

# Addressing the use of generative AI in academic writing

Johan van Niekerk<sup>a,b,\*,</sup>, Petrus M.J. Delpont<sup>a,b,</sup>, Iain Sutherland<sup>a,</sup>

<sup>a</sup> Noroff University College, Tordenskjolds gate 9, Kristiansand, 4612, Norway

<sup>b</sup> Nelson Mandela University, Gqeberha, 6019, South Africa

## ARTICLE INFO

### Keywords:

Evaluation methodologies

Pedagogical issues

Teaching/learning strategies

## ABSTRACT

The rise of generative AI has been a major disruptive force in academia. Academics are concerned about its impact on student learning. Students can use generative AI technologies, such as ChatGPT, to complete many academic tasks on their behalf. This could lead to poor academic outcomes as students use ChatGPT to complete assessments, rather than engaging with the learning material. One particularly vulnerable academic activity is academic writing. This paper reports the results of an *active learning* intervention where ChatGPT was used by students to write an academic paper. The resultant papers were then analysed and critiqued by students to highlight the weaknesses of such AI-produced papers. The research used the Technology Acceptance Model to measure changing student perceptions about the usefulness and ease of use of ChatGPT in the creation of academic text. A statistical analysis indicates the intervention's impact on their behavioural intentions on using ChatGPT for academic writing.

## 1. Introduction

The advent of advanced artificial intelligence systems, particularly generative pre-trained transformers like ChatGPT, has led to significant discussion in the academic world (Lingard, 2023). The emergence of this technology, its widespread accessibility and its use as a tool, raises concerns about its implications for education and learning (Darvishi et al., 2024). This paper aims to explore the impact of Artificial Intelligence (AI) technologies on student learning, particularly in the context of developing critical thinking and academic writing skills.

A key area of concern is the potential for AI technologies to transform how students engage with educational tasks. The ease and efficiency offered by AI technologies in generating written content are undeniably attractive to students, which unfortunately leads to students relying on the use of AI technologies to complete, rather than support, academic tasks as part of the learning process (Darvishi et al., 2024). AI technologies present a tempting shortcut when tasked with academic writing, in particular academic assessments. This reliance comes with significant longer-term drawbacks. If students increasingly outsource their writing and critical thinking to AI technologies, there is a risk of undermining their ability to formulate and express original ideas (Molenaar, 2022). The fundamental skills of structuring thoughts and presenting coherent arguments through written reports are crucial in both academic and

professional settings. The over-reliance on, or inappropriate use of AI technologies could jeopardise the development of these skills, leading to long-term implications for the student's intellectual growth and preparedness for real-world challenges. The convenience that the use of AI technologies offers must be critically examined, to explore ways to achieve responsible AI integration in education (Molenaar, 2022; Moorhouse et al., 2023; Chan, 2023).

The primary objective of the research was to determine whether students can be dissuaded from the *inappropriate use of generative AI* in support of academic writing. The research was based on two underlying assumptions: first, that there are aspects of the academic writing process where the use of generative AI might be appropriate; and second, that inappropriate use and student reliance on AI technologies can be addressed through an active learning intervention, which requires students to identify the shortcomings of AI-generated results by comparing them to established academic criteria. This intervention was introduced as part of an academic writing skills course at a University College in Norway, as further discussed in Section 4.1.

In addressing this growing dependency on AI technologies in academic settings the aim of this paper is to guide students towards a balanced use of technology, where AI supports rather than replaces the critical intellectual processes necessary for academic success. This paper contributes to the discourse on the use of generative AI in education. It

\* Corresponding author.

E-mail addresses: [johan.vanniekerk@noroff.no](mailto:johan.vanniekerk@noroff.no), [johan.vanniekerk@mandela.ac.za](mailto:johan.vanniekerk@mandela.ac.za), [johanvn@mandela.ac.za](mailto:johanvn@mandela.ac.za) (J. van Niekerk), [pieter.delpont@noroff.no](mailto:pieter.delpont@noroff.no), [petrus.delpont@mandela.ac.za](mailto:petrus.delpont@mandela.ac.za) (P.M.J. Delpont), [iain.sutherland@noroff.no](mailto:iain.sutherland@noroff.no) (I. Sutherland).

<https://doi.org/10.1016/j.caeai.2024.100342>

Received 10 June 2024; Received in revised form 5 December 2024; Accepted 8 December 2024

provides an example, supported by statistically significant empirical evidence, demonstrating that it is possible to address student over-reliance on, and inappropriate use of these tools through educational interventions. The use of an active learning approach provides students with the opportunity to identify the most appropriate and effective use of the tool.

The remainder of this paper is structured as follows: Section 2 provides background information necessary to understand the context of this study. Section 3 describes the research process of the study, and the methodology used is described in section 4. The results of the intervention and the implications are discussed in Section 5. Limitations of the study are highlighted in Section 6 and the paper concludes in Section 7.

## 2. Background and theoretical underpinning

This paper concentrated on ChatGPT version 3.5 (hereafter referred to as ChatGPT), as it was the most advanced and freely accessible generative AI available during the data collection period in August and September 2023. When prompted to describe itself, ChatGPT returned the response: “*ChatGPT refers to a language model developed by OpenAI called GPT-3.5. It is designed for natural language understanding and generation, enabling it to generate human-like text based on the input it receives. ChatGPT is often used in applications such as chatbots, virtual assistants, and other conversational interfaces, where it can understand and respond to user queries in a contextually relevant manner*” (OpenAI, 2024). It is a language model trained on a very large language data set (Wang et al., 2024b; Wolfram, 2023). The model computes scores for words within a particular sequence to map the relationships between words, enabling the model to predict the next word in a string of text. It will predict text based on the patterns learnt in training, which might include biased, contradictory, or incorrect information. It cannot provide information from outside the training set (OpenAI, 2024; Rice et al., 2024).

### 2.1. ChatGPT and academic writing

Critically for academic writing, the training set for ChatGPT excludes the majority of academic conferences and journal proceedings, which sit behind the various academic publisher’s paywalls. Nevertheless, ChatGPT is demonstrably useful as studies have shown how well AI technologies can be incorporated into education: For instance, utilising ChatGPT for creating tests (Cotton et al., 2023), for personalised student feedback (Wang et al., 2024b; Bommasani et al., 2021) and the potential reduction of teacher workloads (Guo et al., 2023). Studies have explored how it may be used in support of academic writing. Wang et al. (2024a) examined how academic writing might be supported by ChatGPT in self-directed learning activities.

Irrespective of these benefits, one needs to equally acknowledge the possibility of negative consequences, such as AI’s ability to “hallucinate”, provide incorrect or fabricated information, and more critically the inability to formulate and express original ideas (Molenaar, 2022). In exploring the ethical and responsible use of AI technology in teaching and learning, Chauncey and McKenna (2023) noted that the limitations of the generative AI needed to be understood, and that pushing beyond these limitations, in the case of ChatGPT, resulted in references being fabricated. The online interface and open access to ChatGPT make it inherently easy to use. This has a major impact on the likelihood of a student using this technology. One way to understand the likelihood of students using ChatGPT for academic writing, is to consider the Technology Acceptance Model.

