THE INTERDISCIPLINARY JOURNAL OF STUDENT SUCCESS

CDS PRESS

Open Access **Protocol**

Student Judgements of AI-generated and human-generated educational content [Protocol]

Gary F. Fisher*[,1], Paula Shaw[1], Dean Fido[1,2]

Citation: Fisher, G.F., Shaw, P., Fido, D. (2024). Student Judgements of AI-generated and human-generated educational content [Protocol]. The Interdisciplinary Journal of Student Success.

Founding Editor-in-Chief: Masood Zangeneh, Ph.D.

Editor: Hamid R. Yazdi, Ph.D.

Editor: Mona Nouroozifar, Ph.D.

Received: 01/25/2024
Accepted: 03/12/2024
Published: 05/28/2024

[1]University of Derby (UK)
[2]**ORCiD:** 0000-0001-8454-3042
Corresponding author: Gary F. Fisher, g.fisher@derby.ac.uk

Abstract: Academics are aware of the emergence and rapid growth of accessible and user-friendly artificial intelligence (AI) writing software, such as the popular ChatGPT. Though focus has been placed on the students' usage of such software, this could also present a viable tool for academics. This protocol details a mixed methods approach to systematically and rigorously compare the impact that AI-generated learning content (vs. human-generated learning content) has on students, as well as to understand student perceptions and acceptance of AI-generated content more broadly. Phase one uses AI prompting to generate a piece of learning content comparable to human-generated content as validated via sentiment analysis. Phase two uses analyses of covariance to explore the impact of teaching content type (AI- vs. human-generated) and the associated label accompanying it (congruent vs. incongruent) on students' judgements of said content after controlling for the covariates of age, sex, intrinsic motivation, extrinsic motivation, and general acceptance of AI. Phase three uses semi-structured interviews across multi-disciplinary student samples to explore acceptance of AI-generated teaching content in higher education. This protocol will facilitate large-scale replication of our methods across an international and interdisciplinary landscape.

Keywords: AI, Artificial Intelligence, Educational Content, Student Judgment.

Purpose

The proposed study will establish an iterative approach to generating AI-informed higher educational learning content, and in turn will compare this content to a human-generated equivalent to explore student judgements thereof. Moreover, it will gain a qualitative understanding of said judgements across students who prefer to study online and on-campus using novel comparative thematic analysis procedures.

Rationale

Academics are becoming more aware of the emergence and rapid growth of accessible and user-friendly AI writing software, such as the popular ChatGPT. ChatGPT is a type of transformer-based language model that generates human-like texts, which has gained much interest in research and academic communities (Brown et al., 2020). Generally, this attention has been negative (e.g., Groves et al., 2022), with academics worried about students using AI technology for nefarious means, such as generating essays and contributing work to discussions that is not derived from their own knowledge pool. However, there exists a growing movement of academics who believe that we should embrace ChatGPT. For example, Fido & Harper (2023) wrote on how the use of AI might help to reduce the awarding gap on an international level, by allowing students to test out academic concepts in new environments, compare essay plans, and consolidate complex text into a reduced and novel perspective.

However, to date, there exists no research empirically exploring the judgements of students towards academics beginning to use AI in their learning and teaching strategies (or even using it to generate content). Owing to calls for improving learning content (Ho et al., 2023) and an evergreen understanding that increased workloads are contributing to poor academic staff mental health (Woolston, 2018), ChatGPT might present a useful tool to help academics logistically, whilst further increasing the quality of their teaching content. Of course, in order to proceed with such adaptations, we must understand the student perspective, especially in a world which focuses on value for money. Understanding how students perceive the use of AI, as well as potential contributing factors for their willingness to engage in such changes (such as academic motivation; Kotera et al., 2021), presents an important pedagogical issue facing academics.

Objectives

1. Use the open-access ChatGPT tool to create a set of AI-generated versions of existing, human-written, online teaching content.
2. Produce a comparative linguistic analysis of these two iterations of the same text (one AI-generated, the other human-generated) using the LIWC linguistic analysis software.

