, including superb speed and handwriting recognition capability, ease of use, and accessibility to educators. It performed reliably when evaluating clearly written responses, though human verification remained necessary for very poor handwriting. In rubric-based judgement, Copilot tended to be more generous than other platforms, particularly when compared with Claude that apparently demonstrated professionalism; whereas in marking scheme grading that required the identification of specific terms or concepts, it demonstrated comparable performance with other platform, in general. Copilot interpreted mathematical calculations effectively when supplied with the correct solutions beforehand (instead of working out the solutions itself), proving a good checking tool. However, its limitations became evident when assessing hand drafted graphs or scientific patterns, such as experimental results producing precipitation lines; although it could recognize labels/words and basic structural elements, it explicitly acknowledged its constraints as large language model; which was less specialized for analyzing images, such as graphs and patterns. A further weakness was its limited working memory that restricted it from complex tasks, when compared with, for example, ChatGPT. The study is ongoing; which shall continue to evaluate Grok, Claude, Gemini, ChatGPT, and other emerging platforms.

Keywords: Artificial Intelligence; Automatic Assessment; Large Language Model; Pedagogical Technology; Student Learning Analytics