

ASCILITE 2023

People, Partnerships and Pedagogies

Frenemies - Unleashing the power of ChatGPT in assessments

Matthew Wysel

The University of New England

This paper deals with the educational necessity of collaborative learning and individual assessment. Historically, communication-based inefficiencies in student collaboration supported individual assessment but increasingly content creation technologies are both overcoming these inefficiencies and enabling students to collaborate during assessments. The rapid proliferation in the adoption of Generative AI packages such as ChatGPT is accelerating these challenges faster than universities can respond. This paper presents five, scaffolding assessment designs that are currently being piloted within graduate and undergraduate university classes. Crucially each strategy is predicated on the readily achievable - if scary - notion that authentic assessment and Generative AI services can be *frenemies*. Strategies covering extension across a range of quantitative and qualitative classes are detailed. Preliminary findings and 'next steps' are presented.

Keywords: Generative AI, Collaboration, Authentic Assessment, Assessment Design.

Introduction

Effective learning requires both assessment and collaboration (Gregory et al., 2014; Senge, 1997) as students move through shared learning pathways in the acquisition and appropriation of their learning goals. Effective assessment 'closes the loop' on the learning process that ties individuals to their scholastic performance while collaboration enables creation of new knowledge (Wurman, 2001) whether dialogic, as students work with learning resources such as their classmates or learning systems, or monologic, as students work from these resources (Jabri & Jabri, 2022).

However, the proliferation of communication technologies such as official platforms like Learning Management Systems, Slack or Microsoft Teams, or unofficial platforms like chat groups on Facebook or Discord, rapidly evolve the capabilities of collaborative learning even as they simultaneously empower and complicate the task of assessing an individual student's performance (Geerling et al., 2023). On the one hand, authentic assessment idealises 'real-world' scenarios and their associated technologies, but on the other hand, these real-world scenarios are increasingly intertwined with the availability and utilization of typically free collaborative content creation technologies (Lawrie, 2023). In the middle, students' must balance their responsibilities to official academic integrity policies and opportunities from technology-enabled collaboration.

Further complicating this tension, the rapid adoption of Generative Artificial Intelligence (AI) – or *Gen AI* – systems such as Open AI's *ChatGPT*, Google's *Bard*, Inflection's *Pi*, or Luka's *Replika* enable students to collaborate – albeit with non-humans – at an almost immersive level (Tlili et al., 2023). Significantly, Gen AI services use statistical, natural language models to interpret and create prose, code, imagery, or mathematical outputs that mimic those created by humans (OpenAI, 2022). This makes traditional methods of detecting plagiarism increasingly ineffective (Khalil & Er, 2023). Where formally university markers struggled to detect ghost writing and other forms of contract cheating, the rate of increase in the sophistication of Gen AI, coupled with the unprecedented increase in adoption has left tertiary, secondary and trade educators scratching their heads at the apparent dereliction of what were considered effective assignments less than 12 months ago (Geerling et al., 2023). Importantly and however uncomfortable for educators, this form of content creation still falls within the understanding of 'collaboration'. Students prompt the Gen AI service with a question, which the service interprets before responding in much the same way two students might interact in a study hall. Just like a study partner, the services' responses might be correct, incorrect, or – most typically – a combination of the two. As in any case of collaboration, the student must assess which output they are presented with and how to incorporate that output within their own learning (Tlili et al., 2023).

This concise paper proposes five scaffolding assessment designs that unite collaborative-based learning and individual-centric assessment in a Gen AI empowered world. This approach supports the tenet that increasingly dichotomies between online/offline, natural/artificial, and even human/non-human are becoming less relevant than the educational opportunities afforded by deliberately blending people with technology. This paper notes

that while ‘individual assessment’ and content generating AI might – at first glance – be enemies, with an appropriate assessment design methodology, the two might be united as friends. Crucially, the strategies proposed by this paper move beyond the ‘stop-gap’ attempts that attempt to exploit rapidly vanishing weaknesses in the capabilities of Gen AI services and attempt to reflect the likely endpoint of knowledge workers who operate with AI in a symbiotic relationship of responsible decision making and ever increasingly capable – and personal – AI (Guo & Gil, 2023). In this way, it is hoped the proposed assessment designs contributes to the expanding corpus of AI-infused digital pedagogies that evolve alongside students and educators, encouraging collaboration between both human and non-human agents while retaining the individual accountability that assessment requires.

