“Understood the assignment”: A UX-led investigation into student experiences of GenAI

HEMSWORTH, Kirsty, WALKER, Alex and EVANS, Jayne

Citation:

HEMSWORTH, Kirsty, WALKER, Alex and EVANS, Jayne (2024). “Understood the assignment”: A UX-led investigation into student experiences of GenAI. In: BECKINGHAM, Sue, LAWRENCE, Jenny, POWELL, Stephen and HARTLEY, Peter, (eds.) Using Generative AI Effectively in Higher Education. Sustainable and Ethical Practices for Learning, Teaching and Assessment. Routledge, 54-64. [Book Section]

“Understood the assignment” – A UX-led investigation into student experiences of GenAI

Dr Kirsty Hemsworth, Alex Walker and Jayne Evans

ORCID ID Kirsty Hemsworth: 0009-0001-1460-3430
ORCID ID Alex Walker: 0009-0000-8203-9678
ORCID ID Jayne Evans: 0009-0000-1538-8122

Abstract

This chapter explores how undergraduate students with limited experience of academic writing intuitively engage with generative AI. Using an adapted user experience model for usability testing (Blakiston, 2014), the study investigates how beginner users navigate and interact with LLMs in real-time, focusing on prompt generation and iteration. By analysing screen recordings and audio narrations, the study considers how the process of translating assessment tasks into instructive prompts offers valuable insights into how students approach academic writing tasks with the support of LLMs. Preliminary findings suggest that AI-literacy should be integrated into academic skills-based curricula, building student competence and confidence on the ethical and effective use of generative AI for assessment purposes. This is best achieved through partnership work with current students.

Introduction

The arrival of GenAI in Higher Education institutions (HEIs) and university classrooms will necessitate unprecedented change for academic staff and students alike. From pedagogical taxonomies and assessment practices to referencing and academic integrity, many touchstones of higher education are being interrogated in light of AI capabilities. Many academics and researchers have been early adopters of GenAI, testing the limits of large language models (LLMs) to generate plausible and publishable academic material (Altmäe, Sola-Leyva and Salumets, 2023; Hosseini, Rasmussen and Resnik, 2023; Macdonald *et al.,* 2023), yet we are largely unaware of how students might be testing the limits of these platforms. While there are genuine concerns about GenAI outputs impacting authorship and intellectual property, more intriguing – at least for this chapter’s authors – are how student-AI interactions shape writing processes and academic literacies. We explore how students intuitively engage with GenAI chatbots and consider how capturing this process can offer valuable insight into how undergraduate students conceptualise academic writing and AI-literacy.

Background

Plagiarism and academic misconduct dominated early discussions of student interactions with LLMs (Barnett, 2023; Dien, 2023). Universities were swift to publish guidance on how HEIs could uphold "academic rigour and integrity" (Russell Group, 2023). Concurrently, library and information professionals grappled with issues of acknowledgement and copyright and researchers questioned whether LLMs could be credited in publications as contributing authors (McAdoo, 2023). Another complex issue is student behaviour surrounding GenAI – their motivations for using it, their understanding of its capabilities, and their perceptions of its ethical applications. Knee-jerk responses to LLMs have assumed that students engage with these tools to "cut corners" (Graham, 2023) – a view that risks limiting discussions of student GenAI-use as remedial behaviour, similar to the framing of contract cheating as a marker of "low academic competence" (Sweeney, 2023, p.72). Detection has been proposed as a miracle solution to these perceived threats to academic integrity, but persistent student behaviours regarding contract cheating, along with the recent discontinuation of much-publicised GenAI detection software (Nelson, 2023), suggest that productive interventions regarding AI-use are best placed earlier in the writing process.