### 2.2. Technology acceptance model

The Technology Acceptance Model (TAM) is a frequently used model to understand the intention to use a new technology. In the context of this study, the technology is ChatGPT. In its original form, Davis (1989) states, that a person’s attitude, as originally defined by Fishbein

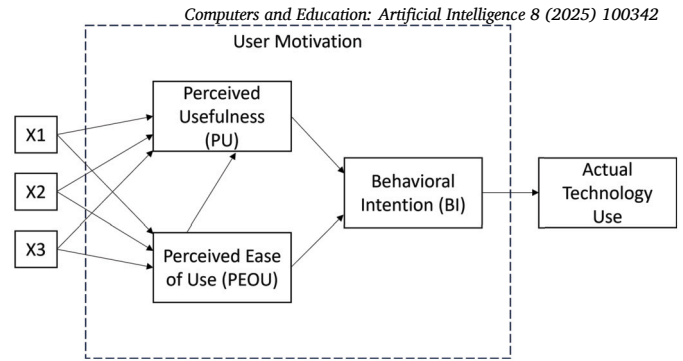


Fig. 1. Relationship of TAM constructs, adapted from Davis (1985).

and Ajzen (1977), towards a particular technology directly affects the intention to use that technology. This is alternatively referred to as Behavioural Intention (BI). In addition, Davis continues to state that the attitudes or BI of a person towards a specific technology are influenced by two main factors, as depicted in Fig. 1.

These factors are firstly the Perceived Usefulness (PU) of the technology, and secondly, the Perceived Ease of Use (PEoU) of the technology. In this context, PU is defined as “*the degree to which a person believes that using a particular system [or technology] would enhance his or her job performance*” (Davis, 1989). In contrast, PEoU, refers to “*the degree to which a person believes that using a particular system [or technology] would be free of effort*” (Davis, 1989). Thus, in summary, one can state that the BI of a person to use a specific technology is based on causal relationships (represented by the arrows in Fig. 1) with a person’s PEoU and PU of a specific technology (represented by X1 - X3 in Fig. 1). This relationship between BI, PEoU, and PU has been utilised by many studies to research the acceptance and actual usage of new technologies (Wang et al., 2023; Rejali et al., 2023). This includes a study by Rahman et al. (2023) that broadly examined the factors triggering students’ intention to use ChatGPT. In addition, work by Saif et al. (2024) incorporated the TAM to examine the relationship between final year degree student’s stress and anxiety and the use of ChatGPT to complete academic assignments. This work highlighted a number of positive correlations, but also an interesting negative correlation between PEoU and BI. The work concluded there was a need for appropriate mechanisms to guide students in the ethical use of ChatGPT to support their learning activities.

This study utilises TAM to explore student perceptions of the usefulness of ChatGPT for academic tasks. It is acknowledged that students exposed to the ChatGPT tool, are likely to view this tool as having both a high PEoU and a high PU which, in turn, results in a student having an increased BI to use this tool. The high PEoU and PU are likely due to ChatGPT requiring only a short prompt to produce grammatically correct, well-written and extensive prose. However, the limitations of the text may not be clearly understood.

The objective of this study is to enable students to discover weaknesses in the accuracy and reliability of content produced by ChatGPT. The goal is to influence students’ PU and PEoU, thereby reducing their BI to use ChatGPT for academic writing by highlighting its limitations. The study aims to encourage students to rely on ChatGPT as a tool for *polishing* rather than *producing* academic texts, promoting the development of their own ideas and structures. The TAM was used to inform the active learning pedagogical strategy in the intended intervention.

### 2.3. Active learning

Active learning methods encourage students to participate in the learning process rather than be passive observers of presented learning material. It can be defined as “*...instructional activities involving students in doing things and thinking about what they are doing.*” (Bonwell & Eison, 1991). Active learning is commonly described as a spectrum from simple activities such as writing and reflection, to more complex expe-

rential learning. There is a substantial body of evidence suggesting that incorporating active learning into a course, with students engaging with the material, results in better student outcomes (Yannier et al., 2021).

Students can engage with learning material in a number of ways; by discussing issues relating to the material or as in the case of this work, by creating an artefact in the form of a ChatGPT essay and then actively investigating relevant issues. In this case the students were asked to examine the form and features of the text, and the veracity of the resources ChatGPT has claimed to use in the essay text. To this extent, an active learning approach is well-suited as an intervention to adequately influence students' PU, PEoU and eventually their BI towards using ChatGPT within their academic writing. The following subsection 2.4 will discuss the core academic writing skills that universities strive to instil in their students, outlining seven key characteristics of good writing focused on by this study.

#### 2.4. Academic writing skills

Writing is the primary means of formal communication in academia. As such, acquiring strong academic writing skills is one of the most important abilities a student can gain. Academic writing skills serve not only as a medium for articulating knowledge and comprehension within a student's fields of study, but also as a framework for constructing and presenting sophisticated arguments and hypotheses. The proficiency in academic writing is tantamount to the ability to engage in critical thinking, synthesise research findings, and evaluate scholarly discourse (Du Preez & Fossey, 2012).

This section outlines the pivotal role that academic writing plays in a student's educational journey, outlining its impact on academic success and the broader implications for professional competencies in various career paths. Robust academic writing abilities require the development of several key characteristics. These characteristics were identified from the existing academic writing skills course material which was based on a variety of resources, including Paltridge et al. (2009), Oshima et al. (2007), Coffin et al. (2005), Purdue Online Writing Lab (2024):

The first characteristic (C1) of academic writing identified was the need for a clear **structure and flow**. The students targeted in the research intervention are developing their knowledge and skills in a variety of computing disciplines including Applied Data Science, Digital Forensics, and Cyber Security. The structure of the writing is closely related to the type of reports they are required to produce. If the type of report is based on empirical data the structure will be different from a report that is based primarily on arguments.

The second characteristic (C2) is that academic writing should be **evidenced and referenced**. This evidence can take a number of forms: It may be primary data supporting research arguments. Evidence could also be mathematical proofs and similar or it could be in the form of literature supporting the research arguments. When citing literature, it is extremely important that the cited sources are valid and of high quality (usually most easily assessed based on the reputation of the author and/or journal).

The third identified characteristic (C3) was the need to show **critical insight**. Rather than regurgitating the opinions of others, students should be reading literature, considering the implications of literature, and expressing their own opinions on the topic in their writing.

The fourth characteristic (C4) is that good academic writing is **balanced and unbiased**. Academic writing should consider all available evidence. When there are contrary opinions, clear reasoning should be given as to why a specific opinion is valued over another. This fourth characteristic is closely related to the second and third characteristics mentioned previously.

The fifth characteristic (C5) is **conciseness**. Well-written academic text is clear and to the point. It should be simple to read and easy to understand. Overly complex language and verbose or imprecise text are inappropriate for academic writing.

The sixth characteristic (C6) that was identified relates to **objectivity** in phrasing. The writer needs to present arguments in the third person, as often required by the computing discipline. The writing should also avoid emotionally laden words or terms when presenting evidence. For instance: "*In his fantastic paper on...*"

Lastly (C7), the writing needs to have a **formal tone**. In addition to being related to the fifth and sixth points above, the text should avoid contractions and colloquial expressions. It should avoid rhetorical questions.

An understanding of the characteristics of good academic writing is a core learning outcome for the academic writing skills course. The authors believe that students will be less inclined to use ChatGPT for their work if they recognise that ChatGPT does not necessarily produce writing that adheres to these characteristics. To ensure the achievement of this shift in opinion, it is necessary for students to actively experience the limitations of ChatGPT. The authors believe the realisation of ChatGPT's limitations will affect the BI to use ChatGPT for their writing by reducing the PU and PEoU.

### 3. Research process

The purpose of the intervention was to disillusion the students; and to clarify their perception of ChatGPT. In particular to reduce their BI to use ChatGPT for tasks relating to academic writing by impacting the way students perceived ChatGPT, both in terms of the PU and PEoU. Influencing the student's perspective on PU requires the student to develop an understanding of the limitations of ChatGPT on academic writing. Once the limitations of ChatGPT are understood then PEoU might also be lowered when students realise additional work is required to determine if the output from ChatGPT is inconsistent and that both facts and supporting references provided by ChatGPT require checking.

