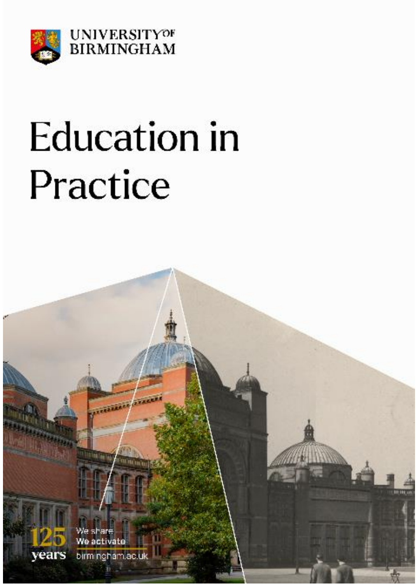

Lawler, CRE (2026) 'Direct Generative AI interaction improves undergraduate critical thinking', *Education in Practice*.

First published online on 20 February 2026

ISSN 2057-2069

https://education-in-practice.co.uk/lawler-cre-2026-direct-generative-ai-interaction-improves-undergraduate-critical-thinking/

Direct Generative AI interaction improves undergraduate critical thinking

Dr Catherine R E Lawler, Department of Biomedical Science, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham; Bath Spa University

Contact: c.lawler@bham.ac.uk

First published online on 20 February 2026

Abstract

Evaluation, a cognitive skill which underlies critical thinking, is a key graduate attribute. Simultaneously, teaching around Generative Artificial Intelligence (GenAI) has been incorporated into Higher Education. Barriers to learning are present in both areas, so it was hypothesised that the teaching of both may be enhanced using Experiential Learning. This study presents an intervention undertaken with multiple groups of undergraduate students with the intended learning outcomes of increasing critical thinking and knowledge of GenAI. During the intervention, students evaluated writing samples in three groups: those who interacted directly with AI, those who interacted indirectly with AI, and those who did not knowingly interact with GenAI. Students increased evaluative behaviour by 170% during the intervention, demonstrating acquisition of critical thinking skills. This effect was 29% stronger in students who interacted directly with GenAI, highlighting the importance of Experiential Learning in skills-based teaching. During the intervention, all students discussed prompt use and increased their understanding of GenAI. This intervention is predicted to be transferrable to many contexts. The results challenge current research that GenAI use by undergraduate students may be detrimental for the development of critical thinking, and highlight the need for further research into student-AI interaction.

Keywords: critical thinking, graduate attributes, experiential learning, higher education, skill retention, generative AI

Introduction

Critical thinking features as a graduate attribute at many universities, including the University of Birmingham, where evaluation underpins critical thinking (University of Birmingham, 2024). Evaluation features as a pinnacle cognitive skill in Bloom's Revised Taxonomy, and is highlighted as an appropriate learning outcome for sessions given to undergraduate students (Adams, 2015), making enhancing evaluation behaviour an appropriate short- and long-term intended learning outcome for undergraduate students. Teaching and learning of critical thinking skills, such as evaluation, can be knowledge-based or application-based, but instilling university students with critical thinking skills faces many barriers (Dwyer, 2023). Experiential Learning has been purported to aid in acquisition of higher order critical thinking skills (such as evaluation) in undergraduate students (Adams, 2015). This study sought to investigate if Experiential Learning was an effective tool facilitating critical thinking development in students whilst they gained knowledge about Generative Artificial Intelligence (GenAI).

GenAI, particularly Large Language Models (LLMs) have influenced higher education (HE), offering affordances and hindrances to instructors and students alike. Simultaneously, there is an increasing demand for AI-literate graduates (Demos, 2023). However, barriers exist to teaching GenAI to undergraduates, such as developing ethical considerations, and rapidly evolving technology (Demos 2023). Recent research showed adults have shown preference for human- over AI-generated text (Zhang and Gosaline, 2023). This study hypothesised that this preference would extend to undergraduate students. This hypothesis shaped the main aim of the study: to examine whether asking undergraduate students to evaluate AI-generated work improves student knowledge of GenAI whilst developing their critical thinking. To evaluate this, a one-hour intervention was developed with the Intended Learning Outcomes of 1) increase knowledge of GenAI and 2) enhance evaluation skills.