Approaching LLMs as tools that can help students navigate higher education offers a more positive view of GenAI and reframes the effective implementation of GenAI platforms as an alternative academic literacy and employability skill (Ewen, 2023). Practical pedagogical frameworks for GenAI use have already been developed in the field of information literacy, where database searching strategies have clear similarities to the process of prompt generation (Lo, 2023). Just as effective literature searching involves the selection of explicit search terms - and the systematic introduction of alternate keywords and inclusion/exclusion criteria - prompt engineering for LLMs benefits from iterative refinement and the

> *"continuous process of testing prompts, analysing the generated AI responses, and making necessary adjustments based on observed patterns and outcomes" (Lo, 2023).*

Previously framed as an information literacy skill, students clearly need to understand the process of prompt generation and refinement to obtain relevant and reliable information from LLMs.

Prompt generation: A new academic skill?

Equally important if students are to create explicit prompts that link to assessment criteria, is the use of instructive language over open questions. Decoding the language of assessment criteria is a core academic skill taught by learning developers, and students are encouraged to interrogate the task set by their markers and adapt their criticality as appropriate. Here, we begin to see the crossover between academic skills, information literacy and AI competencies. Effective prompt generation belongs to a wider set of critical thinking and evaluation skills that enable students to extract information from assessment criteria, to identify useful research materials, and to integrate these into their academic work.

Experimentation with prompt generation for academic writing, largely focused on formulaic writing structures within scientific disciplines, has been successful (e.g. using GenAI tools to generate largely descriptive sections of scientific articles, such as abstracts, introductions and conclusions (Altmäe, Sola-Leyva and Salumets, 2023; Macdonald *et al., 2023*)). Particular success has been found in terms of formality, word usage and sentence structure (Májovský et. al., 2023). When testing LLMs' ability to generate scientific case studies, Buholayka, Zouabi, and Tadinada (2023) found that, while the omission of technical information for a professional audience was a key barrier to generating publication-compliant writing, the structure of the LLM-authored case study was largely accurate. While these studies illustrate the potential for LLMs to create well-structured pieces of academic writing, we note that the prompt iteration and refinement was undertaken in these cases by the authors. Therefore, we must interrogate the identities and experience levels of the 'academic writers' described in much existing research on GenAI (Giray, 2023): do/can undergraduate students possess the ability and contextual knowledge to translate assignment tasks into meaningful AI-appropriate prompts?

Approaching student behaviour as user experience

Screencasts – where students record their interactions with digital platforms and capture all on-screen actions – is an established user experience (UX) methodology in library settings, providing insight into users' information retrieval behaviours (Conrad and Stevens, 2019), use of online library databases (Jones and Wisniewski, 2021) and website navigation (Denton, Moody and Bennett, 2016). They offer a window on how students interact with online platforms and, if recorded in one take, can offer a largely authentic view of student behaviour. Combining screencasts with think-aloud protocols (Mathison, 2015) is particularly useful, offering additional insight into student decision-making in real-time, particularly where participants are working remotely without supervision.

Several studies have applied think-aloud protocols to explore student writing, introducing an additional, reflective layer of qualitative data where written content alone is insufficient to capture the writing process. Research has considered how think-aloud protocols of assignment feedback might foster open dialogue between academic staff and students (Cunningham, 2019; Jones, Georghiades and Gunson, 2012), while narrated recordings of written translation exercises offer insight into how EAL students navigate the intersections of source and target languages when self-translating (Seror, 2013). In all cases, adding think-aloud protocols aims to reduce researcher inference, offering insight into how students engage with digital technologies and writing processes, as expressed in their own words.

Method

Using an adapted user experience model for usability testing (Blakiston, 2014), two Level 4 students were asked to record their screens and narrate their use of ChatGPT-3.5 in response to a simulated assignment brief and marking criteria. Neither of the participants had any prior experience using ChatGPT-3.5: this platform was chosen as it was free to access for the participants and the functionality is largely representative of other LLMs. Both participants had heard of the tool and had expressed interest in exploring the software. This provided added value for participants - using the project as a 'dry run' to test the LLM capabilities. The participants were also beginners in essay-based writing. As Finance, Accounting and Business students, they had experienced a range of assessment formats including report-writing, presentations, and short-answer form exams. Neither had written an extended essay-based assignment and therefore approached the task with limited knowledge of the assignment writing process.