Three areas of the literature were examined to provide context for this study. Firstly, a suitable model was required to frame the students' approach to the technology (ChatGPT) as discussed in Section 2.2 on TAM. The student's skill set most affected by ChatGPT is that of academic writing, one of the most important student skill sets in academia. Therefore, as a second area of literature, seven key aspects of academic writing were identified in Section 2.4. A series of interventions were then required to achieve the desired change in BI. Thus the third area of literature included active learning, as outlined in Section 2.3, which was deemed the most appropriate approach to achieve the intervention with the intention being that the students, through their own work, uncovered the limitations of the ChatGPT application.

The intervention design was a multi-stage instrument as described in Fig. 2 with students being asked to create, and then critique an essay developed from a prompt using ChatGPT. The exercise was part of the assessment strategy for the academic writing skills course in the autumn semester of 2023 (described further in section 4.1). It should be noted that although this study pre-dates the recommendations made by Moorhouse et al. (2023) it complies with the proposed idea of a GAI literate assessment as suggested by Moorhouse et al. (2023). Nevertheless, there are several stages to the intervention.

#### 3.1. Stage one: essay creation

An overview and demonstration of ChatGPT were presented to the students explaining the intended task and briefly outlining the features of ChatGPT. This was a neutral walkthrough of ChatGPT functionality, explaining the concept of a prompt. The academic staff outlined the use and features of the tool without providing any opinion on the tool's suitability for the proposed task. This ensured that all students had an introduction to ChatGPT and had time to create a ChatGPT account prior to the next step in the task being released. The task for stage one, generating the essay was then released to students. This consisted of three contextualised prompts. These were three context-specific prompts to choose from based on the three degree programmes covered by the

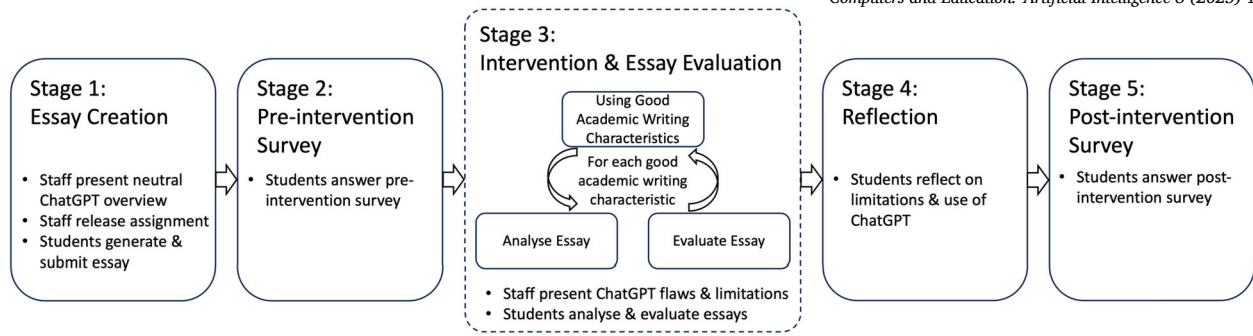


Fig. 2. Multi-stage intervention design.

Table 1  
Essay prompts for ChatGPT.

Degree Programme	Essay Prompt
Cyber Security	"Write 1000 words with at least five in text citations and references on the challenges of malware in Cyber Security"
Digital Forensics	"Write 1000 words with at least five in text citations and references on the importance of tool testing in Digital Forensics"
Applied Data Science	"Write 1000 words with at least five in text citations and references on research methods in Applied Data Science"

academic writing skills course. Students had to use one of the prompts described in Table 1 that was related to their specific degree programme.

When preparing the intervention it was noted that the ChatGPT did not always produce an essay with inline citations. In this case, students were allowed to re-prompt or add additional prompts if the first prompt did not provide references, or if the references were not from academic journals. Prompts they could utilise in this case could be:

- Repeat the request and ask for inline citations
- Repeat the request and ask for journal papers to be cited

Students were assigned an additional task of creating essays as part of the course assessment. They were given one week to complete their essays using ChatGPT, ensuring they had enough time to finish before completing the pre-intervention survey in stage two.

### 3.2. Stage two: pre-intervention survey

Students were asked to complete the pre-intervention survey to capture the students' thoughts on the use of ChatGPT. Anecdotal evidence at this stage was that the students were impressed with the functionality of ChatGPT, in particular those students who had not encountered the technology before.

### 3.3. Stage three: intervention & essay evaluation

In a second tutorial using ChatGPT, students were introduced to common issues with AI-generated text, particularly ChatGPT's tendency to produce unreliable references and "hallucinate" information. After this demonstration, students conducted a critical analysis of the essay generated by ChatGPT in response to a prompt they selected from Table 1, applying the predefined characteristics of good academic writing that they had learned earlier.

The exercise required students to assess the essay against each characteristic, as listed in Section 2.4, through structured questions. These questions were designed to guide the students in their evaluation, detailing specific aspects to consider for each characteristic. For instance, to assess the **formal tone** of the essay (C7), students were provided with additional guidance on how to assess the characteristic. For example, students were instructed that to maintain a **formal tone**, good writing will avoid contractions (e.g., replacing "can't" with "cannot," "I'm" with "I am").

Similarly, to evaluate the characteristic for academic writing to be **evidenced and referenced** (C2), students were tasked with evaluating a

specific journal reference cited in the essay generated by ChatGPT. Their first step was to verify the existence of the reference, ensuring it was not fabricated ("hallucinated") by the AI. If the reference was confirmed to exist, students then assessed its quality using the metrics provided by Scimagojr (2024). Additionally, they were instructed to analyse whether the reference and its use in the essay were contextually appropriate. This task leveraged the knowledge gained from other parts of the academic writing skills course, which covers the correct use of references and citations. Despite being first-semester undergraduates, the students were equipped to make informed judgements about the credibility and relevance of the citations.

Similar guidance was provided for each of the seven underlying characteristics. Based on their assessment, students then rated how well the essay met each academic writing requirement on a scale from 1 (poorly met) to 5 (excellently met). The authors believe that the evaluative process not only reinforced their understanding of academic writing characteristics but also honed their critical reading and analytical skills.

### 3.4. Stage four: reflection

The fourth stage of the intervention aimed to encourage students to reflect on their learning. The value of such reflective practice in learning has become well established (Pretorius & Ford, 2016). Students were first asked to complete a questionnaire on the learning management system (LMS) summarising their thoughts on ChatGPT as part of the mandatory monitoring process for the academic writing skills course. Thereafter, they had to write a short reflection to further crystallise their ideas on the issues identified with ChatGPT.

### 3.5. Stage five: post-intervention survey

The final part of the activity was to complete the post-intervention survey to determine if the interventions had impacted the students' BI regarding the use of ChatGPT. Participation in this was voluntary.

## 4. Methodology

The theoretical underpinnings of this study are provided by the TAM, the previously identified characteristics of good academic writing, and an active learning didactic approach. This section discusses the methodological considerations of this research, including how the different theoretical underpinnings were integrated for this study.



**Table 2**  
Intervention design stages mapped with Bloom's Taxonomy.

Stage	Bloom's Level	Associated Verbs
Stage 1	Level 1 - 2	Remembering, Understanding and Applying
Stage 2	Level 2 - 3	Understanding, Applying, Preparing and Executing
Stage 3	Level 4 - 5	Analysing, Evaluating, Deconstructing and Assessing
Stage 4	Level 5	Criticising, Predicting and Reflecting
Stage 5	Builds toward Level 6	Composing, Creating and Adapting

#### 4.1. Sample and setting

This study was conducted at Noroff University College (NUC), a private institution in Norway known for its applied degree programs. The focus was on undergraduate students from the computing faculty. The course included students from three different degree programmes: a Bachelor's in Cyber Security, a Bachelor's in Digital Forensics, and a Bachelor's in Applied Data Science. Notably, 41% of the students enrolled were female, reflecting a diverse cohort across gender and academic disciplines.