Due to the aforementioned challenges in teaching critical thinking and GenAI, significant instructor interaction was planned to support students in the Zone of Proximal Development (Vygotsky, 1978), encourage engagement, and support reflection as part of Experiential Learning (Quibrantar and Obidimma, 2023). It was hypothesised that enhanced critical thinking would be seen in groups who learned experientially (Adams, 2015), so this was incorporated via direct interaction between students and GenAI. To evaluate if there was a difference in outcome between Experiential and non-Experiential Learning, not all students would interact directly with GenAI. ChatGPT (OpenAI, 2025) was the LLM chosen for use in this intervention due to its free access, popularity and ability to produce text. Prompt engineering would be introduced to students to encourage effective interaction with GenAI tools (Wu *et al*., 2024), via examination of writing generated using different prompts. The intervention was designed to use repetition as a teaching tool for critical thinking (Adams, 2015), aiming to enhance retention of evaluation skills over the short- and long-term. As cognitive skills such as evaluation are considered higher order (Adams, 2015), Level 6 undergraduates were initially selected for this intervention. This intervention was subsequently repeated on second- (Level 5) and third-year (Level 6) undergraduate students. To ensure this intervention was transferrable to different subject areas, and that the outcomes of this study would be applicable across higher education contexts, reflective writing was the type of text chosen for evaluation, as reflective writing is considered a core student activity (Veine *et al*., 2020)

Methodology

Study design

This research project was undertaken during academic years 22/23 and 23/24 at Bath Spa University (which acted as pilot studies), and 24/25 at University of Birmingham (which acted as the main study). Participants were undergraduate students studying Biological or Biomedical Sciences who gave informed consent to participate in this study. The number of participants were: 22/23 n=6, 23/24 n=5, and 24/25 n=27, with Group A n=9, Group B n=9, and Group C n=9. A one-hour in-person intervention was scheduled with the intended learning outcomes of increasing knowledge of GenAI and enhancing evaluative skills. Ethical approval for the research outlined in this paper was given by the University of Birmingham and Bath Spa University.

GenAI use and Experiential Learning

Group C interacted directly with GenAI, Group B interacted indirectly with AI and Group A did not knowingly interact with AI. The manner of this interaction was through evaluation of text produced with GenAI. Students also evaluated their own reflective writing as part of the intervention. Sources of writing for evaluation were: Own Work (students' own writing), Sample 1 (generated by ChatGPT3.5 using a short prompt), Sample 2 (generated by ChatGPT3.5 using a long prompt), Sample 1o (generated by ChatGPT4o using a short prompt) and Sample 2o (generated by ChatGPT4o using a long prompt). Post-reveal indicates an evaluation stage students engaged in after being told the writing they were evaluating was produced by GenAI.

Evaluative behaviour

To measure evaluative behaviour, the metric "edits per minute (edits/min)" was developed. An edit was considered a single addition, deletion, annotation or rearrangement in a piece of writing, and number of edits made was self-reported by each student at the end evaluation stage of the intervention. The intervention was broken into five stages, during which students evaluated writing samples for five minutes. Students also qualitatively evaluated writing using a five-point scale, ranging from "very poor" to "very strong" for content, structure and clarity using a Mentimeter (Mentimeter, 2025). In the pilot studies, students evaluated Own Work, Sample 1, Sample 1 Post Reveal. In the first pilot study, students then analysed Sample 2, whereas in the second pilot study, students then interacted with ChatGPT (OpenAI, 2025). Both cohorts then finished the intervention to returning to evaluate their own work. In the main study, all students analysed Own work and Sample 1o. Group A analysed Sample 1o, Sample 2o and Own Work; Group B analysed Sample 1o Post Reveal, Sample 2o and Own Work; and Group C analysed Sample 1o Post Reveal, followed by a direct interaction with ChatGPT 4.o (OpenAI, 2025), then evaluation of Own Work.

Student Experience

To evaluate their experience of the intervention, students were asked to score statements on a five-point Likert scale (Likert, 1932), ranging from strongly disagree to strongly agree, with statements of "This session was X", with X as "interesting", "important", "fun" and "useful".

Data analysis

To compare between pilot groups, edits/minute was adjusted to relative edits/minute. Data are considered as Starting Evaluation Behaviour (SEB), Maximum Evaluation Behaviour (MEB) (the highest evaluation behaviour observed during the intervention) and Final Evaluation Behaviour (FEB).