Assessment Task

Module title	Finance in Context
Assessment topic	Financial Literacy Education in Universities
Assessment type	Essay
Word count	2000 +/- 10%
Weighting	50%
Submission date	Friday 3 November 2023

Assessment brief

You are required to write and submit a 2000-word essay exploring the topic of financial literacy education in universities, examining the importance and impact of financial literacy training for students experiencing the challenges of the cost-of-living crisis and the effects of the Covid-19 pandemic.

LO1: To identify the importance of financial literacy in Higher Education.

Students will critically assess the significance of financial literacy education in higher education, examining its impact on students' financial well-being and financial success. They will identify key reasons why financial literacy education is an important for universities to provide as part of their student support offers.

LO2: To examine current financial literacy initiatives in UK universities, identifying strengths and challenges.

Students will investigate and compare existing financial literacy initiatives in universities. They will evaluate different approaches, programs, or resources aimed at enhancing students' money management skills, and assess their effectiveness and common challenges.

LO3: Propose practical recommendations or guidelines for universities to enhance their financial literacy support.

Students will synthesise their understanding of the importance of financial literacy and insights from existing initiatives to suggest practical recommendations for universities to enhance financial literacy education.

Figure 1: The assessment brief used by participants

A simulated assignment brief was devised to allay any concerns regarding the use of GenAI chatbots with live assessments. The module title and assessment topic were invented for the purposes of the screen capture activity, following a standardised layout for assignment briefs. The activity was unscripted – participants were asked to begin their screen recording from the input of their first prompt and use the microphone to record an audio explanation of their decision-making while navigating the tool and creating an essay plan. The screen recordings were then analysed and compared by a postgraduate student, focusing on how the participants extracted information from the brief to generate prompts, and how these were refined based on the text outputs in ChatGPT-3.5. This was enhanced by a written reflection submitted by both participants once the analysis was completed to address any gaps in the recordings where researcher inference risked skewing the results.

Screencast analysis

From their opening prompts, we noted that, despite their lack of experience using LLMs, both participants approached the task with preconceptions about the tool's capabilities. Both opened with questions, seeking information on how to approach the *task* of writing an extended essay, rather than generating any content. Participant A seemed to conceptualise ChatGPT-3.5 as a search engine rather than a generative tool. They judged the quality of the content based on its replicability:

> *"I found that each time I asked a question, it showed slightly different answers. So I was hoping for some consistency…unfortunately [there] wasn't a lot. There was a question I asked regarding good recommendations in financial literacy in universities. [The answer] was not what I was hoping for, so I tried a different way of searching."*

The term 'searching' here is particularly revealing - Participant A follows a literature-searching process, demonstrating some awareness of the importance of quality academic sources in their prompt construction. Their subsequent prompts repeated this evaluative language: *"what are good recommendations for "financial literacy education in universities" essays?"* The use of speech marks in these prompts also suggests prior knowledge of phrase searching – a skill commonly used in library database searching – which again alludes to their conceptualisation of the LLM as an information discovery tool.

Later, Participant A narrowed their search by directly quoting keywords from the assessment brief to enhance their prompts. However, the results of this search were deemed unsatisfactory – the prompt was misinterpreted and the LLM generated suggested essay subheadings and content. The student reacted over the microphone, redirecting the LLM to act as search engine, rather than text generator:

> *"this was not what I was looking for – I was looking for some book recommendations or articles…I'm going to do that again"*

However, through this process, Participant A demonstrated basic 'iterative refinement' skills (Lo, 2023) – they followed up with a clear statement to redirect the LLM: *"book recommendation for financial literacy education in universities."*

However, by overlooking ChatGPT-3.5's generative capabilities and viewing the text outputs as information retrieval results, Participant A failed to identify when the tool produced fabricated content. Several book recommendations were hallucinations (Li et al., 2023; Májovský et al., 2023). Participant A missed this, despite spelling mistakes in some of the generated content, and continued to be preoccupied with the specificity and refinement of their inputs using literature-searching strategies – they introduced "higher education" and

"UK" as additional search terms – rather than considering that the quality of the outputs may be compromised.