All computing undergraduates at NUC are required to undertake a foundational academic writing skills course during the first half of their first semester. This course was chosen for this study to provide early intervention at the outset of the students' academic careers. The authors of this study, having either delivered or contributed to the course over the past five years, are well-acquainted with the academic skills course content. Notably, during the 2023-2024 academic year, one of the authors was actively teaching the course.

During the 2023-2024 academic year a total of 173 undergraduate students were enrolled in the academic skills course. Participation in the active learning components of the study was mandatory, as these were integral to the course's curriculum and assessment methods. However, in adherence to the universities research ethics requirements, participation in the pre- and post-intervention surveys was both anonymous and voluntary. The survey was released to all of the students enrolled on the course.

#### 4.2. Intervention design

The research intervention aimed to reduce student reliance on material produced by ChatGPT by exposing weaknesses of using ChatGPT in a series of tasks embedded in the learning process. An active learning approach was used to involve students in both the creation and critique of the ChatGPT material. The intervention was designed with material being released to the students in stages, as previously discussed in Section 3 and depicted in Fig. 2.

The intervention stages were aligned with Bloom's taxonomy. Such alignment is a common pedagogical practice for all courses at the university. For each of these stages, various verbs were included to align with a study by Stanny (2016) on Bloom's taxonomy and associated verbs. Table 2 highlights the mapping of these verbs with the different intervention design stages. These verbs depict the natural student learning journey from level one to level six. For instance, as part of stages one and two, the student was required to *understand* the task at hand, *apply* the instructions for essay creation and complete the pre-intervention survey. Stage three required the student to *analyse, evaluate and assess* the created essay based on each of the academic writing characteristics. The foregoing stages equipped the student to criticise and reflect on how the created essay aligns with the academic writing characteristics, as presented by stage four. Finally, stage five aimed at influencing how students *compose* and *create* future academic tasks by facilitating adaptation of what they have learned, essentially impacting students' BI to use ChatGPT in academic writing tasks.

The initial step was exposing students to ChatGPT, without providing any background on reliability. Subsequently, students were required to *use* ChatGPT to create a thousand-word essay. The prompts in Table 1 were selected so that students would develop material relevant to their

degree programmes. It was also deemed important for all students in a degree program to use the same prompts to ensure comparable results and to demonstrate that ChatGPT responses to a fixed prompt will not necessarily be consistent. The additional prompts around inline citations and requesting journal papers provided an opportunity for individual students to fine-tune the text provided by ChatGPT.

After essay creation, an instructor leads session with staff highlighted some of the issues that might have been experienced with ChatGPT. The next stage in the intervention, continuing with the active learning approach, required students to *analyse* the produced text. Their assessment of the quality of the material was guided by a series of questions as discussed in Section 3.3. As part of the process, students were asked to focus on one of the journal references to verify its existence and, if it was real, assess the quality of the reference. The intention is that this would expose some of the limitations of ChatGPT and thereby reduce the students' reliance on using ChatGPT to find suitable academic references.

#### 4.3. Data generating instruments

The research instruments were designed based on two theoretical frameworks: TAM presented in Section 2.2 and the seven characteristics for good academic writing outlined in Section 2.4. A qualitative questionnaire was developed that included three main questions, each targeting one of the TAM constructs (PU, PEoU, and BI). Each of these main questions was subdivided into seven sub-questions reflecting the characteristics of good academic writing previously identified. Table 3 represents the mapping between the TAM constructs, good writing characteristics and survey question numbers.

Responses were gauged using a Likert scale. For PU and PEoU the scale ranged from "Strongly Disagree" to "Strongly Agree". For instance, the first sub-question (1.1) asked, "Using ChatGPT would improve the structure and flow of my writing.". For BI the scale ranged from "Very Unlikely" to "Very Likely".

The questionnaire was administered at two different times: before the intervention to establish a baseline, and after the intervention to assess any changes in students' perceptions about the utility of ChatGPT for academic writing. To integrate the questionnaires into the course and encourage participation, while maintaining ethical standards, the questionnaires were incorporated into the academic skills course's assessment pack and clearly communicated their non-compulsory nature, as detailed in Section 7.2 on ethical considerations.

The post-intervention questionnaire also had two optional, open-ended questions allowing students to share personal insights in their own words.

In this study, newly developed tools were used to measure PU, PEoU, and BI, rather than relying on existing validated instruments. While many validated instruments exist, they did not align precisely with the unique characteristics of this study, which focuses on academic writing and technology acceptance in the context of generative AI use in education. The newly developed tools were designed to capture specific attributes relevant to the academic writing process, which were considered critical for accurately assessing technology acceptance in this context.

The internal consistency of these newly developed scales was evaluated using Cronbach's *alpha*, which yielded values above 0.78 for all

**Table 3**

Mapping between TAM constructs (PU, PEoU and BI), good academic writing characteristics (C1-C7) and survey question numbers (Q1.1 - Q3.7).

Characteristic	Questions Related to PU	Questions Related to PEoU	Questions Related to BI
C1: Structure & Flow	Q1.1	Q2.1	Q3.1
C2: Evidence & References	Q1.2	Q2.2	Q3.2
C3: Critical Insight	Q1.3	Q2.3	Q3.3
C4: Balanced & Unbiased	Q1.4	Q2.4	Q3.4
C5: Conciseness	Q1.5	Q2.5	Q3.5
C6: Objectivity	Q1.6	Q2.6	Q3.6
C7: Formal Tone	Q1.7	Q2.7	Q3.7

**Table 4**

Cronbach alpha values for TAM constructs (PU, PEoU, and BI).

Construct	Pre-Test Cronbach Alpha	Post-Test Cronbach Alpha
PU	0.818	0.789
PEoU	0.852	0.810
BI	0.852	0.830

constructs (see Table 4). These results indicate that the scales demonstrated *good internal reliability* both before and after the intervention, supporting the validity of the instrument in measuring the TAM constructs (PU, PEoU, and BI). The intervention did not greatly affect the coherence of the questions, but there might have been slight shifts in how participants interpreted or responded to them post-intervention. Future studies could further refine these tools to improve consistency and reliability.

Additionally, changes in students' responses from pre- to post-intervention were analyzed using the *McNemar-Bowker test of symmetry* (Bowker, 1948), which is specifically suited for comparing distributional changes in *ordinal categorical data* across independent groups (Sheskin, 2020). This method was chosen as it directly assesses whether significant shifts occurred in the response patterns between the two time points, rather than focusing on internal scale consistency. The use of the McNemar-Bowker test was appropriate given the study's primary focus on detecting changes in the distribution of Likert scale responses across time (Davies, 2020).

The combination of these tests — Cronbach's alpha for internal consistency and the McNemar-Bowker test for comparing distributions — ensures that the data collected is both reliable and valid for the purposes of this study.

#### 4.4. Data collection

Data was collected anonymously using QuestionPro. Data was collected with two questionnaires. The questions and their relationship to the identified criteria for good academic writing can be seen in Tables 5, 6, and 7.

Participation in the active learning intervention was compulsory and part of the coursework. A total of 173 students participated. However, participation in the research component of the work was voluntary. The pre-intervention survey was answered by  $N = 159$  and the post-study survey by  $N = 95$  respondents.