Statistical analysis

Excel was used to calculate means and standard deviations and perform statistical analysis including paired or unpaired two-tailed T-tests, with Bonferroni's multiple comparison test. GraphPad Prism was used to perform normalcy testing (D'Agostino and Pearson), one-way ANOVA with Bonferroni's multiple comparison test, and repeated measures ANOVA with Sidak's multiple comparison test. For all statistical tests, $\alpha=0.05$ was used, with $P<0.05$ considered statistically significant.

Results and Discussion

Impact of intervention on critical thinking

The first pilot study appeared to show students increased their evaluation behaviour following the intervention (Figure 1a), but this was not statistically significant, possibly due to high variation between students and a low number of participants. The second pilot study showed students increased their evaluation behaviour by 120% (Figure 1b). Students in both pilot studies were more critical of writing which was not their own (Figure 1). Students in the second pilot study were more critical of the same writing when they knew it was AI-generated ($P<0.05$), consistent with the hypothesis that students prefer text they thought to be human- versus AI- generated. This is consistent with research on adult populations (Zhang and Gosaline, 2023).

At the end of the intervention, students in the first pilot showed significantly less evaluation when returning to reviewing their own writing (FEB), compared with the maximum they had shown when they evaluated writing produced by others (MEB) ($P<0.05$). However, students in the second pilot did not. This difference was hypothesised to be due to the nature of the interaction with AI. These interactions were either indirect, as experienced by students in the first pilot study, or direct, as experienced by students in the second pilot study. This supported the hypothesis that Experiential Learning via direct GenAI interaction would be more effective in developing critical thinking in undergraduates than indirect interaction with AI. This was further investigated through revised design in the main study.

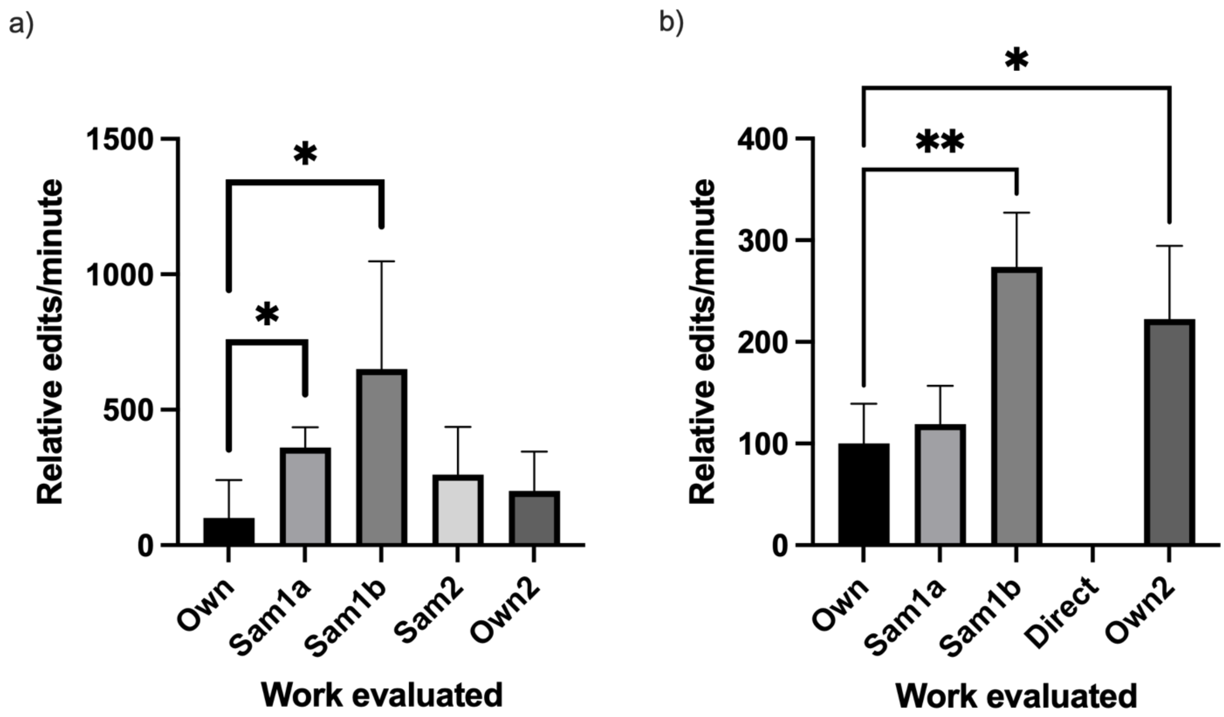

Figure 1: Student evaluative behaviour of writing from different sources with indirect and direct interactions with GenAI.

