Investigating Teacher Evaluative Judgement in an AI-mediated Assessment Landscape: A Phenomenographic Study

Generative AI (GenAI) tools such as ChatGPT are increasingly used by university students worldwide to help with their assessment tasks. This presents unprecedented challenges for teachers to make judgements on various difficult issues – for example, how should teachers judge students' work that has been partially or fully enhanced by GenAI? Do students who complete their work without GenAI assistance deserve a higher grade? These questions point to an urgent need to support teachers in their evaluative judgement – i.e., the process through which teachers assess the quality of students' work – in an AI-mediated assessment landscape. Developing such support necessitates an in-depth understanding of the varying practices and concerns teachers face in their evaluative judgement, which has not yet been captured in empirical research. Therefore, this project proposes three main objectives: (1) To explore how teachers make evaluative judgement of students' work that may to varying degrees be mediated by GenAI; (2) To identify the key factors that shape different teacher evaluative judgement; (3) To generate preliminary insights into the impact of teacher evaluative judgement on students' interaction with GenAI. Building on these objectives, the project aims to develop three major deliverables: (a) a typological model for categorising different patterns of teacher evaluative judgement, (b) a compilation of critical incidents capturing challenging judgement situations encountered by teachers in an AI-mediated context, and (c) research-informed guidelines to help teachers navigate their evaluative judgement effectively. This project is framed as a phenomenographic study. Based on the disciplinary taxonomy (soft/hard, pure/applied), the project focuses on four disciplines – arts, education, sciences and engineering – with the intention of capturing potential variations in teacher evaluative judgement. Data will be collected through individual interviews with 32 teachers from these disciplines as well as four case studies. The case studies will feature four courses in these disciplines, employing think-aloud protocols with follow-up teacher interviews, document analysis, and student interviews. The project will recruit participants from the Greater Bay Area but the AI-related issues being explored have global significance and relevance. The sweeping changes brought by GenAI to higher education make this project timely and urgent. As the first research of its kind on teacher evaluative judgment in an AI-mediated context, the project will provide valuable insights into teacher professional development, assessment designs and AI policies to enable more responsible and effective use of GenAI for teaching and learning in higher education.