3. Survey a diverse student group to provide a quantitative dataset that will facilitate evaluation of how students perceive (e.g., judge, endorse) AI-generated versus human-generated teaching materials and whether this differs as a function of the congruency of the label this is given.
4. Conduct one-to-one interviews with University of Derby students who study either online or on-campus to enable comparative thematic analysis of the themes identified in relation to judgements of AI-generated learning content.

Duration of the Study

Enrolment and data collection for the study is estimated to take approximately three months to complete. For each participant, phase two is estimated to take no longer than 30 minutes to complete and phase three is expected to take between 45-60 minutes to complete.

Methods

Study Design

In phase one, we will employ an iterative, five-step prompting strategy, beginning with a prompt stating the required length and subject area of the work (e.g., "*Write 700-750 words explaining the engineering and motivational approaches to work design*"). We will then incrementally add further parameters and considerations for ChatGPT to incorporate into its response, reflecting the specific context and purpose for which the content was being produced. These will include ensuring the content was directed towards a particular audience (e.g., "…to undergraduate students with a fair grasp of the subject area") and the learning context in which the audience will be accessing the material (e.g., "…*who are studying an online module in Business Psychology as part of an online Bachelors in Business and Management*"). Then, we will specify stylistic and formatting characteristics present within the material (e.g., "*The text should use academic citations as appropriate and deploy real-world examples of the engineering and motivational approaches in practice*") and educational features and approaches to be employed (e.g., "*After explaining each approach, the text should include questions aimed at students that ask them to reflect on the strengths of each approach and their own experience of it*").

In phase two, approximately 400 participants will complete a questionnaire at a single time point, which tasks them with making judgements (agreement with statements) on one of two pieces of learning content that differ as a function of whether it is AI- or human-generated, and which is accompanied by either a congruent or incongruent label. Demographics (i.e., age and sex) and self-reported intrinsic motivation, extrinsic motivation, and general acceptance of AI will be obtained online using survey software Qualtrics.

For phase three, sixteen participants will partake in online semi-structured interviews (facilitated by dynamic prompts) designed to explore their perceptions of AI-generated teaching content and their acceptance of it in higher education settings.

Study Population, Selection Criteria, and Sample Size Justification

All participants will provide informed consent via Qualtrics before completion of the survey and/or the interview. Requirements for both empirical phases (phases two and three) include being over the age of 18 years, fluent in English, and being a university student based in a UK university (to control for variation in learning style).

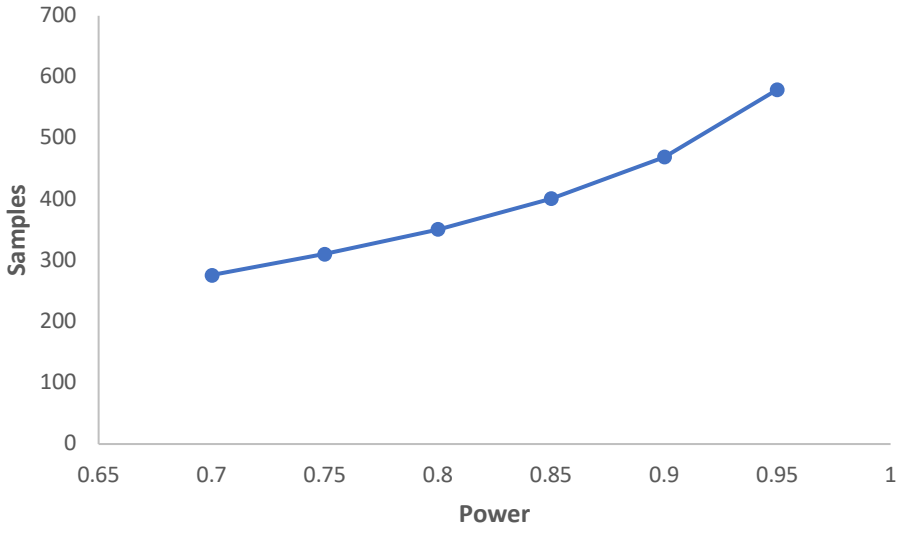
For phase two, an a priori power analysis using G*Power (version 3.1) indicates that a sample of 351 participants will be required to ensure 80% power and to ensure observed effects are of practical importance. The range of sample sizes as a function of power and a visualization thereof are featured in Table 1 and Figure 1, respectively.