Participant B demonstrated a stronger understanding of ChatGPT-3.5's generative capabilities, and described engaging with the tool in conversation:

> *"I was hoping to start a natural conversation like you would with a peer if you didn't know something" (using text outputs to supplement existing subject knowledge)*
> *'I used ChatGPT-3.5 as a guide like talking to a fellow peer on tips or advice.'*

Starting from an opening question regarding essay structure, Participant B adopted a broad-to-narrow approach, narrating how they intended to redirect their prompts towards generating useable content: *"what subheadings should i use to write an academic essay on financial literacy".* Here, they alluded to their knowledge of subheadings as a useful tool for organising academic content but did not refer to the assessment brief, which provided a three-part structure for students to follow. Instead, we see that Student B's essay plan (Figure 2) closely mirrors the presentation of information in the chat window, suggesting that the organisation of AI-generated responses has a direct impact on how students structure their academic work:

1. **Introduction:**
   - Definition and importance of financial literacy
     - Management of money and personal finance
     - Making informed financial decisions
     - Interpret financial information
   - Brief overview of the essay's structure

2. **The Basics of Financial Literacy**
   - Understanding financial literacy
   - Key concepts: income, expenses, budgeting

3. Case Studies and Success Stories

- Examples of individuals benefiting from financial literacy
  - Credit counselling in Singapore case study (2004)
  - Chilean pension reform
  - Mexican financial education for micro-entrepreneurs

Figure 2: Transcription of Student B's essay plan, showing subheading and bullet-point structure copied directly from GenAI.

Participant B's prompt refinement followed a top-down approach: from each response, they selected a keyword and requested more detail. Where the relationship between the generated subheadings was unclear, Participant B sought clarification - *"how does financial literacy relate to income, expenses and budgeting?"* - in an effort to filter out irrelevant information from the long list of content generated by a previous prompt. The result was a detailed essay plan, laid out in a clear bullet-pointed structure that stretched beyond the scope of the initial prompt. However, at no point in the recording did the student return to the brief to check the relevance of the AI-generated content, or whether the volume of content they had amassed would fit the assessment word count. Here, we see an inquisitive but ultimately distracted exploration of topics such as Chilean pension reform and credit counselling in Singapore, facilitated by the LLM.

Both participants recognised that their ability to effectively use ChatGPT-3.5 could be improved with further practice, suggesting AI-use is approached as a skill to be developed. Participant A felt that the key to improving their use of the tool was in refining their questions to produce accurate and replicable information, while Participant B expressed an interest in testing out more complex tasks, such as '*giving ChatGPT-3.5 some background information and then asking it a question about the information.'* These reflections on future use align with the approach of each student: Participant A views the LLM as a search engine and a means of discovering accurate information, while Participant B has a stronger awareness of its generative abilities and applications. After using ChatGPT-3.5 for the first time, both participants reflected on its value and relevance for academic writing, yet their preconceptions about *how* to use the tool remained fixed.

Recommendations for practice

These recorded interactions offer only the slightest of glimpses into how students might engage with GenAI platforms to inform, craft and refine their academic writing. Yet the implications are clear: academic writing, and the set of analytical and evaluative skills that this encompasses, is a process that cannot remain untouched by the transformative effects of GenAI. GenAI has inverted the taxonomy upon which so many of our pedagogical approaches are based: the creation of new knowledge is no longer the preserve of the experts, but a skill opened to all learners with a single prompt (Rivers and Holland, 2023). Yet as our screen recordings revealed, beginner users can quickly move into self-directed conversations with LLMs, with each stage of the exchange determined by the perceived relevance of the generated content, rather than a pre-planned set of prompts or structured set of assessment criteria.

As a result of this study, we propose four key recommendations for developing AI-literacy which we are now integrating into a university-wide academic skills programme, supported by a team of student partners:

1. **Start with prompts**
   Prompt generation is a powerful tool that combines information literacy skills and the instructive language of written assessments. We will include demonstration of how to craft specific and analytical prompts in our core academic writing sessions, using examples of assignment briefs and genuine AI responses to scaffold teaching.

