

Leveraging ChatGPT in academic writing: ChatGPT enhances students' writing quality, writing experience, and ownership

Nicolas Bauer^{a,*}, Tim Fütterer^b, Birgit Brucker^a, Peter Gerjets^a

^a Leibniz-Institut für Wissensmedien, Tübingen, Germany

^b University of Tübingen, Tübingen, Germany

ARTICLE INFO

Keywords:

ChatGPT

Academic writing

Large language models

Higher education

ABSTRACT

The role of ChatGPT in promoting students' academic writing skills – especially when writing without ChatGPT after having used it – has not yet been empirically tested. In an experimental setup, first-year fashion design students engaged in three in-class writing exercises: composing introductions to scientific papers first without AI assistance, second with ChatGPT, and third again without assistance. Our findings demonstrate ChatGPT's statistically significant potential to improve students' writing quality, both during and immediately after its use. ChatGPT appears to be an effective and efficient tool for enhancing quality in academic settings, improving text quality (e.g., examiners' ratings, readability, cohesion) and subjective writing experience (e.g., ease, perceived control, cognitive load). These benefits are observed not only during its use but also when students write again without AI assistance. Moreover, whereas students' sense of authorship is reduced during the use of ChatGPT, their perceived ownership over their work increased in the third writing exercise as compared to their first exercise, suggesting a more nuanced understanding of feelings of ownership, which might help students to find a stronger connection and sense of achievement during academic writing. ChatGPT might help in learning academic writing. Additionally, we discuss findings and implications for using ChatGPT in education.

1. Introduction

Academic writing is essential for helping students develop effective communication skills to convey complex ideas clearly, ensuring readers understand the intended information without confusion. Furthermore, academic writing fosters critical thinking and helps students follow specific guidelines [1,2]. ChatGPT has extraordinary text-generation capabilities, producing short stories and essays that can be more advanced than those of undergraduate students [3]. It can help summarize sources, refine language, and provide formative feedback on grammar, vocabulary, and overall structure, particularly assisting novice scholars to express themselves in an academically valid way [4,5].

Despite the considerable potential this technology offers, significant risks in its use in academic writing include plagiarism, fabricated references, diminished critical thinking, overreliance, and challenges in maintaining originality [6–8]. As user numbers increase, overreliance on ChatGPT during learning can negatively affect students' progress [9] or may lead students to struggle to express their own thoughts [10–13].

This may affect their intrinsic motivation, writing, and learning experience in academic writing [14], as well as their ownership of the texts they produce [15]. Conversely, some research suggests that ChatGPT may improve students' writing experience [16,17]. The relationship between ChatGPT and students' writing experience is a nuanced issue that warrants further investigation, particularly regarding experiences when access to ChatGPT is withdrawn [18].

Recent research has begun to investigate the impact of ChatGPT on students' writing development, with positive increases in writing quality indicators often observed during its use [19,20]. Some studies highlight the potential for learning effects when writing is approached with ChatGPT assistance, although the methods used varied [21–23].

Despite these initial results, few studies have experimentally assessed ChatGPT's impact on students' academic writing quality, either during the writing process or afterwards, when the tool may no longer be available.

In this study, we explore how students use ChatGPT in academic writing settings and how its use affects their writing quality, subjective writing experience, and sense of ownership, both during and after use.

* Corresponding author at: University of Tübingen, Europastraße 6, 72072 Tübingen, Germany.

E-mail addresses: nicolas.bauer@student.uni-tuebingen.de (N. Bauer), tim.fuetterer@uni-tuebingen.de (T. Fütterer), b.brucker@iwm-tuebingen.de (B. Brucker), p.gerjets@iwm-tuebingen.de (P. Gerjets).

<https://doi.org/10.1016/j.caeo.2026.100351>

Received 22 May 2025; Received in revised form 22 March 2026; Accepted 4 April 2026

Available online 5 April 2026

2666-5573/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

We employed an ABA-within-participants experimental design in which students wrote three introductory sections of academic papers. By comparing the results of both the ChatGPT-assisted exercise and the post-ChatGPT exercises with the baseline, we aim to draw valuable conclusions towards potential in-class applications of ChatGPT in academic writing. We have visualized our study's framework in a flowchart (see Fig. 1).

2. Theoretical background

2.1. ChatGPT's potential for improving academic writing performance and quality

2.1.1. Writing performance and quality measures

To assess whether ChatGPT could have a pedagogically functional

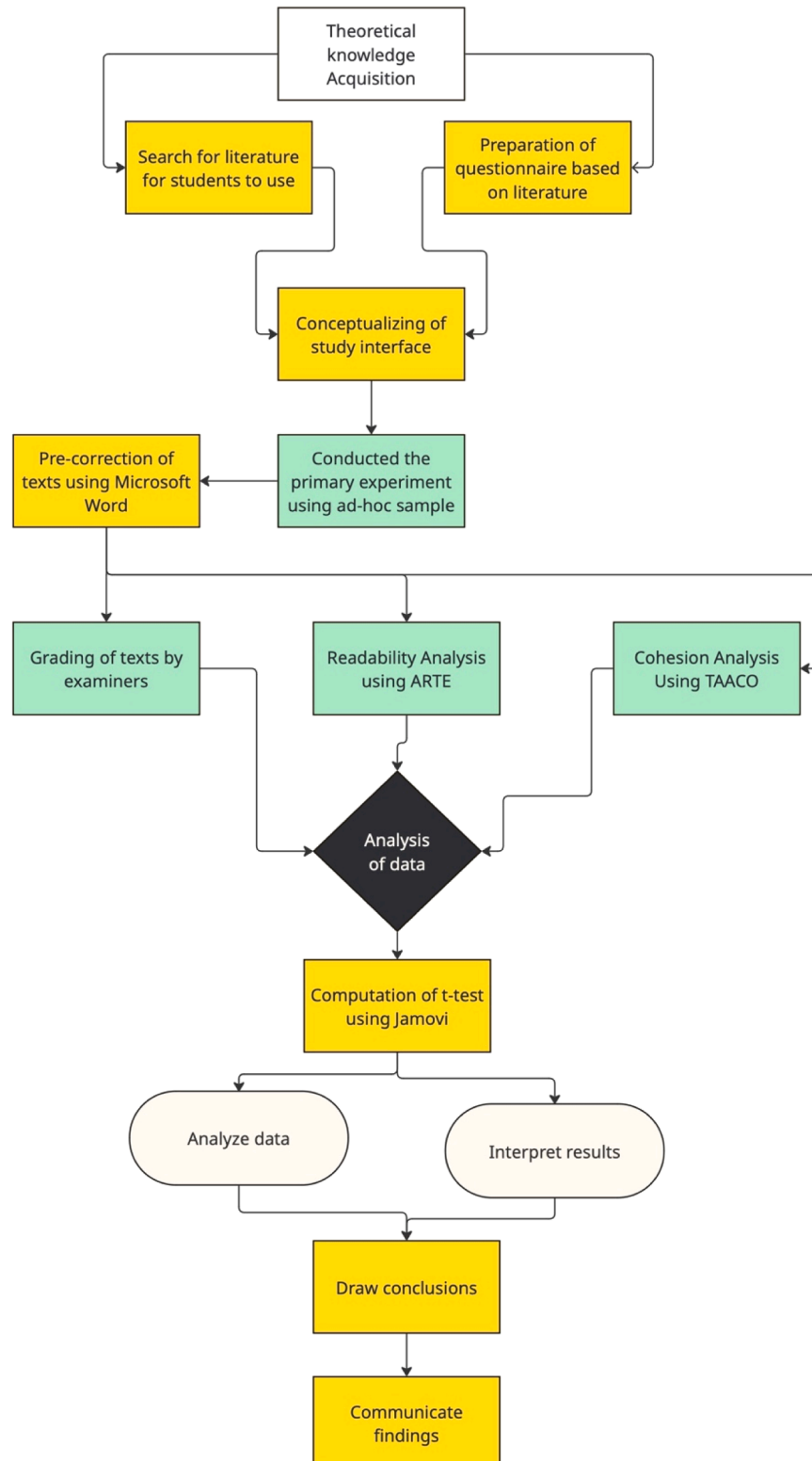


Fig. 1. Study Flowchart.

role in academic writing, we first aim to review studies on students' writing performance and quality improvement without AI. Writing performance, particularly in the academic writing domain, is defined as the ability to produce written text and is related to writing processes and skills [24]. Writing performance is closely linked to a student's writing quality.

Writing quality is defined as the ability of a written text to effectively "communicate or achieve its purpose with its intended audience" [25, 26]. Writing quality comprises measurable components of a student's work that can be assessed by factors such as cohesion, organization, argumentation, linguistic accuracy, and the ability to communicate ideas effectively [27, 28]. Writing quality also includes lexical diversity and syntactic complexity, which are closely related to text readability [29, 30].

2.1.2. Improving academic writing performance and text quality

Regarding improving academic writing quality, three important aspects discussed in the literature may be crucial to consider when evaluating ChatGPT's potential benefits for academic writing.

First, Van Der Loo et al. [31] compared different instructional methods in academic writing. They found that observational learning was more effective than learning by doing, particularly for students categorized as *planners* (outlining before writing), whereas *revisers* (first writing, then revising) showed no difference. Thus, one potential of ChatGPT is that it allows students to observe alternative ways an AI would approach a writing task. If used correctly, both *planners* and *revisers* might benefit from this input.

Second, Braaksma et al. [32] found that weaker students benefited from observing peers with similar skills, whereas good students improved from observing good models. Therefore, ChatGPT's ability to produce text at different levels of aspiration could enable its use as an adaptive learning tool, providing opportunities for observational learning across proficiency levels based on an individual's preexisting skills [33].