Students sequentially evaluated writing produced by themselves or GenAI. Data presented are means of relative edits per minute, ± SD. A) Students interacted indirectly with GenAI, n=6. B) Students interacted directly with GenAI n=5. A one-way repeated measures ANOVA and Sidak's multiple comparison test was used with $\alpha=0.05$, * = $P<0.05$, ** = $P<0.005$.

Comparison between pilot study cohorts was limited by an unexpected five-fold difference in initial edits/minute. This was partially accounted for by normalising to relative edits/minute. However, it was hypothesised that differences in Starting Evaluation Behaviour (SEB) could confound results. Furthermore, the temporal gap between pilot studies (one year) allows for confounding variables, such as changing student perception of GenAI, to occur. Therefore, the main study was designed which would split a single cohort into multiple groups to allow for direct comparison. Group A was a control group, who were not informed the writing they evaluated was produced using GenAI. This allowed for evaluation of the effect of that knowledge of GenAI use could have on development of critical thinking. Group B interacted indirectly with

GenAI: they evaluated writing generated by GenAI, were told it was AI-generated, and asked to evaluate it again. Group C interacted directly with AI, by evaluating writing generated by AI, being told it was AI-generated, and then being asked to use AI to generate similar writing themselves. In contrast to the pilot studies, no significant difference in starting critical thinking was found between groups regarding evaluation of their own work (P<0.9999). This facilitated direct comparison of effect of interaction with GenAI on student evaluation behaviour.

GenAI interaction improves student critical thinking skills

Similarly to both pilot studies, students were more critical of writing they did not produce, compared to their own writing (+49%; P<0.0005). This is consistent with other research, which indicates undergraduate students are less able to be critical of their own work (Guest and Riegler, 2021). Students were also more critical of the same writing when they knew it was produced by GenAI (+93%, P<0.0005) (Figure 2). This was consistent with the hypothesis of this work, and other research on adult populations (Zhang and Gosaline, 2023). All groups increased in evaluation behaviour during this intervention (+214%; P<0.0005) and finished the intervention more critical of their own writing (+170%; P<0.0001) (Figure 2), meeting one of the intended learning outcomes of this intervention (enhanced evaluation).

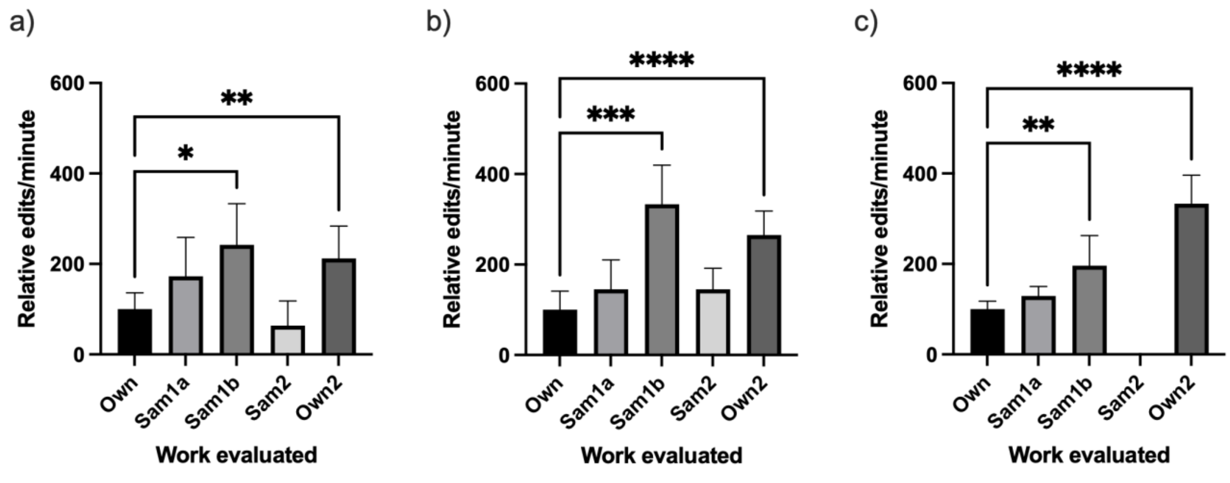

Figure 2: Student evaluative behaviour on writing from different sources, with no interaction, indirect interaction and direct interaction with GenAI.