For phase three, to reach data saturation and to allow us to make thematic comparisons between online and on-campus students, a total of sixteen participants will be sampled.

Table 1. *Power analysis across power levels for phase two*

	Power (1 - β)					
	0.70	0.75	0.80	0.85	0.90	0.95
Sample Size	277	311	351	401	469	580

Figure 1.
Visualizations of power analysis across power levels with small-to-medium effect size (*f* = .15)



Recruitment Method

For phase two, we will capitalize on the targeted recruitment mechanisms available through services such as *Prolific*; a survey distribution service where participants are paid a living wage for their engagement with research surveys. Through this service, only participants who meet our inclusion criteria will see and be able to take part in our research. For phase three, students will be recruited through a series of university-wide internal e-mail advertisements and programme announcements at the University of Derby. This allows for a diverse group of individuals who are undertaking different degrees and who are at different stages of said degrees. We will target both on-campus and online students to enable comparative thematic analyses.

Data Collection and Study Schedule

Data will be collected at a single time point with all data expected to be collected within a period of three months. Participants in each empirical phase (phases two and three) will be provided with study information and asked to affirm their consent prior to participation (this will be achieved through an online survey and via a button press). For phase two, participants will then be asked to provide a unique identification code (comprised of the last 3 digits of their telephone number and the last 3 letters of their name) in case of data withdrawal, and demographic questions (i.e., age and sex). Subsequent pages will contain one of the two pieces of learning content; further divided as a function of the study condition (i.e., congruent and incongruent label) followed by judgement questions and measures of academic motivation and general acceptance of AI. For phase three, a selection of interview dates/times will be provided to participants, with interviews conducted over and recorded via Microsoft Teams. To allow for anonymity, participants will not be required to have their cameras turned on during the interviews.

After both empirical phases, participants will be asked to re-affirm their consent in line with BPS guidance for internet mediated research (BPS, 2021) and will be provided with debrief information. Participants will be informed they have 14 days to withdraw their data by providing their unique ID code. Data will be maintained for a minimum of 7 years in line with GDPR guidance, but participants will be informed that in line with common practices for open and replicable science, data might be kept indefinitely in a fully anonymised form. All identifiable data will be removed, as will all data held within Qualtrics survey software after it has been downloaded and securely backed up on the institution's OneDrive cloud system. Where necessary, quotes used will be paraphrased if they include any data which is deemed to be traceable to any given participant. Should participants wish to withdraw their data during participation or within 14 days following participation, this data will also be permanently destroyed, although consent forms will be kept for auditing purposes.

Expected Outcomes

Phase one outcomes will feature an AI-generated piece of learning content suitable for a higher education audience which will be linguistically-similar (as indexed by sentiment analysis) to human-generated content.

For phase two, we will test a competing hypothesis pertaining to judgement of learning content that differs via its source of generation and three additional a-priori hypotheses. First, if ChatGPT (as informed by our prompting system) is able to capitalise on available data sources and present the resulting information in a clear and meaningful manner, then we would expect more positive student responses to AI- vs. human-generated content after controlling for label congruence, with the opposite result expected if the AI-generated content is not fit for purpose. Moreover, we hypothesise [1] an 'AI-generated' label to attenuate positive judgements, and positive judgements of AI-generated learning content to be associated with [2] extrinsic motivation, and [3] general AI supportive viewpoints.

For phase three, we anticipate more supportive viewpoints of AI-generated learning content from students who choose to learn online, relative to those who choose to learn on-campus.

Adverse Events (AEs)

There is no expectation of any adverse effects on participants within this study. Nevertheless, students will be provided with student-specific services should they wish to reach out for further guidance with their academic studies.

Withdrawals

Reasons for Withdrawal

Participants will be informed of their right to withdraw at several time points throughout the study. Participants can withdraw from phase two during participation (by closing their web browser) or after taking part in the study by emailing the primary researcher using their unique ID code. Participation in this study will be automatically terminated (via Qualtrics) should they decline to give consent. Participants can withdraw from phase three by bringing their intention to close the interview to the attention of the research team or by contacting the research team up to 14 days post-interview. They can also opt to omit any of the data they have provided from the analysis. Importantly, participants will not be expected or asked to provide a reason for withdrawal to remove any barriers for them doing so.