#### 4.5. Data analysis

The results for the pre- and post-intervention questionnaires were analysed using the *McNemar-Bowker test*, implemented via the `statsmodels` library provided by Seabold and Perktold (2010). This test was deemed appropriate because it is specifically designed for *paired ordinal categorical data*, making it ideal for analysing the Likert scale responses gathered in this study. The McNemar-Bowker test is particularly suited for detecting significant changes in the distribution of responses between two time points (pre- and post-intervention) when the data is paired but not necessarily normally distributed. It tests for symmetry in

response shifts, making it highly applicable in this context where there is a need to look for shifts in categorical responses across time.

The McNemar-Bowker test does not require the assumption of normality or homogeneity of variances, making it appropriate for the ordinal nature of Likert-scale data. It is a non-parametric test that evaluates changes in response patterns, which allows for determining whether there were significant shifts in how participants answered the questionnaire before and after the intervention.

Results for these tests are listed in Tables 8, 9, and 10. Additionally, interpretative statistics from the reports provided by the QuestionPro survey platform were used for qualitative interpretation of the results. These results are shown in Tables 5, 6, and 7. Section 5 will discuss these results in more detail.

## 5. Results and discussion

### 5.1. Results

This study focused on using an active learning based intervention to develop students' awareness of the limitations of ChatGPT. The aim is to affect how the students assess ChatGPT in terms of PEoU and the PU. Ultimately, the goal is to dissuade students from over-reliance on and inappropriate use of AI technologies when writing academic material.

A comparison between the pre- and post-intervention survey questions examining PU (displayed in Table 5), PEoU (Table 6) and the impact on BI, (Table 7) indicates a change in student perception. This is clearly indicated in the heat maps presented by these tables which highlight the change in student perception of the underlying constructs of the TAM. The values in these tables show *the change in the underlying Likert scale value between the first and second surveys* as a result of the intervention.

#### 5.1.1. Perceived usefulness

The most significant changes in the students' PU, as detailed in Table 5, is the students' discovery that ChatGPT is unable to locate useful evidence to support academic arguments (Q1.2). Students' realisation that the training data used to create ChatGPT excludes most high-quality academic material led to a decrease in their initially high PU of the tool. Students have also realised that ChatGPT lacks critical insight in the subject area (Q1.3). The intervention had a very limited impact on the students' perception of ChatGPT as a writing tool. The students continued to value the usefulness of ChatGPT in support of improving the structure and flow of written text (Q1.1) and useful in supporting concise writing (Q1.5).

#### 5.1.2. Perceived ease of use

The most notable changes in the students' PEoU as detailed in Table 6 were also related to ChatGPT's ease of use in locating appropriate reference material. Many students were unaware that ChatGPT would "hallucinate" references that "sound plausible" by combining the names of real authors and journals with generated paper titles. After realising this, students no longer trusted the output from ChatGPT. However, the tool is still perceived as an easy way to improve the structure and flow of the text (Q2.1) and a useful tool to ensure a more concise block of text (Q2.5).

**Table 5**

Survey results: change in Perceived Usefulness (PU).

No.	Question Wording	Change
Q1.1	"Using ChatGPT would improve the <b>structure and flow</b> of my writing."	0.18
Q1.2	"ChatGPT can help me find <b>evidence</b> (references) to support arguments (statements) that I make when writing."	1.27
Q1.3	"ChatGPT can provide <b>critical insight</b> into the topics I want to write about."	1.05
Q1.4	"ChatGPT will help me to consider all viewpoints related to a specific issue (it is <b>not biased</b> )."	0.60
Q1.5	"ChatGPT helps me to write <b>concisely</b> (no unneeded text)."	0.17
Q1.6	"ChatGPT helps me to present arguments <b>objectively</b> (it focuses on evidence and not opinion)."	0.58
Q1.7	"ChatGPT helps me to write in a <b>formal tone</b> (good spelling and grammar, the correct tone, appropriate word choices, no colloquial language)."	0.25

**Table 6**

Survey results: change in Perceived Ease of Use (PEoU).

No.	Question Wording	Change
Q2.1	"It is easy to use ChatGPT to improve the <b>structure and flow</b> of my writing."	0.40
Q2.2	"It is easy to use ChatGPT to find <b>evidence</b> (references) to support arguments (statements) that I make when writing."	1.36
Q2.3	"It is easy to use ChatGPT to provide <b>critical insight</b> into the topics I want to write about."	0.73
Q2.4	"It is easy to use ChatGPT to help me consider all viewpoints related to a specific issue (it is <b>not biased</b> )."	0.47
Q2.5	"It is easy to use ChatGPT to help me write more <b>concisely</b> (no unneeded text)."	0.30
Q2.6	"It is easy to use ChatGPT to present arguments in a more <b>objective</b> way (to focus on evidence and not opinion)."	0.65
Q2.7	"It is easy to use ChatGPT to write in a <b>formal tone</b> (good spelling and grammar, the correct tone, appropriate word choices, no colloquial language)."	0.38

**Table 7**

Survey results: change in Behavioural Intention (BI).

No.	Question Wording	Change
Q3.1	"I would use ChatGPT to improve the <b>structure and flow</b> of my writing."	0.57
Q3.2	"I would use ChatGPT to find <b>evidence</b> (references) to support arguments (statements) that I make when writing."	1.36
Q3.3	"I would use ChatGPT to provide <b>critical insight</b> into the topics I want to write about."	0.87
Q3.4	"I would use ChatGPT to help me consider all viewpoints related to a specific issue (it is <b>not biased</b> )."	0.76
Q3.5	"I would use ChatGPT to help me write more <b>concisely</b> (no unneeded text)."	0.56
Q3.6	"I would use ChatGPT to present arguments in a more <b>objective</b> way (to focus on evidence and not opinion)."	0.81
Q3.7	"I would use ChatGPT to write in a more <b>formal tone</b> (good spelling and grammar, the correct tone, appropriate word choices, no colloquial language)."	0.65

**Table 8**

McNemar-bowker test results, mean, and median for perceived usefulness (PU) Q1.1 to Q1.7.

	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q1.6	Q1.7
McNemar-Bowker Value ( $\chi^2$ )	4.88	1.46	6.65	3.34	0.10	1.55	15.20
Degrees of Freedom (df)	4	4	4	4	4	4	4
P-value	0.027	0.227	0.010	0.068	0.752	0.213	<0.001
Pre-Intervention Mean	2.65	2.61	2.79	2.68	2.62	2.85	2.20
Pre-Intervention Median	4.0	4.0	4.0	3.0	3.0	3.0	4.0
Post-Intervention Mean	2.73	1.62	2.03	2.42	2.66	2.40	2.76
Post-Intervention Median	3.0	2.0	3.0	3.0	3.0	3.0	4.0

### 5.1.3. Behavioural intention

The impact of the intervention on the students' PEoU and PU described in the foregoing sections, and detailed in Tables 5 and 6, indicate that the intervention has been successful in regard to influencing aspects of the students' BI. Table 7 indicates that the most notable impact relates to the use of ChatGPT to find evidence to support arguments when writing academic text. The least impact of the intervention was on the students' PEoU and PU in regard to improving the structure and flow of the text and supporting the creation of concise text. This is reflected in the BI scores for questions 3.1 and 3.5.

From the foregoing, the results are analysed to determine if the changes are statistically significant. A statistical analysis using the McNemar-Bowker Test shows a statistically significant change in several of the underlying characteristics. This further reflected a change in the students' PEoU and PU resulting in a shift in students' BI regarding the use of ChatGPT for academic writing.

The McNemar-Bowker Test results for PU are detailed in Table 8 which indicates three statistically significant results. PEoU is detailed in Table 9 which indicates four statistically significant results. BI is detailed in Table 10 which similarly indicates six statistically significant results.