A third crucial aspect concerns the speed at which quality improvements appear in the writing domain. Studies measuring writing improvements typically aim at demonstrating enhanced quality and quantity after several weeks or a semester [34–37]. Regarding the measurement of improved writing skills, Bruton and Kirby [38] emphasize that "per-minute production" should not be equated with writing skills. Instead, the focus should be on idea generation, meta-cognition, and cohesive text creation.

As a substantial improvement in students' academic writing quality on these dimensions does not occur within a few hours of conventional writing instruction, short-term intervention studies lasting mere hours rather than weeks are rare in the literature. However, given ChatGPT's potential to provide numerous, adaptive examples to learners, our study examines whether substantial improvements in students' writing can be achieved through shorter interventions.

2.1.3. Academic perspectives on how ChatGPT could affect academic writing

Numerous researchers have praised ChatGPT for its ability to write college-level essays, explain concepts, and demonstrate almost human-like creativity, thereby performing integral parts of academic writing [11, 39, 40]. Upon ChatGPT's release, initial reactions from scholars were negative, and they assumed that ChatGPT might negatively affect students' academic writing skills [41]. In contrast, some universities recognized ChatGPT's potential for research but initially prohibited its use in coursework [42], whereas other universities allowed its use without regulations. Although many recognized the need to implement strong guidelines, developing and implementing them proved challenging, particularly as ChatGPT rapidly evolved its capabilities through updates [43]. As this technology is making its way into classrooms, understanding its impact on writing development, particularly in academic writing, is crucial. Recent research by Dong [44] investigated L2

students' perceptions and usage of ChatGPT, focusing on their writing proficiency. The study suggests that students with positive expectations about performance and efficiency were more likely to use ChatGPT, although students wanted more structured guidance from teachers on effective strategies for using ChatGPT. The author suggested that negative associations with using ChatGPT in academic writing initially deterred students from experimenting with the tool and prevented teachers to explore its potential in class [44].

2.1.4. Potentials and challenges of ChatGPT in academic writing

Alsaedi [9] found in a systematic review of ChatGPT-related articles that over-reliance on ChatGPT could lead to decreased learning outcomes, whereas improved writing proficiency among students was another central point. Students indicated that ChatGPT improved writing efficiency, but noted issues with originality, wordiness, and logical fallacies when using it [45]. Nevertheless, more empirical research on the opportunities and risks of ChatGPT in academic writing classes is needed.

On the one hand, previous research has shown that potential risks in AI language model responses exist, including hallucinations, biases, ethical questions, and inconsistencies [46, 47]. Moreover, Krzanowski [48] observed that GPT-3 sometimes generated student-quality essays but at other times produced inconsistent results. On the other hand, ChatGPT might have pedagogical benefits, for instance, by aiding students in their writing [49], particularly for non-native English speakers, as seen with text-enhancing tools like Grammarly [50]. Dong [51] investigated how students engaged with ChatGPT responses while reviewing their written texts, highlighting that cognitive engagement and critical evaluation of ChatGPT feedback were crucial for students to benefit from using ChatGPT. Furthermore, ethical engagement with the tool was an important dimension, confirming ideas from previously mentioned sources that emphasized the need to understand ChatGPT's shortcomings.

ChatGPT is likely to play a crucial role in university writing tasks given its widespread adoption among students.

2.1.5. Improvements in writing quality with ChatGPT

Multiple studies have shown quality improvements while using ChatGPT [19], but most have focused on English-language writing abilities [20–22]. Studies specifically examining ChatGPT's impact on academic writing remain rare, though some initial research has indicated potential use cases. Song and Song [52] found that students who used ChatGPT to revise their writing could outperform a control group that did not use the tool. Mahapatra [23] reported that using ChatGPT as a source of feedback on original writing could help students improve their academic writing performance and quality over a semester. Collaborative writing with ChatGPT has been observed in practice but has not yet been extensively studied experimentally in academic writing.

2.2. ChatGPT's potential for students' increasing subjective writing experiences in academic writing

2.2.1. Writing experience

Writing experience in academic writing refers to the cognitive and emotional experiences students encounter while writing, which can be influenced by personal prior experience, writing affinity, perceived skills, motivation, perceived task difficulty, perceived control over the outcome, and technology acceptance [14, 53, 54]. There are various measures to quantify writing experience during an experiment. Measures such as perceived ease and cognitive load can indicate how students perceive task difficulty [55–57].

2.2.2. ChatGPT's potential impact on students' writing experience

Contrary to some researchers' arguments that ChatGPT's efficiency in producing text could potentially demotivate students [58], there is

evidence that ChatGPT might even increase motivation and in-class engagement. For instance, Ali et al. [59] found that ChatGPT increased students' English language writing motivation and ease. Other studies have confirmed ChatGPT's positive effects on students' writing experience [60–62]. Accordingly, Yildiz [17] reported positive effects regarding subjective writing experience among students using ChatGPT compared to a control group without AI support. Besides motivation, increases in self-efficacy during AI use might also be a positive predictor of subjective writing experience [63]. Xu et al. [16] also noted that ChatGPT can increase students' self-efficacy and motivation in certain classes. However, ChatGPT's strong capabilities might backfire when the AI assistant is unavailable, leading to frustration and negatively affecting engagement [18]. Given ChatGPT's capabilities, when AI assistance is no longer available, one might expect a worse subjective writing experience and greater student frustration.

2.2.3. Defining ownership in the context of academic writing

When discussing writing experience with AI, new crucial factors emerge, such as perceived ownership of the text created. Ownership can be described as the feelings of possession, self-creation, self-identity, and control of the creative process [15,64]. A reduced sense of ownership was associated with diminished perceptions of the value of academic writing and intrinsic motivation [14].

2.2.4. ChatGPT's potential impact on students' perceived ownership

Draxler et al. [15] found that users rarely claim full ownership of AI-generated text but do claim co-authorship of AI-assisted work, often without disclosing AI involvement. Therefore, students' sense of ownership may decrease when using ChatGPT, potentially negatively affecting their subjective writing experience. Mikalayeva [65] notes that a heightened sense of ownership can increase students' motivation and class participation. There also seems to be a correlation between the amount of AI content used and, therefore, declining feelings of ownership over texts [66]. Investigating the effects of ChatGPT on students' perceived ownership could reveal whether classroom writing experience is affected by its use.

3. The present study

The present study investigates the effects of ChatGPT on the development of academic writing performance. We aim to understand changes in the quality of student-written texts when using ChatGPT assistance as compared to texts written without AI assistance. Therefore, we are interested in writing situations before students access ChatGPT (condition before ChatGPT), situations while accessing ChatGPT (condition with ChatGPT), and situations after students lose access to ChatGPT (condition after ChatGPT). To stay within the course's time frame, we wanted to address a scenario that covers many aspects of the academic writing process. Therefore, our focus is on writing introductory sections of academic papers, as these sections require high levels of idea generation, metacognition, synthesis of ideas from sources, and the creation of cohesive, argumentative text [67]. We address the research question: *How is the use of ChatGPT associated with the quality of academic texts produced by first-year students in a regular academic writing course (during and after using ChatGPT)?* Regarding this research question, we stated two hypotheses (H):

H1: Students write academic texts, more specifically the introductory sections of a scientific paper, of higher quality when using ChatGPT assistance (compared to writing without ChatGPT). That is, we expected the AI tool to affect performance.

H2: Students write academic texts, more specifically the introductory sections of a scientific paper, of higher quality after having used ChatGPT assistance (compared to writing before ChatGPT). In other words, we expected an AI tool learning effect.

H2 is quite tentative as it has not yet demonstrated that *learning effects* from observing how ChatGPT addresses academic writing tasks

occur. Moreover, one might argue that students might get discouraged when writing on their own again after using ChatGPT for text production. Therefore, this hypothesis assumes that the customized texts generated by ChatGPT for a writing assignment might serve as important worked examples for learners and that their availability would outweigh potential discouragement effects after using the AI assistant.

An additional exploratory issue in our study is collecting process data to observe how students use ChatGPT as an assistant and how they edit AI-generated content. Moreover, we were interested in the degree to which they would experience ownership and how their subjective writing experience changed when writing the introductory sections before, during, and after the Human-AI co-creation with ChatGPT.

4. Method

4.1. Sample

Twenty-five students participated in the study; four were excluded due to incomplete data, leaving 21 participants. Participants were first-year international students enrolled in the fashion design curriculum in Berlin, presenting the authors with an ad hoc sample from a student population that was not too familiar with academic writing to begin with. An ethics committee at our university approved the study. Students used an algorithm to create usernames that only they could retrieve, ensuring participant anonymity while allowing them to request the removal of their data if desired. The students had limited knowledge of ChatGPT and academic writing and rated their experience with ChatGPT at $M = 14.70$ on a scale of 1 to 100 (Table 1). As students had to formulate texts in English in this experiment, their language proficiency was a crucial factor to observe. Students ranked their English proficiency as very good on average ($M = 4.67$, $SD = 0.91$, $range = 1-6$, see Table 1 for the pre-questionnaire results), with some native speakers and no students ranking themselves below a *good* (4) score. Furthermore, students needed to perform well on IELTS, TOEFL, or similar English-language tests to be admitted to this degree program at our university.

4.2. Study design and procedure

To test our hypotheses, we employed a within-subjects experimental design with an ABA structure. The experiment lasted three hours and took place during the sixth session of a regular English-language academic writing course taught by the first author (further contextual information on the academic writing course is provided in Appendix A). Any student enrolled in the class could participate in the study, but

Table 1
Arithmetic Means and Standard Deviations of Variables Used in the Pre-Writing Questionnaire.