Handling of Participant Withdrawal

As previously stated, participants can withdraw at any time during the study (by closing their web browser during phase two or by bringing their wishes to the attention of the research team during phase three) or up to 14 days following participation (by e-mailing the research team with their unique identification code). Participants who withdraw from phase two will

not be replaced unless the sample falls under the requirements to ensure statistical power. Re-sampling of participants in phase three will be determined as to whether saturation of data has been reached.

Premature Termination or Suspension of Study

If there is a sufficient and reasonable cause, the study may be terminated or suspended. In such instances, the primary researcher will notify the University of Derby College of Health, Psychology and Social Care Research Ethics Committee providing an explanation for this action, such as the identification of AEs. The study may continue once the research team and research ethics committee are satisfied that any concerns have been addressed.

Statistical Analysis Plan

All analyses for this study have been determined a priori. For phase two, after cleaning the final dataset and checking parametric assumptions for conducting an ANCOVA, we will report descriptive statistics, as well as bivariate correlations between participant age, sex, academic motivation, acceptance of AI, and content judgements, for the whole sample, and for AI- and Human-generated content separately. Next, a 2x2 between group ANCOVA will be conducted whereby the first IV is content type (AI- vs. Human-generated), the second IV is accompanying label (congruent vs. incongruent), the covariates are participant age, sex, academic motivation, and acceptance of AI, and the DV is judgement score. We will calculate effect sizes and publish an open data set for scrutiny and replication.

Qualitative Analysis

For phase three, comparative deductive thematic analysis will be conducted comparing perceptions AI-generated learning content between the positions of students who choose to study online and on-campus. An epistemological perspective of social constructivism will be adopted as participants will respond to questions based on what they have learnt through social interactions and their interpretation and understanding of the interactions between themselves, academics, and their peers (or general observations made within society). By definition, following data coding, preconceived themes will be developed prior to immersion in the dataset from the online interviews, informed by theory and existing research. The two groups of interviewees (online students; n = 8, on-campus students; n = 8) will be directly compared to one another drawing from the six-step thematic analysis process defined by Braun and Clarke (2006), inspired by the work of Keenan et al. (2021) and Hammond et al. (2023). The diverging and converging themes between the two groups will be compared under the overarching theme of justification, in conjunction with the quantitative arm of the research.

Assessment of Safety

This study will follow the standard definition of AEs and report any AEs to the University of Derby College of Health, Psychology and Social Care Research Ethics Committee for up to 14 days after the final participant has completed the study. Should any AEs be identified, the primary research will assign a level of severity to it and assess the likelihood that the AE is due to study protocols. A risk assessment was completed prior to ethical approval, no risks were moderate or severe. Commercial risk was assessed (i.e., in the event that students reported dissatisfaction with the teaching material), however this is mitigated by students being prompted outside of this research to provide critical yet constructive course material throughout their programmes as a standard means of programme improvement.

Data Monitoring

The study will abide by the standards and requirements advised by the University of Derby, Good Clinical Practice, GDPR, and British Psychological Society. In phase two, due to data collection being anonymous, it will not be possible to follow-up on incomplete data or verify the accuracy of the data provided. However, the information provided to participants prior to participation aims to prevent error by providing clear instructions and a 'request response' function to encourage participants to complete all sections of the questionnaire; to comply with ethical guidelines, participants can skip questions if they do not wish to answer. In phase three, participants will be provided the opportunity to omit or re-word any data they provided to increase accuracy and anonymity. We will not follow-up with participants for ad-hoc clarifications.

Data Handling and Record Keeping

For phase two, data will be collected and maintained on Qualtrics until the required sample size has been achieved. Data will then be exported to an SPSS file format, backed-up, and deleted from Qualtrics. No identifying data will be obtained from participants and unique ID codes will be permanently deleted 14 days after the final participant completes the study. Anonymised data will be used for our analyses.