Background

The context for the development and piloting of these assessment strategies was a series of four classes at the University of New England in Australia. The university teaches approximately 20,000 students through graduate and undergraduate classes that are taught both on campus and online. The four subjects included a qualitative management class (80 undergraduate students), two mixed quantitative/qualitative computer science classes (54 undergraduate students and 79 graduate students), and a project-based business class (10 undergraduate students). The two Computer Science classes were ‘co-badged’, teaching both undergraduate and postgraduate students with an overlapping curriculum, ostensibly similar assessment tasks, and in a common learning environment, while the Management and Business classes were stand-alone classes with independent learning environments. Students’ collaboration occurred within these classes using the official learning management system built around the *Moodle* platform and via several unofficial ‘student-only’ platforms including Discord servers and Facebook groups.

Historically, one of the key factors that has hampered a centralised response to rapid technological change has been the protracted timelines that necessarily accompany significant changes to gazetted assessments. To that end the initial requirement for each proposed assessment design was that any adaptation remain minor enough to be considered akin to ongoing development or evolution of the assessment and not significant enough to trigger a gazette change. For example, where policy noted an assessment as an ‘essay-style response of approximately 2000 words’ the proposed assessment design was adapted to stay within that style and length of response. Accordingly, the scaffolding assessment designs presented in this paper are deliberately couched in terms that emphasize their similarity to existing assessments to aid their incorporation within the normal cycle of assessment evolution that accompanies annual reviews.

Finally in line with other studies (Tlili et al., 2023), students revealed a broad range of proficiency levels pertaining to both general familiarity of Gen AI and its incorporation as a study tool. To avoid overburdening students with an additional technology-based learning curve, workshops and *Ask-Me-Anything* sessions were offered in the first week of the trimester to gauge the cohort’s proficiency and provide basic training. Unsurprisingly, computer science students showed an increased familiarity with the use of Gen AI systems to interrogate assessment questions when compared to their counterparts in the business and management classes.

As summarised in Table 1, the assessments in the three classes included a range of essay-style responses, written reports, oral presentations, technical analysis, and presentation and assessment of visual media – such as a digital poster.

Assessment Design – Something for Everyone

Each of the five assessment designs piloted in the four classes will be briefly summarised in order of increasing pedagogical sophistication. Table 1 summarises the pedagogical philosophy, teaching goal and assessment style each design was piloted within.

Raise the Tide

Raise the Tide normalises the use of Gen AI services across a cohort. Historically, significant diversity in ethical frameworks across a class supported a learning environment that rewarded academic dishonesty as less honest students leveraged unsanctioned resources, such as ghost writers or assignment swapping websites, while more honest students tend to eschew them. Vocal, enthusiastic normalising of readily accessible study resources such as ChatGPT mitigates much of the reward for dishonest behaviour as all students can now safely participate at the level they wish to. In the Management class, the sanctioning and promotion was supplemented with recorded *Ask-Me-Anything (about ChatGPT)* sessions run in Week 1 for all classes. These sessions included demonstrations of ‘copy-and-pasting’ the assignment question into a Gen AI service and an initial discussion

about how to improve responses received. This permitted discussion around questions such as, “won’t I get kicked out if I do that?” and “how do I know the answer is true?”. These sessions concluded with the lecturer leading the class through a brief comparative exercise on tech-enabled collaboration versus ‘non-tech’ collaboration such as that traditionally found in study halls.

Table 1. Assessment Design, Philosophy, Targets, and Use Within the Study

Name	Philosophic Approach	Teaching Target	Class (Assessment / Style)
Raise the Tide	Normalise the technology, contextualize within the curriculum, encourage adoption	Promote equity throughout an ethnically diverse cohort; reduce <i>busy work</i> for both teachers and students	MGMT300 (A1 / Essay) MGMT300 (A2 / Essay)
Old School Collaboration	Require collaboration with Gen AI but also critical analysis of output	As above plus promote critical thinking within the context of collaboration with Gen AI.	COSC300 (A1 / Essay) COSC500 (A1 / Essay)
Personal Responsibility for Collaborative Effort	Encourage maximum utilisation of Gen AI but focus is on personal accountability for the final output	Extending above plus centralise the issue of personal accountability regarding a collaborative exercise.	COSC300 (A2 / Essay & Presentation) COSC500 (A2 / Essay & Presentation) BUS300 (A1 / Extended Abstract, Presentation, & Q/A session) BUS300 (A2 / Report, Presentation & Q/A session)
Invert the Problem	Ability to utilise Gen AI becomes a key part of the assessment	Extending above, demonstrate proficiency managing a task that includes Gen AI as a component in a broader process	COSC500 (A3 / Report presentation of quantitative analysis)
Immersive Assessment:	Collaboration becomes open-ended as students must use Gen AI to <i>learn</i> class theory.	As above, but students must determine when they can trust Gen AI in an open-ended task such as their own learning.	COSC500 (A4 / Report and graphical presentation of quantitative analysis)