2. **Foster GenAI literacy as a skillset**
   Approaching AI literacy as a competence-based skillset opens up the possibility for all learners to approach GenAI-use as a developmental process. We will develop new content on GenAI applications that connect academic and employability skills, using relevant frameworks to encourage students to track their digital competencies.

3. **Encourage critical evaluation**
   Students require skills and confidence to assess the relevance of LLM outputs, and to decide about their application within their specific academic disciplines. We will update the examples in our academic skills sessions to engage students in up-to-date conversations around ethical applications of GenAI.

4. **Embrace user experience methodologies**
   Understanding how students interact with GenAI tools is crucial to develop responsive approaches to teaching AI literacy. We will expand the UX methodology used in this study to build a database of student interactions with GenAI to inform teaching, learning and assessment approaches.

Although brief, our recordings offer a glimpse of GenAI-student interactions. They suggest that user experience methodologies have the potential to capture specific moments in which GenAI tools influence and inform student approaches to academic writing. Perhaps most exciting is the possibility that this approach can offer insight into how less visible students – those hesitant to seek support or engage in 'remedial' development activities – might capitalise on the anonymity of interactions with GenAI tools to experiment with, and develop, their academic writing skills.

Critical Appraisal

Zayne Knight, MSc Cancer Biology

On my course, we have been informed of the unreliable nature of some GenAI tools and issues of plagiarism, but not discouraged from using them. GenAI has the potential to be an excellent tool and is already useful for mathematics and identifying data trends in my field. I feel that the same methodology which this chapter has used would be a reliable way to test the capabilities of GenAI tools for programming and more advanced data analysis.

I have used GenAI tools to offer more information on selected topics, particularly for definitions and explanations. An example of a prompt I would use is: "What are nanoparticles and explain how they are able to treat cancer" and I would ask a follow-up question such as, "Could you provide academic references to support this information"? As I have experienced, the GenAI tool will provide references, however, many of these references do not work or have been redacted from papers.

When reflecting on the study chapter, I realised that the prompts used by beginners are quite general and could be broken down into more specific searches.

References

Altmäe, S., Sola-Leyva, A., and Salumets, A. (2023) 'Artificial intelligence in scientific writing: a friend or a foe?', *Reproductive BioMedicine Online*, 47(1), pp.3-9. Available at: https://doi.org/10.1016/j.rbmo.2023.04.009.

Barnett, S. (2023) 'ChatGPT-3.5 Is Making Universities Rethink Plagiarism', *Wired*, 30 January. Available at: https://www.wired.com/story/ChatGPT-3.5-college-university-plagiarism/ (Accessed 20 September 2023).

Blakiston, L. (2014) *Usability Testing: A Practical Guide for Librarians*. Maryland: Rowman & Littlefield.

Buholayka, M., Zouabi,R. and Tadinada, A. (2023) 'The Readiness of ChatGPT-3.5 to Write Scientific Case Reports Independently: A Comparative Evaluation Between Human and Artificial Intelligence', *Cureus*, 15(5). Available at: https://doi.org/10.7759/cureus.39386

Conrad, S. and Stevens, C. (2019) '"Am I on the library website?": A LibGuides Usability Study', Information Technology and Libraries, 38(3), pp.49-81. Available at: https://doi.org/10.6017/ital.v38i3.10977

Cunningham, K. (2019) 'Student Perceptions and Use of Technology-Mediated Text and Screencast Feedback in ESL Writing', *Computers and Composition*, 52, pp.222-241. Available at: https://doi.org/10.1016/j.compcom.2019.02.003

Denton, A., Moody, D. and Bennett, J. (2016) 'Usability Testing as a Method to Refine a Health Sciences Library Website', *Medical Reference Services Quarterly*, 35(1), pp.1-15. Available at: https://doi.org/10.1080/02763869.2016.1117280.

Dien, J. (2023) 'Editorial: Generative artificial intelligence as a plagiarism problem', *Biological Psychology*, 181. Available at: https://doi.org/10.1016/j.biopsycho.2023.108621.

Ewen, M. (2023) 'Mapping the potential of AI in the age of competence based higher education', *WonkHE*, 20 February. Available at: https://wonkhe.com/blogs/mapping-the-potential-of-ai-in-the-age-of-competence-based-higher-education (Accessed: 10 September 2023).