Undergraduate students sequentially evaluated their own writing and writing produced by GenAI. A) Group A: control; B) Group B: indirect interaction with GenAI; C) Group C: direct interaction with GenAI. Data presented are mean relative edits per minute, ±

SD; n=27. A one-way repeated measures ANOVA and Tukey B post-hoc test was used with α=0.05, * = P<0.05, ** = P<0.005, *** = P<0.0005.

Experiential Learning enhances development of critical thinking in undergraduate students

Students who interacted with GenAI during this intervention had significantly increased final evaluation compared with students who did not (+41%; P<0.0001) (Figure 3a). Furthermore, students who interacted directly with GenAI showed significantly critical thinking than those who interacted indirectly (+26%; P<0.05) (Figure 3a). This supports the hypothesis that direct interaction with AI enhances evaluation behaviour. When returning to evaluate their own writing at the end of the intervention, students who interacted directly with GenAI were now most critical of their own writing, whereas students who did not interact directly with GenAI were still least critical of their own writing (P<0.0005) (Figure 3b). These outcomes support the hypothesis presented in the introduction: Experiential Learning enhances development of critical thinking.

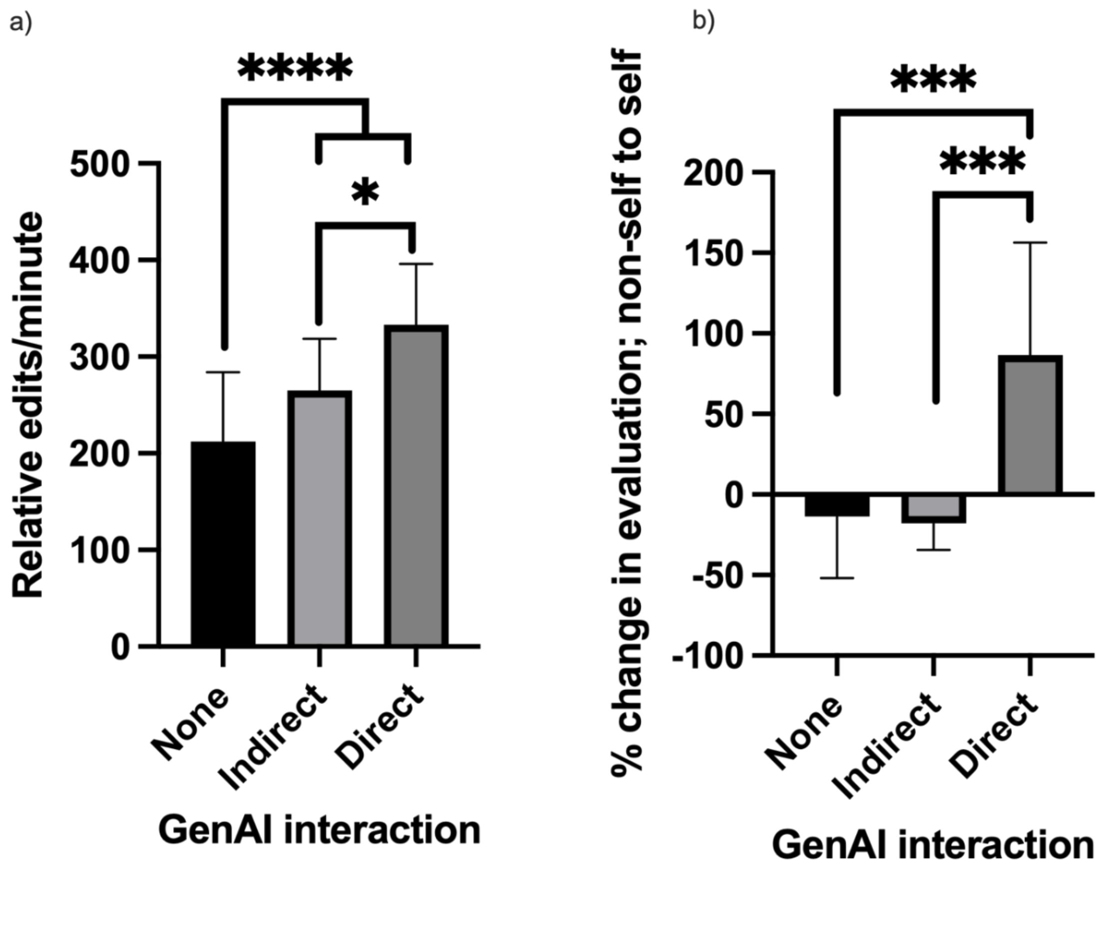

Figure 3: Effect of direct vs indirect interaction with GenAI on student evaluation.

A) Student evaluation of their own writing following different interactions with GenAI. Data presented are mean relative edits per minute ± SD; n=27. A two-tailed heteroscedastic T-test was used, with Bonferroni's multiple comparison correction. α=0.05, * = P<0.05, **** = P<0.0001. B) Change in student evaluation from writing produced by others (non-self) to their own writing (self), following different interactions with GenAI. Data are presented as % change in evaluation ± SD; n=27. A one-way ANOVA with Sidak's multiple comparison test was used with α=0.05. *** = P<0.0005.

The findings of the main study corroborated the pilot studies, supporting the hypotheses that interaction with GenAI would enhance critical thinking, and that Experiential Learning via direct interaction with GenAI further enhances evaluative behaviour in undergraduate students. The intervention was successful over the short-term as one of the intended learning outcomes of increasing student evaluative behaviour (critical thinking).

Prompt length influences student perception of AI-generated work

The second intended learning outcome of the intervention was increasing knowledge of GenAI. Students were presented with two samples of writing which were produced by GenAI: one which was generated using a short prompt and one using a long prompt. All students evaluated these pieces of writing prior to knowing they were produced by GenAI, scoring them out of five across the criteria structure, content and clarity. Writing produced using a long prompt was scored higher by students in structure (P<0.0001) and content (P<0.05) than writing produced using a short prompt (Figure 4). There was no difference in perceived clarity of work produced using a short or long prompt. Following evaluation of the writing, the prompts used to generate the writing were revealed to non-control cohorts. Students were informed they had scored writing produced using long prompt higher than writing produced using a short prompt. This served as a mechanism of stimulating discussion of prompt engineering, enhancing student engagement with, and knowledge of, GenAI tools.