Old School Collaboration

Old School Collaboration included the explicit mandate to work with a Gen AI service to answer a predefined question. Students were required to submit a templated response that included: 1) their prompt(s) to the Gen AI service, 2) the reply(ies) received, and 3) their critique of the replies received according to established class theory. While this design was written for qualitative assessments it could easily be adapted for quantitative assessments.

To simplify marking, a template including each section was provided. In stark opposition to the current suite of defensive assessment techniques used to combat Gen AI usage that include hyper-specific questions, niche questions, contexts or writing styles, the assessment questions in this design were deliberately made generic for the specific purpose of promoting quality in the output. ‘Vanilla’ questions permit the Gen AI system to provide more informative responses and greater degrees of insight. Therefore, students attempting to short-cut the process found the service provided more robust answers that required a greater understanding of class theory to critically analyse. As per normal, students were marked on their ability to demonstrate an understanding of established class theory across both the prompts and critiques.

Personal Responsibility for Collaborative Effort

Personal Responsibility for Collaborative Effort can be used with either qualitative or quantitative assessments and is designed to extend students’ previous experience with earlier collaborative assessments. This scaffolding experience permitted students to extend their AI-collaboration skills but focused on individual accountability for the outcome of the collaboration.

In the Computer Science class, the original assessment style mandated an essay structure. To meet this requirement students were required to collaborate with ChatGPT to generate an essay-style script which they read over the top of a *Pecha Kucha* presentation. Students were required to submit both the script and the pre-recorded presentation for marking to limit the significant increase in marking time that often accompanies presentation-style assessments (Akimov & Malin, 2020). Where collaboration with Gen AI services was an explicit component in the previous assessment designs, the focus of this assessment design shifted to personal accountability for any collaboration that might have happened. This shifts the central issue for students from ‘how can I get away with it?’ to ‘how can I get the most value from it?’ (Khalil & Er, 2023). Having been trained to use the services, students were now implicitly required to determine the extent of their reliance on collaboration, ostensibly without oversight, but with the increase in focus on individual accountability that accompanies a personal presentation.

All variants of the assessment required students to maintain a personal ownership of the output of their collaborative efforts and ensure they were able to articulate that output in their own words. Emerging findings from the Business class show students investing significant effort in integrating the output from Gen AI services into their own theoretical understandings, often producing a deeper understanding compared to previous years.

Invert the Problem

Invert the Problem inverts the normal problem of Gen AI too-easily solving human-generated assignments by requiring students to collaborate with a Gen AI service to first create, and then solve, their own problem. In this case, students were required to “use a Gen AI package and [the prescribed analytics package] RapidMiner to prepare a Weekly Sales Report for a fictitious retailer that operates a chain of 5 Pet Supply stores.”

To complete this assessment students had to first use a Gen AI service to generate a sample dataset, that upon analysis would yield insights of the type a medium sized retailer would include in a Weekly Sales Report. These requirements mapped directly to established class theory while the requisite skills were taught in a workshop. Students had to feed the results of their analysis back into their chosen Gen AI service to reinforce the importance of designing quality into each stage of collaboration. Collaborating with their chosen Gen AI service a second time, students had to produce the prose and graphics contained in their personal Weekly Sales Report. Students submitted the sample dataset (to ensure it was unique) and the resulting Weekly Sales Report. Students were graded on their ability to create an insightful analytical report from unique data.

As before, this assessment design inverted the normal danger of *specificity* in generic assessment questions. In this design, generic questions produce better quality answers which require greater understanding of established class theory to analyse and/or extend.

Immersive Assessment

Finally, *Immersive Assessment* immersed students into a multifaceted assessment where they were free to use Gen AI services in whatever manner they saw fit. This specific assessment was a combination of qualitative and quantitative analysis that required students to learn class theory before designing a bounded question that they would be able to answer using a graphical analytics package taught during class workshops.