The implications of these results are discussed in the next section.

### 5.2. Discussion

The analysis of the pre- and post-intervention questionnaires indicates a statistically significant impact on student perception related to a number of the underlying characteristics of good academic writing. This is further supported by the qualitative analyses and, anecdotally, by open-ended comments made by students in the post-intervention survey. For the sake of clarity, these results will be discussed per characteristic C1-C7 (rather than specific questions, e.g. C1 rather than Q1.1, Q2.1 and Q3.1).

This study was aimed at impacting the BI of the students' use of AI technologies, in particular ChatGPT for academic writing-related tasks. The results of the study are represented visually in Table 11 which provides the pre-intervention percentage distribution of the students' responses across the Likert scale for BI for each of the academic writing characteristics. For comparison, Table 12 provides the post-intervention percentages for the same academic writing characteristics. As indicated by the McNemar-Bowker test results there is a significant shift in student intention towards the 'Very Unlikely' option of the Likert scale for some of the writing characteristics. The results of the McNemar-Bowker test statistical analysis were further supported by various open-ended comments provided by the students on the use of ChatGPT in academic

**Table 9**

McNemar-bowker test results, mean, and median for Perceived Ease of Use (PEoU) Q2.1 to Q2.7.

	Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q2.6	Q2.7
McNemar-Bowker Value ( $\chi^2$ )	21.33	0.01	1.20	1.38	7.29	5.31	18.75
Degrees of Freedom (df)	4	4	4	4	4	4	4
P-value	<0.001	0.924	0.272	0.241	0.007	0.021	<0.001
Pre-Intervention Mean	2.48	2.84	2.95	2.73	2.66	2.88	2.57
Pre-Intervention Median	4.0	4.0	4.0	3.0	4.0	3.0	4.0
Post-Intervention Mean	2.98	1.68	2.45	2.55	2.77	2.60	2.86
Post-Intervention Median	3.0	2.0	3.0	3.0	3.0	3.0	4.0

**Table 10**

McNemar-bowker test results, mean, and median for Behavioral Intention (BI) Q3.1 to Q3.7.

	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	Q3.7
McNemar-Bowker Value ( $\chi^2$ )	8.78	4.96	6.14	0.16	7.77	5.14	7.44
Degrees of Freedom (df)	4	4	4	4	4	4	4
P-value	0.003	0.026	0.013	0.693	0.005	0.023	0.006
Pre-Intervention Mean	0.64	0.51	0.72	0.90	0.70	0.88	0.60
Pre-Intervention Median	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Post-Intervention Mean	0.77	0.39	0.81	0.60	0.88	1.06	0.78
Post-Intervention Median	3.0	3.0	3.0	3.0	3.0	3.0	3.0

**Table 11**

Pre-intervention: percentage distribution of responses across likert scale related to Behavioural Intention (BI) per identified academic writing characteristic C1 - C7.

	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
C1	2.60	10.39	21.43	37.66	27.92
C2	9.52	17.01	17.01	35.37	21.09
C3	6.71	20.81	24.16	33.56	14.77
C4	5.52	13.10	30.34	37.93	13.10
C5	4.86	15.97	23.61	38.19	17.36
C6	4.23	16.90	29.58	35.21	14.08
C7	2.26	9.77	20.30	39.85	27.82

**Table 12**

Post-intervention: percentage distribution of responses across likert scale related to Behavioural Intention (BI) per identified academic writing characteristic C1 - C7.

	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
C1	7.61	20.65	26.09	34.78	10.87
C2	41.76	30.77	12.09	10.99	4.40
C3	26.37	27.47	25.27	19.78	1.10
C4	20.88	28.57	18.68	29.67	2.20
C5	12.50	25.00	28.41	27.27	6.82
C6	19.32	26.14	34.09	19.32	1.14
C7	9.20	20.69	25.29	34.48	10.34

writing, several key themes and sentiments emerged and these have been included in the statistical analyses. The results of the following characteristics (C1-C7) will now be discussed individually.

#### 5.2.1. C1: structure and flow

The analyses show that ChatGPT is perceived to be both useful (PU) and easy to use (PEoU) to improve the **structure and flow** of academic writing. Pre-intervention 65.58% of students indicated that they had a BI to use ChatGPT for this purpose, in contrast to 45.65% of students who intended to use it for this purpose post-intervention. Based on the McNemar-Bowker Test, changes in this intention were statistically **highly significant** ( $P\text{-value} \leq 0.01$ ). Furthermore, many students found ChatGPT helpful in structuring their writing and improving the flow of their work (C1). Positive comments highlighted its usefulness in organising and formatting essays and providing sentence and word suggestions as indicated by some students, such as:

*"I found it nice to use as a tool to structure my writings."*

#### 5.2.2. C2: evidence and references

The perception relating to both how useful (PU) and easy to use (PEoU) ChatGPT is to help find supporting **evidence and references** for academic writing has not been significantly changed by the intervention. However, the BI to use it for this purpose has changed. Pre-intervention, 56.46% of students had the BI to use ChatGPT to acquire evidence and references. Post-intervention only 15.39% had this BI. This change in BI was shown to be statistically **significant** ( $P\text{-value} \leq 0.05$ ). A significant number of students expressed concerns about the reliability of references generated by ChatGPT. Instances of incorrect or fabricated references led to a decreased trust in the tool for sourcing evidence, as highlighted by the following comment:

*"I did not know it could be as wrong as this. I've used it often in work relations for cleaning up text / helping me read through my text, but I never really thought it would give me made-up (wrong) references."*

#### 5.2.3. C3: critical insight

The data also indicates that the intervention had a statistically **highly significant** impact on the PU, but **no impact** on the PEoU. Subsequently, the BI to use ChatGPT to provide **critical insight** in academic writing showed a statistically **significant** ( $P\text{-value} \leq 0.05$ ) change. Pre-intervention, 48.33% of students had the BI to use ChatGPT for this purpose. Post-intervention only 20.88% had this BI. Some students noted that ChatGPT lacks critical thinking abilities, often producing content that seems robotic or lacking depth in analysis, as supported by the following comment:

*"Yes, it taught me that ChatGPT can be faulty if not instructed correctly. Also, the wording can be seen as robotic. What I was made most aware of was the lack of critical thinking which is vital in making scientific material."*

#### 5.2.4. C4: balanced and unbiased

The BI related to using ChatGPT to provide a **balanced and unbiased** viewpoint in academic writing changed from 51.03% to 31.87%. This was statistically **not significant**, suggesting students did not fully grasp the issues of potential bias with ChatGPT. However, several comments indicated a perceived bias in ChatGPT's responses, with students expressing scepticism about blindly accepting its results, as indicated in the following:

*"My opinion on using ChatGPT did not change (I will not use it), but I got a better understanding of WHY I should not use it."*



### 5.2.5. C5: conciseness

There was a statistically **highly significant** ( $P\text{-value} \leq 0.01$ ) change in BI to use ChatGPT to make writing more **concise**. However, there was **no significant change** in the underlying PU of ChatGPT for this purpose. This aligns well with the researchers' opinions that the tool is observably suitable and useful for this purpose, but not necessarily easy to use correctly. Some students appreciated ChatGPT's ability to generate concise talking points and summaries, aiding in clear and focused writing. This is supported by the following comment:

*"The advantages that I intend to utilise are the AI's ability to quickly output lists of concise talking points on specific subjects, and basic road-maps on organising and formatting essay-type documents."*

### 5.2.6. C6: objectivity

The number of students who had the BI of using ChatGPT to improve **objective presentation** in their writing changed from 49.29% to 20.46%. This change was statistically **significant** ( $P\text{-value} \leq 0.05$ ). While some students found ChatGPT useful for maintaining an objective tone, others were wary of the false information it sometimes provided, stressing the need for fact-checking. This is highlighted in the following student's comment:

*"Before using ChatGPT for this assignment, I thought that the references and sources would be real. But now I know that a lot of the information is false, and therefore it is important to check the sources before 'trusting' ChatGPT."*

### 5.2.7. C7: formal tone

Finally, the BI related to using ChatGPT to improve the **formal tone** of academic writing changed from 67.67% to 44.82%. This is a statistically **highly significant** ( $P\text{-value} \leq 0.01$ ) change, however, if one discards neutral responses, there are still more students intending to use ChatGPT for this purpose than those not intending to use it.