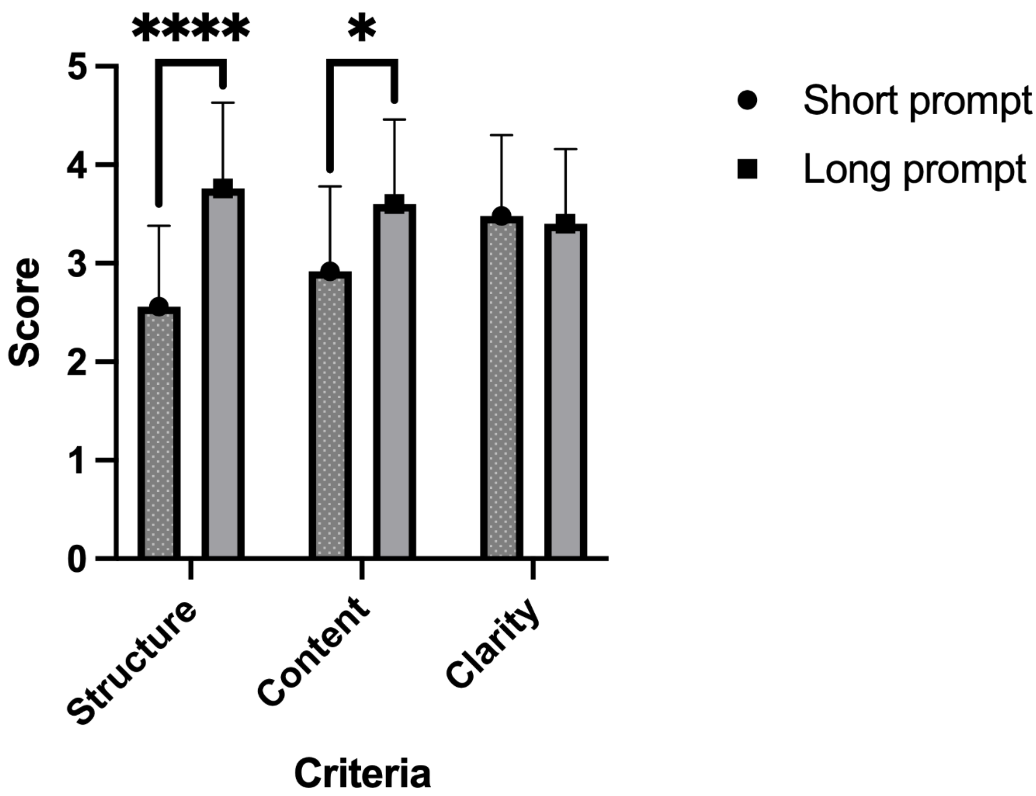

Figure 4: Effect of prompt length on student score of writing produced using GenAI.

Student evaluation of writing produced using GenAI, against criteria: structure, content and clarity. Data presented are mean score ± SD; n=27. A two-tailed heteroscedastic T-test was used, with Bonferroni's multiple comparison correction. α=0.05, * = P<0.05, **** = P<0.0001.

Experiential Learning facilitates engagement in GenAI

To evaluate student experience and acquisition of knowledge, students were asked if they found the sessions interesting, important, fun and useful. They scored responses using a Likert scale (Likert, 1932) on Mentimeter (Mentimeter, 2025). Students scored the session a mean of 3.5/5 for interesting, 4.5/5 for important, 2.7/5 for fun and 4.3/5 for useful (n=22). Students who interacted directly with AI (Group C) scored the session higher across all metrics than those who did not (P<0.005). All groups had a similar level of student: instructor interaction during the intervention, so the increase in student engagement (as measured by interest and fun) appears to be due to direct student interaction with GenAI. Students who interacted with AI before discussing prompt engineering appeared to gain more knowledge on GenAI than students who only discussed prompt engineering, scoring the intervention higher on importance and usefulness (P<0.005). These findings support the use of Experiential Learning when teaching GenAI and critical thinking to undergraduate students.

Discussion and Conclusion

This study has developed an intervention which enhances evaluation skills to develop critical thinking in undergraduate students, whilst improving their knowledge of GenAI. In doing so, it demonstrated students acquire and retain greater critical thinking when interacting directly with GenAI (compared with indirect interactions). This is consistent with Experiential Learning theory and supports the recommendation of a practical approach to teaching critical thinking and GenAI skills to undergraduates.

Application in Different Contexts

Four main considerations have been identified when applying this intervention to different contexts. The first consideration is subject specificity. Whilst this study was performed with students studying Biological and Biomedical Sciences, GenAI knowledge and evaluation skills are graduate attributes which are near-ubiquitously desired, so this intervention is likely highly transferrable to other courses. The second contextual consideration is Level of study. Regarding the hierarchical development of skills in Bloom's Taxonomy (Adams, 2015) during undergraduate courses, this intervention is recommended for students in their second or third year of undergraduate study. However, emerging drive to embed GenAI throughout curricula may encourage the development of a similar intervention for first year undergraduate students, focussing on increasing GenAI knowledge and development of lower strata of critical thinking e.g. application skills (Adams, 2015). Class size is the third considered in using this intervention: significant staff: student interaction to facilitate the Zone of Proximal Development (Vygotsky, 1978) and to support Experiential Learning (Adams, 2015) was employed in this intervention. Future investigation could examine if similar outcomes can be achieved with less instructor: student interaction and a lower staff: student ratio. The final consideration for applying this intervention is format - this session has only been administered during face-to-face teaching sessions due to diminished frequency and quality of student: instructor interaction observed during online teaching (Wut and Zu, 2021), and the potential for this to negatively impact students successfully occupying the Zone of Proximal Development (Vygotsky, 1978). If this intervention were to be administered online, technology which stimulates and facilitates instructor: student interaction should be used to maintain support for students.

Teaching Perspectives

Future interventions of this type will continue to use direct interaction between students and GenAI, in line with Experiential Learning, to support students to enhance their critical thinking and gain knowledge about GenAI usage. Discussion around prompt engineering will continue to be incorporated, to give students critical insight into how GenAI can used effectively. This discussion will be expanded to include ethical considerations around how GenAI should be used.