Variable	Mean	SD	Range	
Age	21.7	3.2	17–29	
Do you have any reading/writing disability, e.g., dyslexia?	1.9	.436	0–1	yes, no, prefer not to say
How would you rate your English proficiency?	4.67	.913	1–6	poor, fair, average, good, very good, native speaker
How often do you use GPT/ChatGPT or similar AI tools in the context of writing?	14.7	18.4	1–5	never – seldom – sometimes – often – almost always
Ability Beliefs Scale	2.88	.408	1–5	disagree, disagree, neutral, agree, strongly agree
Writing Affinity Scale	3.29	.80	1–5	disagree, disagree, neutral, agree, strongly agree
Perceived Value of Academic Writing	3.12	.604	1–5	disagree, disagree, neutral, agree, strongly agree

participation was voluntary. In the weeks preceding this session, strategies for using ChatGPT in academic writing were covered in class, including conducting AI research, summarizing information, answering questions about a topic [68], assisting with writing, paraphrasing, and proofreading. The instructor emphasized that AI-generated content may be incorrect (e.g., through hallucinations) even when it appears correct, and that it should therefore always be verified. In class, the instructor showed students examples of the strategies discussed in the previous week's lecture to illustrate how ChatGPT can be used for academic writing. Although students may have used personal ChatGPT accounts at home prior to the course, the study constituted their first structured, supervised in-class use of the tool. No additional prompt engineering training was provided because the study aimed to observe students' naturalistic use of ChatGPT based on the general strategies already taught in class, rather than trained expert prompting.

Furthermore, students were taught how to write introductions to scientific papers, explicitly addressing the following elements: a hook to generate interest, context to explain the topic's relevance, and a main thesis statement [69]. Given that the current study took place in a first-semester course, we could assume a comparable level of prior knowledge among participants. This assumption is supported by the results of the pre-questionnaire, in which writing affinity ($M = 3.29$, $SD = 0.80$, $range = 1-5$), beliefs in one's own abilities regarding academic writing ($M = 2.88$, $SD = 0.41$, $range = 1-5$), and the previous experience with ChatGPT ($M = 14.70$, $SD = 18.40$, $range = 1-100$) showed similar overall levels across students, with slight to moderate standard deviations, relatively low values on the given scales, and no extreme outliers. Participants received €5 as compensation.

Participants were required to write three introductions to new, original academic papers on three different topics (within-subjects design). Students wrote the first introduction without ChatGPT (baseline), the second with ChatGPT using a version of the GPT-3.5 API embedded in the experiment interface, and the third introduction again without ChatGPT to test for changes after the AI-assisted phase (ABA design: writing before ChatGPT = baseline, writing with ChatGPT, writing after ChatGPT). Students received an explanation of the experiment's sequence before the experiment began.

The three topics assigned to students were related to their major (i.e., fashion design) and included: *the self-image in virtual avatars (topic A)*, *AI's potential in fashion design (topic B)*, and *VR embodiment and the metaverse in the fashion industry (topic C)*. For each topic, students received

three preselected articles as background scientific literature. Students were instructed to rely exclusively on these sources during writing to avoid a confound between writing a scientific paper and performing a literature review. We permuted the topic order across participants to avoid sequence effects, and students spent 35–50 min writing each introduction (For information about the writing interface, instructions, and examples, see Appendix B).

The study began with a pre-questionnaire about demographic information, students' experience with ChatGPT, and academic writing (see Fig. 2). Subsequently, the three writing tasks were administered, each followed by a short questionnaire about students' experiences. After the three writing tasks and their respective post-task questionnaires, an overarching questionnaire on participants' situational interest and the perceived usability of the AI tool concluded the study.

4.3. Measures

4.3.1. Dependent variables for comparing experimental conditions

Three groups of dependent variables were measured to compare the outcomes of the writing task: quality-related variables, subjective writing experience, and perceived ownership.

4.3.2. Quality-related variables

To test our two hypotheses, we obtained three variables to measure text quality: human rater scores, readability scores, and text cohesion. Academic writing in studies is commonly assessed using both automated and human-rated metrics. Readability and cohesion are important for approximating human-rated writing quality [70–72], but human ratings remain the gold standard in research [73,74]. To achieve the best possible assessment of academic writing quality, we used a mixed-methods approach that included readability metrics, cohesion analysis, and human rater scores to capture both surface-level and deeper textual qualities [70,72].

Human Rater Scores: We created the *total text score* (dependent variable DV1) from the sum of two academic writing teachers' ratings on five dimensions, namely precision of word choice (DV1a), language and grammar (DV1b), scope (DV1c), logic in outline (DV1d), and argumentative cohesion (DV1e). Teachers provided *suggested grades* (DV2) for each introduction using the German school grading system, ranging from 1 (*best*) to 6 (*worst*). To ensure neutrality, all introductions were presented to teachers in a random order, without indicating whether AI

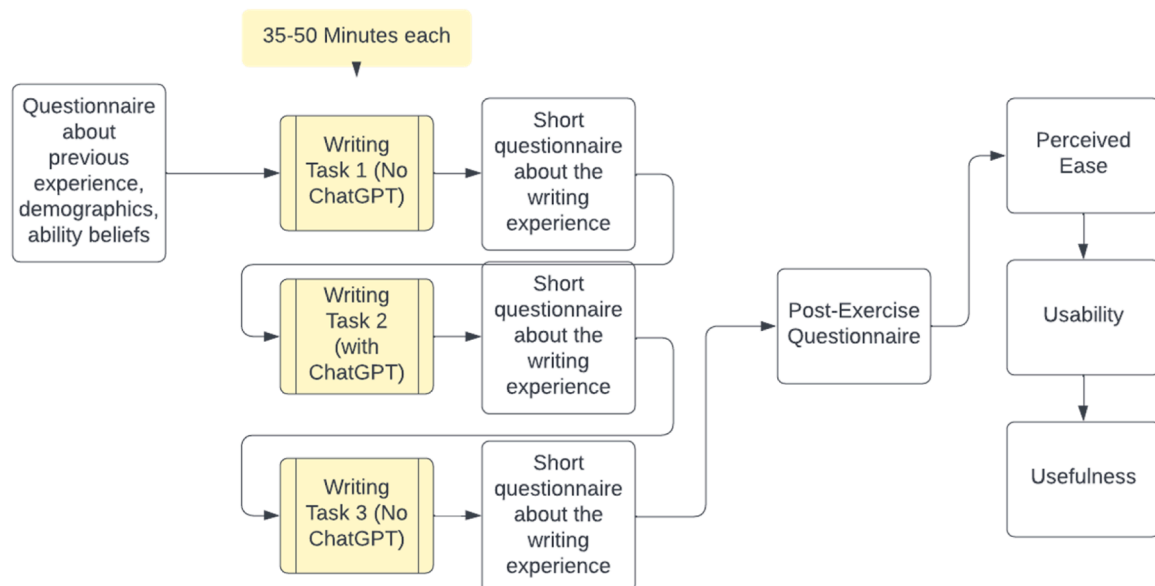


Fig. 2. Study Procedure.

assistance was used.

As AI-generated texts rarely contain spelling or grammar mistakes [75], all texts were corrected using Microsoft Word before being presented to the raters. This ensured a fairer comparison by preventing lower-order text cues from influencing raters' quality assessments based on higher-order text cues (see [76]). This correction procedure simulated typical student practices before submitting papers. From this process, we obtained *spelling mistakes* (DV3) and *grammar mistakes* (DV4) for all texts.

Readability Scores: In addition to human ratings, software tools provided additional text-quality scores. We integrated six readability scores (e.g., Coh-Metrix L2 Readability Index (CML2RI), Flesch-Kincaid Grade Level (FKGL); for a complete overview, see Appendix C) from the Automatic Readability Tool For English (ARTE) toolbox [77], Grammarly's text score and the reading level from Wordcounter.net (ranging from school levels with the lowest here being 7th to 8th grade to college graduates) into one *readability scale* (DV5; Cronbach's $\alpha = .95$). Further data obtained from Wordcounter.net were *total words* (DV6), *unique words* (DV7), *average word length* (DV8), and *total characters* (DV9; [78, 79]).

Cohesion: Cohesion, which is used to predict reader comprehension difficulty and is linked to text quality and expert ratings, was measured using the Tool for the Automatic Analysis of Cohesion (TAACO; [71, 80]). TAACO measures different aspects of local, global, and overall cohesion (in-sentence, within-paragraph, full-text; [71]). Given the short length of most texts, we focused on within-paragraph cohesion, as it might prove more reliable than in-sentence cohesion. Crossley et al. [81] found that this is particularly relevant for international students, rendering this measure relevant to our study, as the participants were mainly international students. We combined several relevant variables (e.g., Overlap of synonyms between adjacent sentences, etc.; for a complete overview, see Appendix C) into one overall *cohesion scale* (DV10; Cronbach's Alpha [α] = .80), along with the *text givenness* (DV11), which is a cohesion measure that indicates how much information can be inferred from previous discourse. We assessed givenness by analyzing the use of pronouns, assuming that the use of pronouns suggests that the information they refer to is already known or given in the context (the measure contains eight items, referring to different types of pronouns; Cronbach's $\alpha = .75$; [71,81]).

4.4.3. Post-writing questionnaire – subjective writing experience and ownership

After each writing exercise, students answered a short post-exercise questionnaire on a 5-point Likert scale. This questionnaire included a scale with three items on the *topic understanding* (DV12; e.g., "It was easy for me to understand the topic.", Cronbach's $\alpha = .83$), and a scale containing three items addressing the *task's perceived ease* (DV13; e.g., "It was easy for me to complete this task.", Cronbach's $\alpha = .88$). Based on the cognitive load theory [57] and the NASA-TLX instrument from Hart & Staveland [55], *cognitive load* (DV14) was measured using five questions regarding mental, physical, and temporal demands, including frustration levels (rated from 1 to 100; e.g., "How much time pressure did you feel [...]?", Cronbach's $\alpha = .86$). Next, one question asked how *successful students felt in their task* (DV15). Moreover, we administered the *perceived control scale* (DV16; [15]; e.g., "The resulting text happened without my intention." [reversed polarity], Cronbach's $\alpha = .67$). Finally, the *ownership scale* (DV17; adapted from Draxler et al. [15] on students' sense of authorship concluded the questionnaires to investigate ChatGPT's impact on feelings of ownership (e.g., "I am the main contributor to the content of this text.", Cronbach's $\alpha = .85$; [82,83]).

