UNIVERSITY OF PORTSMOUTH
CENTRE FOR ACADEMIC AND DIGITAL INNOVATION

Can Copilot be your Wingman for Marking and Feedback?

Sally Cheng, Dr Martin Hoskins, Gavin Knight, Claire Perry, Dr Mary Watkins

Jisc pilot overview

- Sector-wide concern about workload
- Growing interest in AI tools for teaching and assessment
- Need for evidence, not hype
- Collaborative learning opportunity with other HEIs and FEs involved in pilot

UoP's involvement in pilot

Could AI add value to marking and feedback?

Approach

- Small scale pilot
  - three modules
  - S&H, CCI and HSS

- Controlled Copilot environment
- Reflective practice
- Students able to 'opt out'
  - 15 out of 40 students

What we did

Why Copilot Agents?

- Integrates with our digital ecosystem
- Closed box environment and control over data
- Staff already using Copilot

How it works

- Knowledge documents on assessment information
- Detailed instruction prompt that directs agent
- Programmed to analyse submission, give detailed feedback and grade

What we did

Challenges

- New technology within our institution
- Unable to perform some skills needed with specific progamming
- Issues sharing Agent within our ecosystem
- MS issue with uploads
- Little guidance available

What worked

- Straightforward to build
- Performs the task programmed (albeit rigidly)
- Can be limited to knowledge base documents
- Performs task quickly, most outputs in under 5 mins
- Flexibilty for feedback to be as indepth as required

What we learned

MARTIN, CLAIRE, GAVIN

- Assessment types
- What worked
- What didn't work
- Criteria alignment
- Reflection
  - AI as starting point not end point – copilot, not autopilot
  - Design and prompting matter
  - Clear marking criteria

Claire: what worked and what didn't

- Our marks did not match at all, often a grade band apart – closest 58/55 (AI was meaner)
- It failed one paper at 38, I gave it 65. Failed to recognise a quote in italics.
- It was very absolutist in penalising spelling and grammar.
- We largely agreed on feedback strengths/weaknesses. AI wrote this more quickly and more formally.
- Areas to develop were largely unsuitable and sometimes went against what was required.

What we learned

Martin – L4 Childhood and Youth Studies

Assessment Type: Essay (2,000 words)- Focus on historical change, causation, and consequence across different educational settings

What Worked

- Useful comparison between human and AI-generated marking.
- AI appeared to apply marking criteria consistently and objectively.
- Feedback closely matched the published marking criteria.
- Highlighted potential as a standardisation tool for essay content/criteria mapping

What Didn't Work

- AI performance was inconsistent and often unreliable.
- Grade disparities in over 60% of assessments
- Variations ranged from one grade boundary higher/lower to two boundaries lower.
- Written feedback not always clear, appropriate, or in line with assessors' own expectations
- Reduced confidence in AI-generated grades.

Reflections

- AI can support marking but should not replace human judgement.
- Particularly important in Humanities and Social Sciences, where interpretation and nuance matter.
- Prompt design significantly influences feedback quality.
- Clear criteria provide a framework, but effective feedback still requires human expertise and context.

What we learned

Gavin – L6 Pathological Scienes 3

Assessment Type: Students create an app based on a clinical guideline. The assessment includes a formative element where students receive feedback on their app plan, and the summative submission must include annotation acknowledging how that feedback was used.

- Technical issues with MS Copilot created delays and difficulties in the use of this Agent
- Clinical guidelines meant larger knowledge base – added complexity
- Students submitting flow diagrams

Next steps

Reflect and share

- Work with MLs to better understand **if AI could add value to marking and feedback**
- Refine Agent build to see if results can be improved upon
- Share experiences and reflections with Jisc and other HEIs and FEs involved in pilot