For phase three, participant's email addresses will be stored on OneDrive separately from transcripts and deleted 14 days following participation. Participants will be allocated pseudonyms so that they are not identifiable from their responses during the publication process. Interviews will be conducted by a member of the research team who does not have any student-facing job role requirements to prevent conflicts of interest and to increase the likelihood of complete and honest student responses.

Research Ethics Committee

The protocol, associated documents, questionnaire, and interview questions will be submitted to the University of Derby College of Health, Psychology and Social Care Research Ethics Committee for review,

feedback, and approval. Approval is required prior to any participation. Any amendments to the protocol will be subject to further review and approval by the ethics committee before any changes are implemented. Given data is anonymous, any data collected prior to such amendments will be treated in accordance with the procedures for which consent was obtained.

Consent Process

For both empirical phases of this research, consent will be sought via Qualtrics (online survey management software) following the provision of information about the study (e.g., inclusion criteria, process of withdrawal, data management, and contact details of the research team and services). Participants must affirm their consent to continue via a button press. If they fail to consent, Qualtrics will end their participation via an automated process and participants will be thanked for their time. Participants will not be expected to sign, date, or provide any identifiable information other than an e-mail address in phase three which will solely be used to communicate an appropriate time for interview via Microsoft Teams.

Protocol Deviation

Any protocol deviations from the ethically approved study will be reported to the University of Derby College of Health, Psychology and Social Care Research Ethics Committee in writing at the first available opportunity. Protocol deviations may be a consequence of the research team or participants; however, deviations are unlikely given the nature of the study.

Publication and Data Sharing Policy

The research team intends to publish the findings of this study in written and verbal form. The research team may also use the findings of this study as a guide for future research. At all stages, participants will remain anonymous and unique ID codes (phase two) and email addresses (phase three) will be deleted 14 days after the last participant has completed the study.

Study Personnel and Roles

Table 2 documents the members of the research team and their responsibilities throughout this project.

Table 2
Outline of research team personnel, role, and responsibilities

Personnel	Role	Responsibilities
Dr Gary F. Fisher	Primary Researcher	Study design; mentoring; study two data analysis; manuscript writing
Paula Shaw	Researcher	Study design; mentoring; study two data analysis; manuscript writing
Dr Dean Fido	Researcher	Study design; mentoring; study one data analysis; manuscript writing; protocol writing

Funding
This study has been funded internally by the University of Derby.

Declaration of conflict of interest
No conflict of interest.

Ethics Approval and informed consent
This study has received approval from a University of Derby research ethics committee (ETH2223-2632). All participants provided written informed consent prior to enrolment in the study.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... &Neelakantan, A. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165

Fido, D., & Harper, C. A. (2023). How to use ChatGPT to help close the awarding gap. Times Higher Education. Retrieved from: https://www.timeshighereducation.com/campus/how-use-chatgpt-help-close-awarding-gap

Groves, M. (2022). If you can't beat GPT3, join it. Times Higher Education. Retrieved from: https://www.timeshighereducation.com/blog/if-you-cant-beat-gpt3-join-it

Hammond, C., Fido, D., & Keenan, J. (2023, February 23). Exploring converging and diverging opinions of rehabilitative interventions for individuals who have committed serious criminal offences: The need for forensic-specific education in the general public. https://doi.org/10.31234/osf.io/49wq6

Ho, H. C. Y., et al. (2023). Promoting preservice teachers' psychological and pedagogical competencies for online learning and teaching: The T.E.A.C.H. program. *Computers & Education, 195,* 104725.

Keenan, J., Rahman, R., & Hudson, J. (2021). Exploring the acceptance of telehealth within palliative care: A self-determination theory perspective. *Health and Technology, 11*, 575584. https://doi.org/10.1007/s12553-021-00535-9

Kotera, Y., Conway, E., & Green, P. (2021). Construction And factorial validation of a short version of the Academic Motivation Scale, British Journal of Guidance & Counselling, https://doi.org/10.1080/03069885.2021.1903387

Woods, M. (2011). Interview Grouping for research and analysing qualitative data: An overview. [PowerPoint]. Massey University.

Woolston, C. (2018). Feeling overwhelmed by academia? You are not alone. Nature. Retrieved from: https://www.nature.com/articles/d41586-018-04998-1