4.4. Additional variables independent of experimental conditions

4.4.1. Pre-writing questionnaire

Beyond demographic variables (cf. Table 1), we assessed three motivational pretest variables prior to the writing tasks. We used the

expectancy-value theory as a framework for assessing achievement-related choices and motivational perspectives that might affect students' performance and quality in academic writing or in our experiment. According to Eccles and colleagues [84], expectancy for success based on ability beliefs, previous experience in a subject, overall affinity for a subject, intrinsic motivation, and perceived usefulness are predictors of motivation and achievement-related choices [85–87]. We aim to use this data to quantify whether prior exposure or attitudes towards the subject affect writing quality or experience, and whether attitudes change over the course of the experiment.

All items (unless otherwise noted) were answered on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*; all results in Table 1). The scales measured *ability beliefs* (pre-1; $M = 2.88$; $SD = 0.41$; e.g., "I just do not have talent for academic writing." [reversed polarity]; [88]), *writing affinity* (pre-2; $M = 4.83$, $SD = 1.32$; e.g., "How often do you write for your studies?"; based on [89], on a 6-point scale from 1 = *never* to 6 = *very frequently*), as well as *perceived value of academic writing* (pre-3) according to the Expectancy-Value theory ($M = 3.12$, $SD = 0.60$; e.g., "I find the things we learn in academic writing useful for my life."; [84, 85]).

4.4.2. Variables on the use of ChatGPT

After writing with ChatGPT, additional questions on sense of leadership were assessed (adapted from [15]; e.g., "I was writing, and AI was assisting me."; rated from 1 to 100). These values help clarify students' perceptions of their own contributions and their understanding of AI's limitations. Moreover, we calculated values for directly copied text from ChatGPT by comparing the final text to the full ChatGPT output using a plagiarism checker and labeled it as *AI plagiarism* (AI1). Finally, we calculated the *percentage of the total ChatGPT output that students used in their texts* (AI2) by checking the proportion of the full ChatGPT output they used, again using a plagiarism checker.

4.4.3. Final questionnaire

A final questionnaire comprising five measures of experience with ChatGPT concluded the study. We mainly used the technology acceptance model (TAM) to determine how students would accept a class application of ChatGPT. Information gathered in this questionnaire could, in turn, provide insights into how LLMs could benefit academic writing classes in the future [90]. This questionnaire included a *scale for usability* (post1) using the control scale and the affect scale from the software usability measurement inventory (SUMI) and the system usability scale (SUS; 18 items on a 5-point scale, e.g., "The system's speed system is fast enough.", $\alpha = .83$; [91–93]). The technology acceptance model by Davis [94] was used to measure the *usefulness* (post2; 19 items, $\alpha = .98$), and the *ease of use* (post3; 16 items, e.g., "I found it easy to get ChatGPT to do what I want it to", $\alpha = .86$), both on a 5-point scale [95]. Subsequently, we measured the *situational interest in ChatGPT and academic writing* (post4) with six items ($\alpha = .84$; [96]). Finally, we investigated future *intention to use ChatGPT* (post5) on a 5-point scale (1 = *very unlikely* to 5 = *very likely*).

4.5. Statistical analyses

All data was analyzed using Jamovi [97]. We used one-sided paired-sample *t*-tests to compare questionnaire responses and quality-related variables before and during ChatGPT use, and before and after ChatGPT use. The statistical significance level for this study was set at $p = .050$. All variable values were z-standardized within each topic to account for differences in difficulty across the three topics.

5. Results

First, we present results comparing introduction texts written before ChatGPT and those written with ChatGPT (Hypothesis 1). Second, we provide insight into ChatGPT's potential influence on text written after

its use, compared to texts written beforehand without it (Hypothesis 2). Third, we will describe the additional descriptive data independent of experimental conditions, including usability and perceived usefulness. For a snapshot of the most important results, see Fig. 3.

5.1. Hypothesis 1: does ChatGPT improve the writing quality and experience compared to the baseline?

5.1.1. Writing quality gains with ChatGPT

5.1.1.1. Examiner ratings. Examiners rated total text scores (DV1) statistically significantly higher for introductions with ChatGPT ($d = 1.18$, $p < .001$; see results for H1 in Table 2; cf. Fig. 3) and suggested grades (DV2) also improved compared to before ChatGPT ($d = -0.68$, $p = 0.003$). Statistically significant improvements were observed in precision of word choice (DV1a; $d = 1.11$, $p < .001$), language and grammar (DV1b; $d = 1.49$, $p < .001$), and argumentative cohesion (DV1e; $d = 0.53$, $p = .012$). However, scope (DV1c; $d = 0.26$, $p = .121$) and logic in outline (DV1d; $d = 0.18$, $p = .209$) showed only descriptive differences in the expected direction but yielded no significance.

5.1.1.2. Computed writing quality indicators. Although spelling mistakes (DV3) were reduced by a factor of five when using ChatGPT, this effect was not statistically significant ($d = -0.26$, $p = .125$, see Table 2), whereas grammar mistakes (DV4) recognized by Microsoft Word decreased statistically significantly during ChatGPT use ($d = -1.12$, $p < .001$). With ChatGPT, the readability scale (DV5) improved statistically significantly as compared to before ChatGPT ($d = 1.09$, $p < .001$).

Moreover, students tended to write more total words (DV6; $d = 0.35$, $p = .060$) and significantly more unique words (DV7; indicating higher levels of lexical diversity) with ChatGPT ($d = 0.43$, $p = .031$). The average word length (DV8) also significantly increased from before ChatGPT to writing with ChatGPT ($d = 1.02$, $p < .001$), corresponding to a significant increase in total characters (DV9; $d = 0.50$, $p = .017$). The cohesion scale (DV10) showed statistically significant improvements with ChatGPT ($d = 0.88$, $p < .001$). Similarly, givenness (DV11) also significantly increased with ChatGPT compared to before ($d = 0.97$, $p < .001$).

5.1.2. Changes in writing experience with ChatGPT

Regarding the post-writing questionnaires, topic understanding (DV12) did not change when ChatGPT was used ($d = 0.18$, $p = .211$, see Table 3). Students rated the perceived ease of the ChatGPT-assisted exercise (DV13) significantly higher ($d = 1.04$, $p < .001$; Fig. 3). A statistically significant reduction in cognitive load (DV14) due to ChatGPT use was also observed ($d = -1.59$, $p < .001$; Fig. 3). With ChatGPT, students felt more successful in completing their tasks (DV15) than before ($d = 0.69$, $p = .002$). Perceived control (DV16) was statistically significantly reduced when writing text with ChatGPT as compared to before ChatGPT ($d = -0.56$, $p = .009$). Finally, perceived ownership (DV17) was significantly lower when working with ChatGPT than before ($d = -0.52$, $p = .014$; Fig. 3).

5.2. Hypothesis 2: do students experience lasting gains in writing quality and experience after ChatGPT is no longer available?

5.2.1. Writing quality improvements after ChatGPT use

5.2.1.1. Examiner ratings. Examiner's total text scores (DV1) increased statistically significantly from writing text before ChatGPT to writing text after ChatGPT ($d = 1.12$, $p < .001$; see Table 4 for all values for H2; Fig. 3), as did suggested grades (DV2; $d = -0.86$, $p < .001$). Thus, overall teacher judgements of text quality improved by nearly one standard deviation after less than one hour of academic writing practice assisted by ChatGPT. More specifically, precision of word choice (DV1a; $d = 0.61$, $p = .005$), language and grammar (DV1b; $d = 1.56$, $p < .001$), and argumentative cohesion (DV1e; $d = 1.14$, $p < .001$) improved. Moreover, for this comparison, logic in outline (DV1d) also statistically significantly improved ($d = 0.81$, $p < .001$), whereas scope levels (DV1c) remained non-significant ($d = 0.30$, $p = .094$).

5.2.1.2. Computed writing quality indicators. We found no significant differences in spelling mistakes (DV3; $d = -0.17$, $p = .217$; see Table 4). However, students made statistically significantly fewer grammar mistakes after ChatGPT than before (DV4; $d = -0.73$, $p = .002$). Readability (DV5) statistically significantly increased after ChatGPT as compared to before ChatGPT ($d = 0.49$, $p = .018$). Moreover, after ChatGPT, total words (DV6; $d = 0.32$, $p = .077$) and unique words (DV7; $d = 0.30$, $p = .092$) increased marginally as compared to before ChatGPT, whereas average word length (DV8) increased statistically significantly ($d = 0.38$, $p = .048$). Total characters (DV9) were also marginally higher ($d = 0.36$, $p = .055$). The cohesion scale (DV10) shows statistically significant improvements from before ChatGPT to after ($d = 0.49$, $p = .019$), just like givenness (DV11; $d = 0.64$, $p = .004$).