Future Research

Continuation of this research will seek to investigate the long-term effects of this intervention, and how this intervention could be improved to maximise retention of evaluation skills over a longer period. The results of this study contrast with some current research which hypothesises a negative effect of AI-interaction on development of critical thinking in undergraduate students (Tian and Zhang, 2025). Importantly, students in this study gained insight into prompt engineering, which is considered critical for effective interaction with GenAI in Higher Education (Wu *et al*., 2024). Therefore, when considering the results of this study alongside other works, it is hypothesised to be the type or quality of the interaction which students are having with GenAI which may be influencing their critical thinking. Therefore, more research into the nature of the GenAI: student interaction and how this effects the development of critical thinking in undergraduate students is required.

Acknowledgements

The author acknowledges the support of, and stimulating discussions with, colleagues at University of Birmingham and Bath Spa University.

References

Adams N.E. (2015) 'Bloom's taxonomy of cognitive learning objectives' *J Med Libr Assoc*, 103(3), pp. 152–153. DOI: 10.3163/1536-5050.103.3.010.

Demos (2023) 'The AI Generation: How universities can prepare students for the changing world'. Available at: https://demos.co.uk/research/the-ai-generation-how-universities-can-prepare-students-for-the-changing-world/ (Accessed: 22 Dec 2024).

Dwyer, C.P. (2023) ‘An Evaluative Review of Barriers to Critical Thinking in Educational and Real-World Settings’, *Journal of Intelligence,* 11(6), 105. DOI: https://doi.org/10.3390/jintelligence11060105.

Likert, R. (1932) ‘A technique for the measurement of attitudes’, *Archives of Psychology*, 140, pp. 1-55.

Mentimeter (2025) Available at: https://www.mentimeter.com/ (Accessed: 8 October 2025).

Kolb, D. A. (1984) *Experiential Learning: Experience as the Source of Learning and Development*, New Jersey, USA: Prentice Hall.

Quibrantar, S. M. and Obidimma, E. (2023) ‘Evaluating student engagement and experiential learning in global classrooms: A qualitative case study’, *Studies in Educational Evaluation*, 78,101290. DOI: https://doi.org/10.1016/j.stueduc.2023.101290.

OpenAI (2026) ChatGPT. Available at: https://chat.openai.com/ (Accessed: 24 June 2025).

Guest, J., and Riegler, R. (2022) ‘Knowing HE standards: how good are students at evaluating academic work?’ *Higher Education Research & Development*, 41(3), pp. 714–728. DOI: https://doi.org/10.1080/07294360.2020.1867516.

Tian, J and Zhang, R. (2025) ‘Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy’, *Acta Psychologica*, 260 105725. DOI: https://doi.org/10.1016/j.actpsy.2025.105725.

University of Birmingham (2024) *Graduate Attributes*. Available at: https://intranet.birmingham.ac.uk/student/personal-development-and-opportunities/graduate-attributes.aspx (Accessed: 17 February 2026).

Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes*, Massachusetts, USA: Harvard University Press

Veine, S., Anderson, M.K., Andersen, N.H., Espenes, T.C., Søyland, T.B., Wallin, P. and Reams, J. (2020) ‘Reflection as a core student learning activity in higher education - Insights from nearly two decades of academic development’, *International Journal for Academic Development*, 25(2), pp. 147-161.DOI: 10.1080/1360144X.2019.1659797.

Wut, T. and Xu, J. (2021) 'Person-to-person interactions in online classroom settings under the impact of COVID-19: a social presence theory perspective', *Asia Pacific Education Review*, 22(3), pp. 371-383. DOI: 10.1007/s12564-021-09673-1.

Wu, D., Wang, D., Yung, T., and Guo, K. (2024) 'Effects of a Prompt Engineering Intervention on Undergraduate Students' AI Self-Efficacy, AI Knowledge, and Prompt Engineering Ability: A Mixed Methods Study', *Computers and Society*, arXiv:2408.07302. DOI: https://doi.org/10.48550/arXiv.2408.07302.

Zhang, Y. and Gosaline, R. (2023) *Human Favoritism, Not AI Aversion: People's Perceptions (and Bias) Toward Generative AI*. Massachusetts, USA: Sloan School of Management. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453958 (Accessed: 22 Dec 2024).