Crucially, this assessment design changed the rationale students might use when deciding whether to collaborate with Gen AI services. The purpose of collaboration goes from ‘show me the answer’ to ‘how might I go about this assignment’. In so doing, Gen AI services transition from a sometime-conspirator to a potential mentor as students become immersed in a collaborative effort to use Gen AI to empower – rather than disempower – their learning journey. This assessment was scheduled to run at the end of the semester, so students were expected to use their growing understanding of *how* Gen AI services operated to determine what level of time and trust to invest in securing in any advice. The critical analysis students had to demonstrate when collaborating with Gen AI at the start of the semester no longer needed to be formally assessed as students had become immersed in a learning process where Gen AI services were simply one-of-many potential collaboration partners that they must differentiate between.

Preliminary Findings

As this study will continue to run until November 2023 the current findings are only preliminary in nature. Nonetheless, indicative outcomes can be divided into two main categories.

Outcomes for Students and Teachers

Students report significantly reduced busy-work as they spend less time attempting to understand the specific requirements of assessments. They now freely copy-and-paste assessment questions into Gen AI services with the expectation that those services will provide a perspective – but not necessarily an answer – on the assignment. Additionally, when prompted these services also provide sample answers to questions which the students must first critically answer with their emerging understanding of class theory and secondly develop as they see fit.

The immediate benefit for teachers is fewer ‘gimme’ questions from students who attempt to solicit more and more hints from teachers. Now teachers can simply respond with their own question, “What did ChatGPT tell you, and what did you think about that?” Additionally, this utilisation of Gen AI serves as a journeyman’s tutor and especially helps weaker students who historically tended to consume the most time. Lastly, those students who also avail themselves of Gen AI tutelage are also beginning to show increases in the quality of their work.

Outcomes for Assessment Design

In all classes the assessment design incentivised students to begin preparation of the assignments in collaboration with a Gen AI service and to continue critiquing its output throughout the process. Normalising the use of Gen AI services meant just as in ‘real-world’ collaboration, once students no longer found collaboration useful, they reverted to progressing their assignments individually. The utilisation of Gen AI ceased being a threat to assessment design and became an indicator that students possessed the ability to manage their own learning pathway. In other words, *when* students ceased using collaboration services stopped being relevant, rather *that* students chose to both commence and cease collaborating became an important factor in their learning pathway.

Concluding Remarks

The contribution of this paper lies in its exploration of potential strategies that reinstate the twin requirements of collaboration and assessment in students’ learning pathways. This paper offers five scaffolding assessment designs that unite students’ technologically empowered ability to collaborate with both human and non-human partners while retaining individual assessment. In so doing, assessments remain authentic even as the capabilities of Gen AI services continue to expand, and academic integrity becomes a natural extension of assessments design rather than an externally imposed requirement. The pilot study continues.

References

- Akimov, A., & Malin, M. (2020). When Old Becomes New: A Case Study of Oral Examination as an Online Assessment Tool. *Assessment & Evaluation in Higher Education*, 45(8), 1205-1221.
- Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT Has Aced the Test of Understanding in College Economics: Now What? *The American Economist*, 05694345231169654.
- Gregory, B., Uys, P., & Gregory, S. (2014). The Role of Instant Feedback in Improving Student Understanding of Basic Accounting Concepts. *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings of ascilite Dunedin 2014*.
- Guo, S., & Gil, E. (2023). *No Priors Ep. 16 with Mustafa Suleyman, Founder of Deepmind and Inflection*. YouTube. <https://youtu.be/g4VszCFonPk>
- Jabri, M., & Jabri, E. (2022). *Managing Organizational Change*. Bloomsbury Publishing.
- Khalil, M., & Er, E. (2023). Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection. *arXiv preprint arXiv:2302.04335*.
- Lawrie, G. (2023). Establishing a Delicate Balance in the Relationship between Artificial Intelligence and Authentic Assessment in Student Learning. *Chemistry Education Research and Practice*, 24(2), 392-393.
- OpenAI. (2022). *Dall.E 2 Promotional Video* [Online]. <https://openai.com/product/dall-e-2>
- Senge, P. M. (1997). The Fifth Discipline. *Measuring Business Excellence*, 1(3), 46-51.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, 10(1), 15.
- Wurman, R. S. (2001). *Information Anxiety 2*.

<p>Wysel, M. (2023). Frenemies - Unleashing the Power of ChatGPT in Assessments. In T. Cochrane, V. Narayan, C. Brown, K. MacCallum, E. Bone, C. Deneen, R. Vanderburg, & B. Hurren (Eds.), <i>People, partnerships and pedagogies</i>. Proceedings ASCILITE 2023. Christchurch (pp. 614–618). DOI: https://doi.org/10.14742/apubs.2023.653</p>

Note: All published papers are refereed, having undergone a double-blind peer-review process. The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.