5.2.2. Changes in writing experience after ChatGPT use

Results from the post-writing questionnaires show no significant differences regarding topic understanding (DV12) from before ChatGPT to after ($d = 0.08$, $p = 0.365$). The perceived ease (DV13; Fig. 3, see Table 5) increased significantly after ChatGPT as compared to before ChatGPT ($d = 0.60$, $p = .006$). Cognitive load (DV14) significantly decreased after ChatGPT ($d = -0.42$, $p = .035$; Fig. 3). Moreover, students felt significantly more successful in completing their task (DV15)

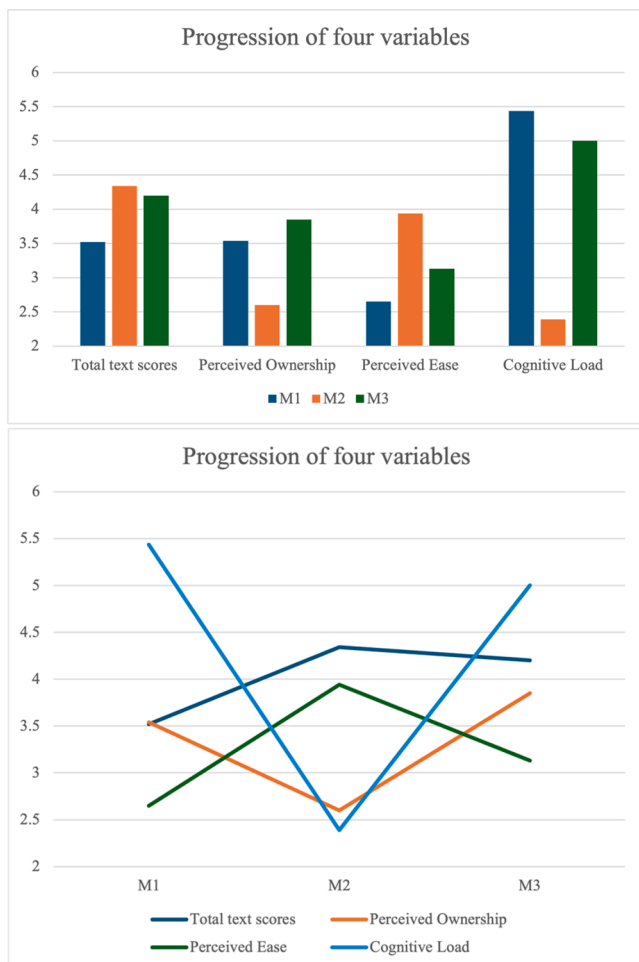


Fig. 3. Main results.

Table 2
Statistics for all quality indicators regarding Hypothesis 1.

Scale	<i>p</i>	<i>d</i>	<i>t</i>	<i>z(m1)</i>	<i>z(m2)</i>	<i>SD m1</i>	<i>SD m2</i>	<i>df</i>	<i>M1</i>	<i>M2</i>	<i>H y/n</i>	<i>Range</i>	Explanation
Total text scores	<.001	1.18	5.424	−0.621	.385	1.11	0.589	20	17.6	21.7	Y	0–25	Teacher Rating Total
Formulation, Choice of words, Precision	<.001	1.106	5.069	−0.609	0.631	1.14	0.520	20	3.53	4.81	Y	0–5	Teacher Rating
Language and Grammar	<.001	1.492	6.837	−0.848	.674	0.987	0.402	20	3.44	4.86	Y	0–5	Teacher Rating
Scope	.121	.263	1.205	−0.189	.103	0.978	0.881	20	3.44	3.90	N	0–5	Teacher Rating
Logic in Outline	.209	0.180	0.825	−0.309	−0.146	1.13	0.810	20	3.63	3.82	N	0–5	Teacher Rating
argumentative cohesion	.012	0.532	2.440	−0.498	0.123	1.12	0.884	20	3.51	4.26	Y	0–5	Teacher Rating
Suggested Grade	.003	−0.682	−3.125	.408	−0.152	1.09	0.786	20	2.83	2.06	Y	1–6	Teacher Rating
Spelling amount	.125	−0.258	−1.18	0.148	−0.183	0.993	0.967	20	3.86	0.762	N	0–38	Microsoft Word Grammar Checker
Grammar mistakes	<.001	−1.119	−5.13	0.542	−0.616	0.961	0.637	20	4.90	1.10	Y	0–10	Microsoft Word Grammar Checker
Readability	<.001	1.092	5.00	−3.58	4.91	6.00	6.00	20			Y	−15, 15	Combined 6 Variables by ARTE Toolbox
Total Word Count	.060	0.354	1.62	−0.246	0.221	0.976	1.03	20	228	303	N	50–683	
Unique Words	.031	0.432	1.98	−0.246	0.257	1.03	0.992	20	131	165	Y	38–330	
Word Length	<.001	1.019	4.67	−0.488	0.739	0.878	0.873	20	4.90	5.65	Y	3.3–6.3	
Characters	.017	0.496	2.27	−0.308	0.330	0.930	1.05	20	1361	2042	Y	280–4664	
Cohesion scale	<.001	0.880	4.03	−0.295	0.316	0.501	0.419	20			Y	−1.6–1.04	Combined of relevant values of TAACO Toolbox
Givenness	<.001	0.097	4.432	−2.68	2.37	3.58	2.83	20			Y	−10–7	Combined of relevant values of the TAACO Toolbox

Note: 1 refers to the writing exercise before ChatGPT, 2 refers to the writing exercise with ChatGPT; *z(m)* refers to the mean values of the *z*-standardized data. Where applicable, the mean values of the raw data were presented for more straightforward interpretation. *H* refers to whether the hypothesis is confirmed or not.

Table 3
Statistics for all writing experience indicators regarding Hypothesis 1.

Scale	<i>p</i>	<i>d</i>	<i>t</i>	<i>z(m1)</i>	<i>z(m2)</i>	<i>SD m1</i>	<i>SD m2</i>	<i>df</i>	<i>M1</i>	<i>M2</i>	<i>H y/n</i>	<i>Range</i>	Explanation
Understanding of the topic	.211	0.179	0.819	−0.091	0.090	0.883	0.720	20	3.68	3.89	N	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Perceived Ease	<0.001	1.042	4.776	−0.476	0.502	0.796	0.549	20	2.65	3.94	Y	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Cognitive Load	<0.001	−1.587	−7.271	2.62	−3.93	3.72	2.81	20	59.8	26.3	Y	1–100	Slider feature from 0 worst to 100 best
Successful in the task - How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	.002	0.694	3.179	−0.385	0.420	1.02	0.862	20	40.4	64.7	Y	1–100	Slider feature from 0 worst to 100 best
Perceived Control	.009	−0.561	−2.572	0.865	−1.78	2.93	3.11	20	3.71	2.96	Y	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Perceived Ownership	.014	−0.515	−2.358	1.02	−3.79	5.45	6.72	20	3.54	2.6	Y	0–5	strongly disagree, disagree, neutral, agree, strongly agree

Note: 1 refers to the writing exercise before ChatGPT, 2 refers to the writing exercise with ChatGPT; *z(m)* refers to the mean values of the *z*-standardized data. Where applicable, the mean values of the raw data were presented for more straightforward interpretation. *H* refers to whether the hypothesis is confirmed or not.

after using ChatGPT than before ($d = 0.50$, $p = .017$). No significant difference was found for perceived control (DV14) when comparing writing before and after ChatGPT ($d = 0.02$, $p = .473$).

Finally, we found marginal gains in perceived ownership (DV17) from before to after ChatGPT ($d = 0.30$, $p = .095$; Fig. 3). However, some individual ownership items suggest that students may feel greater ownership after the ChatGPT exercise than before. For instance, the item

"I made substantial contributions to the text" showed a statistically significant difference before and after ChatGPT ($d = 0.64$, $p = .004$).

5.3. Pre-writing questionnaire

Three pre-writing scales measured *ability beliefs* (pre-1; $M = 2.88$; $SD = 0.41$; e.g., "I just do not have talent for academic writing." [reversed

Table 4
Statistics for all quality indicators regarding Hypothesis 2.

Scale	<i>p</i>	<i>d</i>	<i>t</i>	<i>z(m1)</i>	<i>z(m2)</i>	<i>SD m1</i>	<i>SD m2</i>	<i>Df</i>	<i>M1</i>	<i>M2</i>	<i>H y/n</i>	<i>Range</i>	<i>Explanation</i>
Total text scores	<.001	1.124	5.149	−0.621	.236	1.11	0.896	20	17.6	21.00	Y	0–25	Teacher Rating Total
Formulation, Choice of words, Precision Language	.005	0.613	2.809	−0.609	−0.022	1.14	0.791	20	3.53	4.08	Y	0–5	Teacher Rating
Scope	<.001	1.556	7.132	−0.848	.174	0.987	0.775	20	3.44	4.33	Y	0–5	Teacher Rating
Logic in Outline	.094	0.297	1.360	−0.189	.09	0.978	1.10	20	3.44	3.87	N	0–5	Teacher Rating
Argumentative cohesion	<.001	0.809	3.709	−0.309	.455	1.13	0.848	20	3.63	4.30	Y	0–5	Teacher Rating
Suggested Grade	<.001	1.136	5.207	−0.498	.375	1.12	0.734	20	3.51	4.42	Y	0–5	Teacher Rating
Spelling amount	<.001	−0.862	−3.95	.408	−0.256	1.09	0.960	20	2.83	2.00	Y	1–6	Teacher Rating
Grammar mistakes	.217	−0.174	−0.797	0.148	0.035	0.993	1.01	20	3.86	3.62	N	0–38	Microsoft Word Grammar Checker
Readability	.002	−0.732	−3.353	0.542	0.0737	0.961	0.979	20	4.48	3.38	Y	0–10	Microsoft Word Grammar Checker
Total Word Count	.018	0.493	2.258	−3.58	−1.33	6.00	6.38	20			Y	−15, 15	Combined 6 Variables by ARTE Toolbox
Unique Words	.077	0.323	1.48	−0.244	0.022	0.976	0.933	20	228	248	N	50–683	
Word Length	.092	0.301	1.38	−0.246	−0.012	1.03	0.903	20	131	141	N	38–330	
Characters	.048	0.382	1.751	−0.488	−0.251	0.878	0.760	20	4.90	4.99	Y	3.3–6.3	
Cohesion scale	.055	0.364	1.668	−0.308	−0.0219	0.930	0.907	20	1361	1499	N	280–4664	
Givenness	.019	0.486	2.23	−0.295	−0.0203	0.501	0.453	20			Y	−1.6–1.04	Combined of relevant values of TAACO Toolbox
	.004	0.639	2.93	−2.68	0.312	3.58	3.51	20			Y	−10–7	Combined of relevant values of TAACO Toolbox

Note: 1 refers to the writing exercise before ChatGPT, 2 refers to the writing exercise with ChatGPT; *z(m)* refers to the mean values of the *z*-standardized data. Where applicable, the mean values of the raw data were presented for easier interpretation.