The authors believe that this is likely due to students, with English as a second language, particularly valued ChatGPT for its assistance in formalising their language, enhancing grammar, and refining tone. This is supported by the following student's comment:

*"ChatGPT gives me the option to look at several different sentences and words that would fit better in whatever I am writing. ChatGPT can also give a better structure and formal use of language."*

The foregoing discussions highlighted a mix of enthusiasm and scepticism on the overall BI to use ChatGPT. Some students were excited about integrating ChatGPT into their academic workflows for specific tasks, while others were cautious or outright dismissive, citing reliability issues and a preference for manual research and writing, as indicated by the following student's comment:

*"It did not change my opinion; as I have used it before for my IT hobbies. Still, it made me more aware of its inherent issues."*

In essence, students found ChatGPT particularly beneficial for enhancing the structure and flow of their writing, making their work more concise, and helping non-native English speakers achieve a more formal tone. However, concerns were raised about the reliability of references, the lack of critical insight, and the potential objectivity of responses. These issues led to decreased trust and a cautious approach to using ChatGPT. Consequently, while some students are enthusiastic about using ChatGPT for specific academic tasks, others remain sceptical and highlight the importance of careful fact-checking and limited reliance on the tool.

This implies an effect on the students' attitudes toward the PU and PEoU of ChatGPT. Subsequently, this indicates that the applied intervention was successful in reducing the BI to use ChatGPT for aspects of

academic writing. To completely dissuade students from the inappropriate use of AI technologies like ChatGPT, some further approaches need to be considered.

When reflecting on the heat map results in Tables 11, and 12, noticeable features between the pre- and post-intervention surveys include the student's perception of the use of ChatGPT to support C1, C3, C5, and C7 of academic writing; relating to the use of ChatGPT to support the structure and flow of writing, concise writing, critical insight, and improving the formal tone of the text. The change in C4, the ability to present a balanced and unbiased view, though visible, is not statistically significant. However, the related change in C6, objectivity, is significant. This might be related to a lack of understanding about what is meant by writing being balanced and unbiased and how it differs from being objective. In considering this cohort of students, the results may be impacted by the students' familiarity with the English language. Although language competency was not surveyed in this study, it is possible to examine the nationalities of the students in the academic skills course used in the study. As a Norwegian institution, unsurprisingly 71.0% of the 173 students on the course are Norwegian. The remaining 29.0% come from 30 other countries. Only six students (3.6%) come from countries where English is the first or an official language. It is possible to assume that for the majority of these students (96.5%), English is a second language with varying degrees of competency. One notable feature of ChatGPT is the quality of text from a grammatical and lexical perspective (Herbold et al., 2023). Students who have English as a second language may value the linguistic quality of the ChatGPT-produced text, as noted by Baek et al. (2024), but at the same time struggle to criticise the output from ChatGPT, and find it harder to recognise the output in some cases as too generalised.

This may also explain the slight increase in post-intervention scores for C5, which relates to producing concise text (see Table 12). These scores reflect the students' BI to use ChatGPT to improve their writing style, including tone, appropriate word choices, and avoiding colloquial language.

## 6. Limitations of the study

This was a single sample from one student cohort in the 2023-2024 academic year. Further longitudinal studies, following this, and other cohorts, would provide an indication of the overall success of the intervention. Students resort to using bad academic practices; tools and tactics such as the use of ChatGPT, due to the pressures of poor time management and imminent assessment deadlines. Under these circumstances, students may still be tempted to use ChatGPT to provide material for an assessment.

One limitation of the study was the lack of detailed demographical data regarding the participants. The authors did not gather data regarding possible prior exposure to AI or digital tools. Due to the relative novelty of ChatGPT at the time of the study, the authors assumed that the students would have had limited to no such prior exposure. It would be beneficial for future studies to gather such data. Another limitation of the study is the observed dropout rate between the pre- and post-intervention questionnaires, which could introduce potential biases in the results. A reduction in the number of participants between the two time points can lead to non-response bias, where the participants who completed the post-test may not be fully representative of the original group. This selective attrition could skew the results if, for example, students with a more positive or negative view of the intervention were more likely to complete the post-test. Since the authors did not gather comparable demographic data in the pre- and post-tests, the possibility of unmeasured factors influencing dropout remains. Future research could address this limitation by employing strategies such as incentivising participation to improve response rates, using longitudinal methods to track participants, or employing imputation techniques to account for missing data.

This paper focused on the freely accessible version of ChatGPT (version 3.5). Applications like ChatGPT continue to evolve in terms of functionality which may reduce the effectiveness of the intervention. However, the authors believe that the underlying philosophy of the intervention would be useful in future interventions. It would be extremely difficult to completely prevent the use of AI in academic writing since these tools clearly have a high PU and PEoU. Interventions should thus rather focus on teaching students how to use such tools appropriately and allow students to discover that AI cannot replace their thinking. An active learning approach is especially suitable for this intervention.

## 7. Conclusion

The Technology Acceptance Model (TAM) highlights a user's motivation, and attitude, to adopt a specific technology. This motivation is affected by the Perceived Ease of Use (PEoU) and Perceived Usefulness (PU). Affecting either of these factors can increase or reduce the user's Behavioural Intention (BI), the desire to use that specific technology. This study aimed to reduce the user's BI to use ChatGPT for academic purposes.

A series of interventions were used that incorporated active learning methods to encourage the students to engage with the material and uncover the limitations of ChatGPT. These interventions impacted PU by highlighting the limitations of ChatGPT in academic writing. The same pattern of interventions also impacted the student's PEoU for ChatGPT as using the generated text would at least necessitate the checking of all references, and possibly searching for replacement references.

The interventions used in this study have been shown to have a statistically significant impact on the BI of students to use ChatGPT and thus to dissuade students from over-relying on and inappropriately using ChatGPT for academic writing purposes. The results of the paper demonstrate that it is possible to address student over-reliance and inappropriate usage of generative AI tools through educational interventions.

### 7.1. Implications

The results of this paper demonstrate the effectiveness of addressing student over-reliance on and inappropriate usage of generative AI tools through educational interventions. Students require clear guidance on the use of generative AI (Saif et al., 2024). This paper implies that a holistic approach is required; in addition to ethical guidelines, students also require exposure and training in the use of the technology. This is consistent with similar empirical research on the use of generative AI in writing within self-directed learning (Wang et al., 2024a). Students were able to show a high degree of responsibility towards the use of ChatGPT including the critical assessment of AI-generated information, but Wang et al. (2024a) noted that this required some subject matter expertise and the cognitive skills to make the assessment of the material.

The educational interventions as described in this paper provide an opportunity to explore the strength and weaknesses of the technology and to understand appropriate use. Given the fast pace at which generative AI currently evolves, the specific intervention might not remain effective for much longer. However, the underlying principle of using active learning approaches whereby students engage with the technology and are guided towards discovering the weaknesses in answers produced by such technology, is one which the authors believe could be generalised to future interventions.

