

Shape the Future Leaders Coalition 2025-26 Case Study

School: Warwick Schools Foundation, Warwick

Project Leaders: Jenny Parkinson-Mills and Sherrie Lavender

Research Strand: AI & Digital Pedagogy

Research Question: How can generative AI be used by teachers to develop learners' metacognition during the evaluation stage of the metacognitive process, in Years 9 and 10?

Timeline: Autumn Term 2025 – Summer Term 2026

Participants: Year 9 and 10 students in Geography, English, and Philosophy & Theology

Key Data Collection:

- Baseline and post-intervention MS Forms surveys on students' self-assessed evaluative skills
- Student interviews (stratified by self-rating at baseline; degree of improvement at close)
- Short follow-up forms after each fortnightly SchoolAI session
- Chatbot interaction data downloaded from SchoolAI
- Teacher questionnaires and follow-up interviews
- Classroom observations during intervention and post-withdrawal phases
- Follow-up survey approximately one month after the intervention ends

The Challenge

Warwick Schools Foundation has arrived at an important inflection point. An internal survey of 137 staff found that only around a third have used AI in the classroom — yet even among those who haven't, a significant majority said they wanted to learn more. The barrier is not one of resistance but confidence and knowledge. Against this backdrop, Jenny Parkinson-Mills and Sherrie Lavender identified their pedagogical question: can generative AI scaffold the most cognitively demanding moment in learning — the point at which students must evaluate their own work and decide how to improve? Their research targets the evaluation stage of the metacognitive process, focusing on humanities subjects — Geography, English, and Philosophy & Theology — where the ability to reflect critically on the application of disciplinary and substantive knowledge is central yet often difficult to develop. The project works with Year 9 and Year 10 classes, using SchoolAI chatbots as structured thinking partners to support students through this evaluative moment before gradually stepping back.

The Action Research Approach

The project follows a three-phase design. In the initial data collection phase, students complete a survey assessing their own evaluative skills across three dimensions — how much they reflect on the knowledge used in a task, the skills applied, and the overall outcome — with follow-up interviews with a selection of students. During the intervention phase, teachers first meet to refine subject-specific prompts for their SchoolAI chatbots, then deploy the tool fortnightly following a task, with students completing a short reflection form after each session and teachers downloading chatbot interaction data for analysis. Whole-school and departmental training is deployed to ensure consistent implementation across different subject contexts and teacher confidence levels, with prompt refinements documented throughout. The intervention then ends deliberately: a final round of surveys and interviews mirrors the baseline, and a follow-up survey approximately one month later tests whether any metacognitive gains have persisted without AI support. Classroom observations both during and after the intervention allow the

team to assess not just what students report but what independent self-evaluation actually looks like in practice once the scaffold has been removed.

The Broader Significance

The deeper ambition of this project is to establish AI as a tool that strengthens rather than substitutes for thinking. The purpose of implementing a scaffold and then removing it is purposeful: it's one thing to show that students evaluate better when AI is present; the research asks whether students continued that metacognitive evaluation once AI is gone. If they do, the project offers a replicable model for using generative AI not as an answer machine but as a temporary cognitive scaffold, for developing abilities that students will need to deploy on their own. The focus on humanities subjects is also significant, since these are disciplines where evaluative thinking is incredibly important but also hard to teach and measure. The structured evaluation design — with baseline and post-intervention surveys, fortnightly data collection, interviews, and a delayed follow-up — reflects a level of methodological rigour and ambition that goes beyond typical classroom AI pilots. For the Coalition, the Warwick project hopes to contribute direct evidence on a question that is extremely relevant to current policy debates: not whether AI improves performance in the moment, but whether it can develop the durable metacognitive capacities that support learning across a lifetime.