Table 5
Statistics for all writing experience indicators regarding Hypothesis 2.

Scale	<i>p</i>	<i>d</i>	<i>t</i>	<i>z(m1)</i>	<i>z(m2)</i>	<i>SD m1</i>	<i>SD m2</i>	<i>df</i>	<i>M1</i>	<i>M2</i>	<i>H y/n</i>	<i>Range</i>	<i>Explanation</i>
Understanding of the topic	.365	0.077	0.350	−0.091	0	0.883	0.944	20	3.68	3.76	N	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Perceived Ease	.006	0.596	2.73	−0.576	−0.026	0.796	0.721	20	2.65	3.13	Y	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Cognitive Load	.035	−0.416	−1.91	2.62	1.32	3.72	3.65	20	59.8	55	Y	1–100	Slider feature from 0 worst to 100 best
Successful in the task - How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	.017	0.498	2.28	−0.385	−0.034	1.02	0.932	20	40.4	50.8	Y	1–100	Slider feature from 0 worst to 100 best
Perceived Control	.473	0.015	0.069	0.865	0.918	2.93	3.13	20	3.71	3.18	N	0–5	strongly disagree, disagree, neutral, agree, strongly agree
Perceived Ownership	.095	0.297	1.36	1.02	2.78	5.45	5.21	20	3.54	3.85	N	0–5	strongly disagree, disagree, neutral, agree, strongly agree

Note: 1 refers to the writing exercise before ChatGPT, 2 refers to the writing exercise with ChatGPT; *z(m)* refers to the mean values of the *z*-standardized data. Where applicable, mean values of the raw data were presented for easier interpretation.

polarity]; [88], see Table 1), *writing affinity* (pre-2; $M = 4.83$, $SD = 1.32$; e.g., "How often do you write for your studies?"; based on [89], on a 6-point scale from 1 = *never* to 6 = *very frequently*), as well as *perceived value of academic writing* (pre-3) according to the Expectancy-Value theory ($M = 3.12$, $SD = 0.60$; e.g., "I find the things we learn in academic writing useful for my life."; [84,85]).

5.4. Variables on the use of ChatGPT

5.4.1. Students' use of ChatGPT

Questions on sense of leadership showed that students rated "AI was writing, I was assisting" at $M = 3.43$ ($SD = 1.47$; see Table 6), while "I was writing, AI was assisting" was rated at $M = 2.76$ ($SD = 1.55$). Students asked ChatGPT an average of 4.76 questions per writing task ($SD = 2.21$). Most prompts (28) were intended to prompt ChatGPT to generate novel text without using the source PDFs provided to the students. Of these, around half aimed to understand or gather information about the

Table 6
Arithmetic Means and Standard Deviations of Variables regarding AI use.

Variable	N	Mean	SD	Range
Who took the lead?	21	3.33	1.39	1–5
I was writing, AI was assisting	21	2.76	1.55	1–5
AI was writing, I was assisting	21	3.43	1.47	1–5
AI Plagiarism	21	0.711	0.305	0–1
Percentage of AI content used	21	0.399	0.270	0–1

topic the students were required to write about. 19 prompts summarized text from the sources, 12 paraphrased source texts, and 10 generated text based on the sources. Students also used ChatGPT to improve their own writing (13 prompts), while one asked for a synonym of a word. Finally, 12 prompts were merely conversational or attempted to clarify the previous question when ChatGPT misunderstood the task. Qualitative examples of students' use of ChatGPT (improving self-written text, generating text without source input, and asking ChatGPT questions about the topic) are presented in [Appendix D](#).

5.4.2. Plagiarism from ChatGPT's output

71% of their text was directly copied from ChatGPT, i.e., AI plagiarism (AI1), with many students relying heavily on ChatGPT, hence a 95% median (see [Appendix E](#) for distribution). AI plagiarism was measured by comparing the student's final text to their full ChatGPT output using a plagiarism check. However, the *percentage of content from the total ChatGPT output that students used in their own texts* (AI2) was only 30.5%, indicating that students revised and selected their texts to some degree through prompting before copying the AI text into their own introductions. We calculated AI2 by dividing the number of plagiarised words from the ChatGPT output by the total number of words generated by ChatGPT. This value indicates how much content students generated before they were satisfied that ChatGPT's output was sufficient for creating their final introduction. This result also suggests that students often prompted ChatGPT multiple times and only used selected parts of the AI-generated text, indicating some level of editing, filtering, or iterative refinement before incorporating it into their final versions.

5.4.3. Correlations between ChatGPT use patterns and quality indicators

There was a statistically significant negative correlation between the number of chat queries and the amount of AI text used (AI2; $r = -0.376$, $p = .046$; see [Table 7](#)), indicating that students who generated more text needed more ChatGPT content until they felt their needs for the task were met. This result suggests that students may need to act more as editors than authors, underscoring the importance of critically re-reading ChatGPT-generated texts. Readability of texts (DV5) with ChatGPT had a positive correlation with AI plagiarism (AI1; $r = .607$, $p = .002$), as did readability of texts (DV5) written after using ChatGPT ($r = .427$, $p = .027$), but not readability of texts (DV5) written before using ChatGPT ($r = .145$, $p = .734$). Perceived ownership (DV17) of ChatGPT-texts had a significant negative correlation with AI plagiarism (AI1; $r = -0.579$, $p = .003$). We found no statistically significant correlations between either AI plagiarism (AI1) and perceived ease (DV13), perceived control (DV16), or cognitive load (DV14).

5.5. Final questionnaires

5.5.1. Usability, usefulness, and ease of use

On a 5-point scale, students rated ChatGPT's usability (post1) as $M = 3.85$ ($SD = 0.54$, see [Table 8](#)), usefulness (post2) as $M = 3.67$ ($SD = 1.05$), and ease of use (post3) as $M = 3.80$ ($SD = 0.60$; see [Table 8](#)). Therefore, the user experience seems to have been quite positive.

Table 7
Pearson's Matrix - Comparisons of Pre- and Post-Questionnaire, Plagiarism, and Item Interactions.

Variable	AI Plagiarism		Percentage of AI content used		Direction
	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	
Chat how many queues	-	-	.046	−0.376	3
01 Readability Scale	.266	.145	.323	.107	2
02 Readability Scale	.002	.607	.055	.358	2
03 Readability Scale	.027	.427	.072	.330	2
01 Ownership Scale	.603	.061	.326	−0.104	3
02 Ownership Scale	.003	−0.579	.072	−0.331	3
03 Ownership Scale	.429	−0.041	.358	−0.084	3
02 Total Words	.042	.386	-	-	2
02 Characters	.031	.414	-	-	2
02 Average Word Length	.018	.458	-	-	2
02 Grammar Mistakes	.003	−0.584	-	-	3
Who took the lead?	.008	.519	-	-	2
AI was writing. I was assisting	.006	.541	-	-	2
Usability Scale	.131	.256	-	-	2
Usefulness Scale	.030	.416	-	-	2
Ease Scale	.128	.259	-	-	2
Situational Interest	.166	.222	-	-	2

Note. We found no statistically significant correlations between the AI-usage variables and the control, topic understanding, ease, and cognitive load scales. Directions 2 and 3 indicate a positive and negative correlation, respectively.

Table 8
Arithmetic Means and Standard Deviations of Variables Used in the Post-intervention Questionnaire.

Variable	N	Mean	SD	Range
Usability Scale	21	3.85	.536	1–5
Usefulness Scale	21	3.67	1.05	1–5
Ease of Use Scale	21	3.80	.581	1–5
Situational Interest Scale	21	3.69	.857	1–5
Intention to Use ChatGPT in the future	21	3.57	1.40	1–5

5.5.2. Situational interest in ChatGPT, academic writing, and future use of AI

Participants reported a substantial situational interest in ChatGPT and academic writing on a 5-point scale (post4; $M = 3.85$; $SD = 0.86$; see [Table 8](#)). Significant correlations emerged between pre- and post-experiment and pre-exercise measurements of situational interest in academic writing, for example between "Academic writing is interesting to me" (Situational interest scale, post-exercise; see [Fig. 4](#)) and the pre-experiment items "I like writing academic texts" ($r = .645$, $p < .001$), "I enjoy writing academic texts" ($r = .445$, $p = 0.022$), and "Academic writing is not for me" ($r = -0.569$, $p = .004$). As seen in [Fig. 5](#), there was a statistically significant correlation between lower self-rated talent in academic writing before the experiment and belief that ChatGPT could improve one's grades ($r = 0.57$, $p = .003$).

On a 5-point scale, students rated their intention to use ChatGPT for academic writing again in the future (post5) as $M = 3.57$ ($SD = 1.40$). Therefore, the interest in AI-assisted writing following the experiences in the study appears quite substantial.