### 7.2. Future work

As mentioned, a limitation of this study is that work was undertaken with ChatGPT3.5, a repeat of the study with newer versions of GPT will indicate the continued viability of the intervention. In addition, further work is required to determine if the effects of this intervention have a lasting influence on the students participating in the study. Similar

studies can be undertaken with this student cohort at later stages in the degree programme, but this will also need to account for evolving technology. Further work is also required to develop interventions that can be used in other subject areas impacted by ChatGPT such as programming.

## CRedit authorship contribution statement

**Johan van Niekerk:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Petrus M.J. Delpoort:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Conceptualization. **Iain Sutherland:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization.

## Statement on open data and research ethics

This research project adheres to all requirements for research ethics as stipulated by the Norwegian National Committees for Research Ethics (2014) and received ethics approval from Noroff University College. The study was approved by the ethical committee with ID: NUC-2023-01. Informed consent was obtained from all participants, and their privacy rights were strictly observed. The data can be obtained by sending request e-mails to the corresponding author.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Baek, C., Tate, T., & Warschauer, M. (2024). "chatgpt seems too good to be true": College students' use and perceptions of generative AI. *Computers and Education: Artificial Intelligence*, Article 100294. <https://doi.org/10.1016/j.caeai.2024.100294>. <https://www.sciencedirect.com/science/article/pii/S2666920X24000973>.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint. arXiv:2108.07258.
- Bonwell, C. C., & Eison, J. A. (1991). Active learning: creating excitement in the classroom. ASHE-ERIC Higher Education Report No. 1. Washington, D.C.
- Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43, 572–574.
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20. <https://doi.org/10.1186/s41239-023-00408-3>.
- Chauncey, S. A., & McKenna, H. P. (2023). A framework and exemplars for ethical and responsible use of AI chatbot technology to support teaching and learning. *Computers and Education: Artificial Intelligence*, 5, Article 100182. <https://doi.org/10.1016/j.caeai.2023.100182>. <https://www.sciencedirect.com/science/article/pii/S2666920X23000619>.
- Coffin, C., Curry, M. J., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2005). Teaching academic writing: A toolkit for higher education. Routledge.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1–12. <https://doi.org/10.1080/14703297.2023.2190148>.
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers and Education*, 210, Article 104967. <https://doi.org/10.1016/j.compedu.2023.104967>. <https://linkinghub.elsevier.com/retrieve/pii/S0360131523002440>.
- Davies, R. S. (2020). Data analysis pre-post example. [https://edtechbooks.org/designing\\_surveys/data\\_analysis\\_pre\\_po](https://edtechbooks.org/designing_surveys/data_analysis_pre_po). (Accessed 15 October 2024).
- Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results. Ph.D. thesis. Massachusetts Institute of Technology.

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319. <https://doi.org/10.2307/249008>.
- Du Preez, I., & Fossey, A. (2012). Developing academic writing skills as part of graduate attributes in undergraduate curricula. *South African Journal of Higher Education*, 26, 346–357.
- Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.
- Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How close is chatgpt to human experts? Comparison corpus, evaluation, and detection. *arXiv:2301.07597*.
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus chatgpt-generated essays. *Scientific Reports*, 13, Article 18617.
- Lingard, L. (2023). Writing with chatgpt: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education*, 12, 261.
- Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, 57, 632–645. <https://doi.org/10.1111/ejed.12527>.
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, Article 100151.
- OpenAI (2024). ChatGPT (4) [large language model]. <https://chat.openai.com>. (Accessed 21 February 2024).
- Oshima, A., Hogue, A., et al. (2007). Introduction to academic writing. Pearson/Longman.
- Paltridge, B., Harbon, L., Hirsh, D., Shen, H., Stevenson, M., Phakiti, A., & Woodrow, L. (2009). *Teaching academic writing: An introduction for teachers of second language writers*, vol. 10. Ann Arbor, MI: University of Michigan Press.
- Pretorius, L., & Ford, A. (2016). Reflection for learning: Teaching reflective practice at the beginning of university study. *International Journal of Teaching and Learning in Higher Education*, 28, 241–253.
- Purdue Online Writing Lab (2024). Purdue University. Academic writing. [https://owl.purdue.edu/owl/general\\_writing/academic\\_writing/index.html](https://owl.purdue.edu/owl/general_writing/academic_writing/index.html). (Accessed 16 April 2024).
- Rahman, M. S., Zhang, J., Moral, I., & Hossain, G. (2023). Examining students' intention to use chatgpt: Does trust matter? *Australasian Journal of Educational Technology*, 39, 51–71. <https://doi.org/10.14742/ajet.8956>.
- Rejali, S., Aghabayk, K., Esmaeli, S., & Shiwakoti, N. (2023). Comparison of technology acceptance model, theory of planned behavior, and unified theory of acceptance and use of technology to assess a priori acceptance of fully automated vehicles. *Transportation Research Part A: Policy and Practice*, 168, Article 103565. <https://doi.org/10.1016/j.tra.2022.103565>. <https://linkinghub.elsevier.com/retrieve/pii/S0965856422003160>.
- Rice, S., Crouse, S. R., Winter, S. R., & Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. *Technology in Society*, 76, Article 102426. <https://doi.org/10.1016/j.techsoc.2023.102426>. <https://linkinghub.elsevier.com/retrieve/pii/S0160791X23002312>.
- Saif, N., Khan, S. U., Shaheen, I., ALotaibi, F. A., Alnfai, M. M., & Arif, M. (2024). Chat-gpt; validating technology acceptance model (tam) in education sector via ubiquitous learning mechanism. *Computers in Human Behavior*, 154, Article 108097. <https://doi.org/10.1016/j.chb.2023.108097>. <https://www.sciencedirect.com/science/article/pii/S074756322300448X>.
- Scimagojr (2024). Scimago journal & country rank. <https://www.scimagojr.com/>. (Accessed 2 February 2024).
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th python in science conference*.
- Sheskin, D. J. (2020). *Handbook of parametric and nonparametric statistical procedures*. CRC Press.
- Stanny, C. J. (2016). Reevaluating bloom's taxonomy: What measurable verbs can and cannot say about student learning. *Education Sciences*, 6. <https://doi.org/10.3390/educsci6040037>. <https://www.mdpi.com/2227-7102/6/4/37>.
- Wang, C., Ahmad, S. F., Bani Ahmad Ayassrah, A. Y., Awwad, E. M., Irshad, M., Ali, Y. A., Al-Razgan, M., Khan, Y., & Han, H. (2023). An empirical evaluation of technology acceptance model for artificial intelligence in E-commerce. *Heliyon*, 9, Article e18349. <https://doi.org/10.1016/j.heliyon.2023.e18349>. <https://linkinghub.elsevier.com/retrieve/pii/S2405844023055573>.
- Wang, C., Li, Z., & Bonk, C. (2024a). Understanding self-directed learning in AI-assisted writing: A mixed methods study of postsecondary learners. *Computers and Education: Artificial Intelligence*, 6, Article 100247. <https://doi.org/10.1016/j.caeai.2024.100247>. <https://www.sciencedirect.com/science/article/pii/S2666920X2400050X>.
- Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024b). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity*, 51, Article 101440. <https://doi.org/10.1016/j.tsc.2023.101440>. <https://linkinghub.elsevier.com/retrieve/pii/S1871187123002080>.
- Wolfram, S. (2023). *What is ChatGPT doing... and why does it work?* Wolfram Media, Incorporated.
- Yannier, N., Hudson, S. E., Koedinger, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Munakata, Y., Doebel, S., Schwartz, D. L., Deslauriers, L., McCarty, L., et al. (2021). Active learning: "hands-on" meets "minds-on". *Science*, 374, 26–30.