Giray, L. (2023). ‘Prompt Engineering with ChatGPT-3.5: A Guide for Academic Writers’, *Annals of Biomedical Engineering*. Available at: https://doi.org/10.1007/s10439-023-03272-4

Graham, A. (2023) ‘ChatGPT-3.5 and other AI tools put students at risk of plagiarism allegations, MDU warns’, *BMJ*, 381, p.1133. Available at: https://doi.org /10.1136/bmj.p1133

Jones, E. and Wisniewski, C. (2021). ‘Video killed the multiple-choice quiz: Capturing students’ literature searching skills using a screencast video assignment’, *Journal of the Medical Library Association,* 109(4). Available at: https://doi.org/10.5195/jmla.2021.1270.

Li, Y *et al.* (2023). ‘Can large language models write reflectively?’, *Computers and Education: Artificial Intelligence,* 4. Available at: https://doi.org/10.1016/j.caeai.2023.100140.

Lo, L. S. (2023). ‘The Art and Science of Prompt Engineering: A New Literacy in the Information Age’, *Internet Reference Services Quarterly*, pp. 1–8. Available at: https://doi.org/10.1080/10875301.2023.2227621

Lo, L. S. (2023). ‘The CLEAR path: A framework for enhancing information literacy through prompt engineering’. *The Journal of Academic Librarianship*, 49(4). Available at: https://doi.org/10.1016/j.acalib.2023.102720

McAdoo, T. (2023) ‘How to cite ChatGPT-3.5’, *American Psychological Association*, 7 April. Available at: https://apastyle.apa.org/blog/how-to-cite-ChatGPT-3.5 (Accessed 20 September 2023).

Jones, N., Georghiades, P. and Gunson, J. (2012) ‘Student feedback via screen capture digital video: Stimulating student’s modified action’, *Higher Education*, 64, pp. 593-607. Available at: https://doi.org/10.1007/s10734-012-9514-7

Macdonald, C. *et al.* (2023) ‘Can ChatGPT-3.5 draft a research article? An example of population-level vaccine effectiveness analysis’, *Journal of Global Health*, 13. Available at: https://doi.org/10.7189/JOGH.13.01003

Májovský, M. *et al.* (2023) Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora’s Box Has Been Opened’, *Journal of Medical Internet Research*, 25. Available at: https://doi.org/10.2196/46924

Mathison, S. (ed.). (2005) *Encyclopedia of Evaluation.* New York: SAGE Publications.

Nelson, J. (2023) ‘OpenAI Quietly Shuts Down Its AI Detection Tool’, *Decrypt*, 24 July. Available at: https://decrypt.co/149826/openai-quietly-shutters-its-ai-detection-tool (Accessed 11 September 2023).

Rivers, C. and Holland, A. (2023) ‘How can generative AI intersect with Bloom’s taxonomy?’, *Times Higher Education*, 30 August. Available at:

https://www.timeshighereducation.com/campus/how-can-generative-ai-intersect-blooms-taxonomy (Accessed 19 September 2023).

Russell Group (2023) 'New principles on use of AI in education', *Russel Group*, 4 July. Available at: https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/ (Accessed 8 September 2023).

Seror, J. (2013) 'Screen capture technology: A digital window into students' writing processes / Technologie de capture d'écran: une fenêtre numérique sur le processus d'écriture des étudiants', *Canadian Journal of Learning and Technology / La revue canadienne de l'apprentissage et de la technologie,* 39(3). Available at: https://doi.org/10.21432/T28G6K

Sweeney, S. (2023). 'Who wrote this? Essay mills and assessment – Considerations regarding contract cheating and AI in higher education.' *The International Journal of Management Education*, 21(2). Available at: https://doi.org/10.1016/j.ijme.2023.100818

Youmans, R.J. (2011) 'Does the adoption of plagiarism-detection software in higher education reduce plagiarism?', *Studies in Higher Education*, 36(7), pp.749-761. Available at: https://doi.org/10.1080/03075079.2010.523457