6. Discussion

The present experimental study demonstrates statistically significant

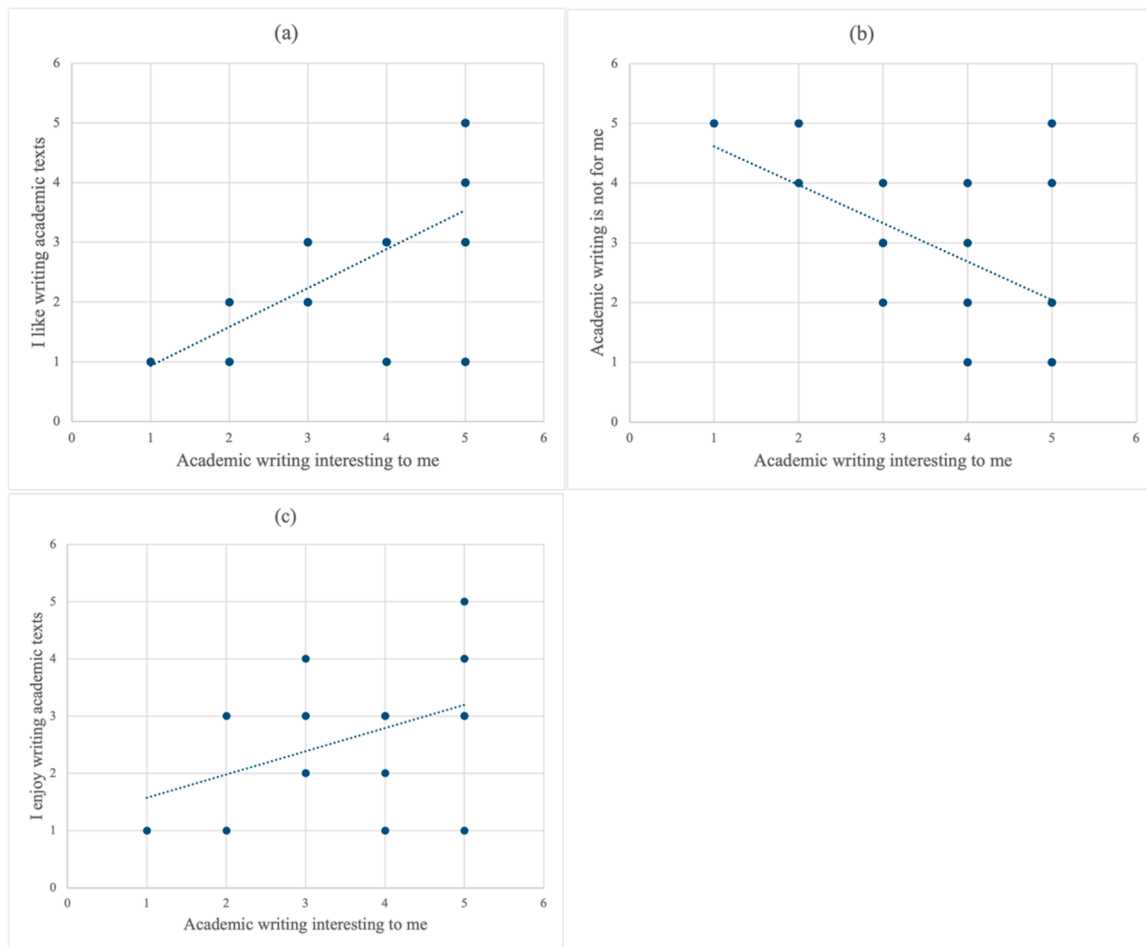


Fig. 4. Correlations between 'Academic writing is interesting to me' and other selected variables.

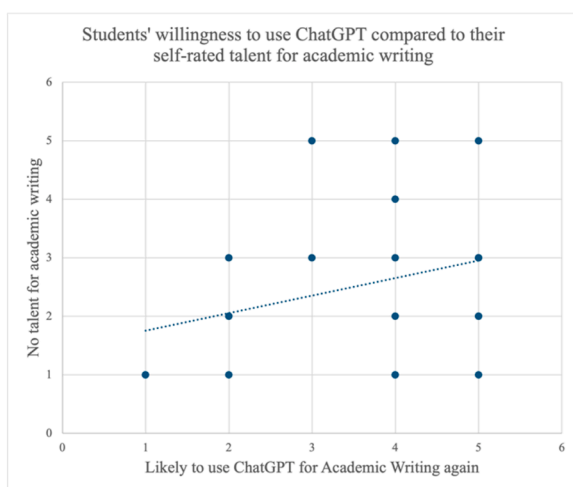


Fig. 5. Correlations between 'no talent in academic writing' and 'intention to use ChatGPT in the future'.

improvements in first-year international fashion design students' academic writing quality when using ChatGPT as an assistant and, even more importantly, after using ChatGPT, as evidenced by comparisons of resulting text quality with student texts written before using an AI assistant. Fig. 3 provides an overview of the study's most significant findings. Teachers' ratings (total ratings (DV1) and suggested grades (DV2)), as well as computed quality indicators (readability (DV5),

cohesion (DV10), givenness (DV11)) improved statistically significantly in texts written with and after ChatGPT compared with those written before ChatGPT. Suggested grades improved (DV2) by more than one German school grade with ChatGPT and by more than half a German school grade after ChatGPT (see Tables 2 and 4).

Similarly, average word length (DV8), associated with higher writing quality, increased significantly during and after ChatGPT compared with before ChatGPT [78,79]. Thus, H1 is confirmed: in our study, first-year international fashion design students write better introductions to scientific papers when using ChatGPT as an assistant than before. Even more important, our crucial second hypothesis, H2, was clearly confirmed: Our sample of first-year international fashion design students significantly improved on all the aforementioned measures not only when using ChatGPT, but also after using it, compared to before. Regarding our research question—*how using ChatGPT influences students' academic texts (during and after use)*—we have a clear answer: ChatGPT has a positive impact on our sample of first-year international fashion design students' academic writing quality during and after its use.

Furthermore, first-year international fashion design students' subjective writing experience was perceived as easier with ChatGPT, and this perception improved further after using it (as evidenced by statistically significant changes in perceived ease (DV13), cognitive load (DV14), and perceived success (DV15)). Their sense of ownership over their texts was marginally higher after writing with ChatGPT than before. Still, their sense of ownership (DV17) and perceived control (DV16) were, as expected, significantly lower when they wrote with ChatGPT than before. Fig. 3 clearly shows that students' ease increases with ChatGPT and remains higher afterward, whereas the opposite is evident for Cognitive Load. It also illustrates students' hesitation to claim

ownership of the ChatGPT texts, followed by a subsequent rise in ownership.

6.1. ChatGPT's potential to improve writing quality

Our study has shown that ChatGPT helped students improve their academic writing quality. Although researchers initially assumed that relying on ChatGPT would have adverse effects on students' learning [9, 41], our findings suggest that ChatGPT could become a valuable learning companion for writing. These results are in line with other relevant studies, which suggested that students might learn from using ChatGPT [98–100] and, more specifically, that ChatGPT might improve students' writing skills and performance sustainably [6, 101, 102]. We could provide preliminary experimental evidence to substantiate this claim. Other current studies provide similar results, indicating that ChatGPT could have a positive impact on academic writing, although the methods of application and frequency of use varied widely across studies [21–23, 52]. Further implications of these results are discussed in the following chapters.

6.2. Students' subjective experiences during and after ChatGPT use

Our experimental findings on first-year fashion design students' writing experiences are in line with other (some quasi-experimental) experimental studies showing that ChatGPT might reduce their anxiety while learning [103], increase their learning motivation [59], increase perceived ease [102], and positively affect their perceived success [104]. Recent studies have shown that ChatGPT may improve students' overall writing experience [16, 17]. The effects of ChatGPT withdrawal require further investigation, but research shows mixed results regarding adverse effects on motivation and perceived ease following access withdrawal [18]. In our study, improvements in ease of use during ChatGPT were sustained in the post-ChatGPT exercise, and students did not report increased frustration afterwards.

6.3. ChatGPT and ownership

The concern that students might rely heavily on ChatGPT when writing and may even claim or perceive AI-generated output as their own work has been raised frequently [105]. However, recent studies (e. g., [15, 106]) have found that people are generally hesitant to claim full ownership of AI-generated content. Similarly, students in our study did not consider ChatGPT-generated texts to be their own work. Their sense of ownership (DV17) was statistically significantly lower with ChatGPT than when writing a text before using the tool. However, their sense of ownership was marginally higher after writing with ChatGPT than before.

This descriptive finding – that students identified their work as their own more strongly after using ChatGPT than before – did not reach statistical significance. However, there was auxiliary anecdotal evidence from students' in-class comments, indicating that academic writing felt rather impersonal to them, as they saw their role as merely compiling other authors' ideas. Therefore, if students were initially largely unaware of their own contributions when writing an introductory section for a new paper, this may have changed after using ChatGPT, given the tool's demonstration of transforming three papers into an introductory section for a new paper. In line with this reasoning, other research suggests that AI may increase people's sense of identity when working without it [107]. A stronger connection to their work might affect intrinsic motivation, learning motivation, and the perceived value of academic writing [14]. Research has shown that the more content students copied from AI, the lower their ownership was [66]. Therefore, writing instructions should be crafted so students can learn from ChatGPT while still adding their own ideas and effort, thereby

increasing their sense of ownership. This could mean that students do not write with ChatGPT, but it asks questions about their text to hint at potential areas of improvement [14].

However, it is also possible that students' increasing reliance on and familiarity with ChatGPT and other AI tools could lead to significant changes in writing behavior, potentially altering users' sense of ownership when generating content with AI [108, 109].

6.4. Improvements in grammar versus logic

ChatGPT excels at producing grammatically correct text with precise word choice [110]. In our study, spelling errors (DV3) and grammatical mistakes (DV4) decreased during ChatGPT use. However, students still need to pay attention to the logic of the text (DV1d). Teachers rating the texts noted that, whereas ChatGPT-assisted texts improved in many technical details, they sometimes included repetitive elements or lacked essential introductory components, both of which impaired the logical structure. Thus, although ChatGPT introductions were well worded, students still need to ensure that ChatGPT-authored content does not encounter logical problems in its intended context. This caveat regarding logical fallacies aligns with previous research [111], particularly concerning the repetition of sentences, words, or constructs [112, 113]. Therefore, teachers must provide students with practical strategies for revising ChatGPT-assisted texts. Sok and Heng [114] also note that for students to use ChatGPT effectively, they need the skills to critically evaluate its output. Future research could focus more on the revision strategies of students who achieved the best results with ChatGPT, particularly in condensing ideas and avoiding logical mistakes, as shown in our experiment.

6.5. Homogenization of texts during and after ChatGPT use

Our findings indicate that ChatGPT might lead to more homogeneous student-written texts. Before ChatGPT, we found more outliers in total words (DV6) and unique word counts (DV7), whereas after ChatGPT, the texts became more similar to each other. Students wrote more words both with and after ChatGPT than before. The standard deviation of total words (DV6) decreased, and the graph (see Fig. 6) shows an almost normal distribution after ChatGPT. We observed similar homogenization effects in readability scores (DV5), with improved readability levels, reduced standard deviation, and fewer outliers after ChatGPT use. Outliers in both directions shifted towards the middle of the distribution. The readability (DV5) graphs in Fig. 7 show outliers before ChatGPT, whereas after ChatGPT, the values are mainly concentrated around the median.

Although in our sample, students' overall quality improved with and after ChatGPT, these homogenization effects might imply that particularly the weaker students benefit most from the AI tool, whereas it could even hinder some of the high performers (i.e., the positive outliers before ChatGPT use). Zhu and Zou [115] reported similar results for creativity outcomes: creativity scores declined for stronger students but not for weaker ones. Custom-trained GPT models might help cater to different starting levels, offering personalized interaction. Here, weaker students could rely on more guidance by ChatGPT, using ChatGPT in multiple areas of writing, such as getting custom examples or templates to suit their specific task, whereas higher performers should be encouraged to explore their ideas on their own initially and use ChatGPT later to add to their existing ideas or enhance their written text. Still, the primary goal of using ChatGPT in class should be to empower all students to develop their writing skills, with and without AI assistance, while ensuring that they develop critical thinking skills and that their understanding of the topic is not outsourced entirely.

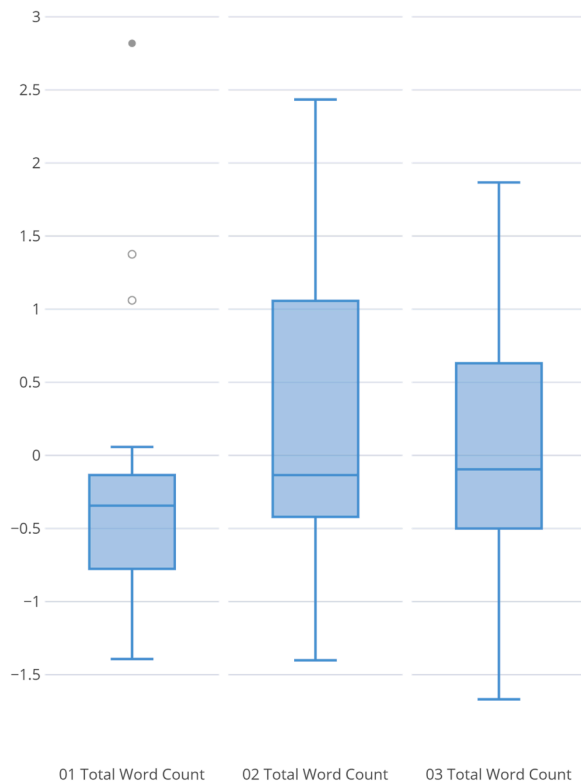


Fig. 6. Distribution of the variable 'total words'.

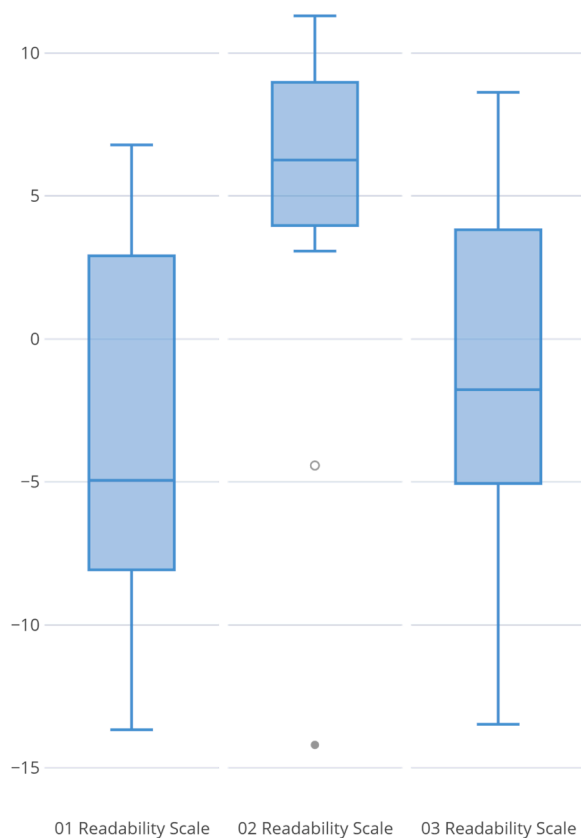


Fig. 7. Distribution of the readability scale.

6.6. ChatGPT's application in class and students' acceptance

In our study, we found that ChatGPT had a positive impact on first-year fashion design students' academic writing quality and writing experience, both during and after access. These findings suggest the potential to integrate ChatGPT (or specialized applications) into classrooms to support different aspects of the writing process. According to the TAM, Students rated ChatGPT to be both useful (post2) and easy to learn (post3). ChatGPT could quickly enhance students' subjective writing experience and increase interest in academic writing. Therefore, we recommend that students use ChatGPT to develop their understanding of academic writing and to improve their ability to express ideas. Dedicated prompt-training strategies may help increase productive and pedagogically useful outcomes for students [51].

6.7. Strengths and limitations

A significant strength of our study is that it demonstrates both quality gains with AI and skill transfer after AI use, indicating that learning effects from ChatGPT are possible. The ABA design, which compares writing processes before, during, and after ChatGPT use in an ecologically valid context, enabled us to focus on situations where students no longer have access to ChatGPT. This allowed us to determine whether quality gains persist beyond AI use and how students' writing experience is affected when they must write autonomously again. By establishing baselines for academic writing quality and perceived writing experience, we could directly compare the effects of the ChatGPT intervention on students' writing quality. Our study is among the first to directly examine the potential impact of ChatGPT on academic writing quality.

Another strength of our study is the triangulation of dependent variables used to measure academic writing quality and experience. By focusing on objective computational analysis, expert human rater scores, and students' self-assessed writing experience, we were able to depict nuanced changes in writing quality and experience more holistically. Employing multiple measures improved the validity of our results.

Furthermore, the marginal gains observed in ownership in the post-ChatGPT exercise relative to the baseline represent an important contribution, as they indicate possible positive psychological impacts on students' sense of authorship in academic writing. This finding challenges narratives that equate the use of AI in writing with a wrongful sense of authorship and reduced student engagement in class. Although preliminary, this finding on ownership – particularly that students did not claim ownership over AI-assisted texts – highlights the complex nature of creating with ChatGPT.

As students participated in the study during their regular academic writing class, we created a realistic picture of the student population, unlike studies that enrolled only particularly interested students.

However, some issues might limit the generalizability of our findings. First, the sample size of 21 students limits the generalizability of the study, although the high effect sizes shown in the results point in the right direction. Using G*Power [116], we confirmed that the sample sizes for paired-samples t-tests were large enough; however, a larger sample would increase the generalizability of our findings. Another issue with the sample is that it is highly specific to a single group of students. Only fashion design students participated in the experiment, who might have different skills and intrinsic motivation regarding academic writing than students of other degrees.

Second, students interacted with ChatGPT for only 50 min, potentially yielding novelty effects and minimizing nonsensical responses that might occur in prolonged interactions. Further studies should explore how perceptions might change as novelty wears off and students encounter some useless responses, as well as whether the learning effects observed in this session persist in the weeks after or are fleeting. The performance effects observed in our session do not prove lasting learning. Longer, semester-based interventions using a similar design

would provide deeper insights into sustained use and any sustained learning effects on writing development and improvement.

Third, since our data collection in May 2023, ChatGPT experience levels in general might likely have changed with ChatGPT's growing maturity and popularity, so that students might show different responses today than in our findings.

Fourth, the improvement observed after using ChatGPT might be due to natural learning effects in academic writing rather than to prior AI assistance. Although previous studies showed learning effects only after much more extended periods (i.e., several weeks rather than minutes; cf. the literature review), the possibility of a normal learning effect in this ABA setup remains and warrants further investigation. Therefore, follow-up studies should introduce control groups to determine whether students achieve comparable learning gains when writing three comparable assignments without an AI tool assisting them in the second assignment. Without this control group, the positive results observed in H2 are preliminary.

This study focused specifically on introductory sections of scientific papers. To increase the generalizability of our findings, future studies should conduct similar assessments across different sections of academic papers. This could mean focusing on using AI to write conclusions, abstracts, or other parts of academic papers, or to gather the best possible results, a full academic paper. However, we intentionally focused on introductions in this study, as introductory sections require high levels of proficiency in skills relevant to other parts of academic papers (i.e., idea generation, metacognition, and the creation of cohesive and argumentative texts).

Another important limitation is that our study involved multiple statistical tests, increasing the risk of Type I error. We did not apply a multiple-comparison correction (e.g., Bonferroni or FDR), so individual p-values should be interpreted with caution. The findings should be considered preliminary until replicated with a design that reduces or corrects for multiple comparisons.

Due to the fact that students' introductions were very short, we had to exclude some of TAACO's most relevant cohesion indices. Therefore, all results regarding cohesion (excluding givenness) should be considered preliminary and verified by studies using longer text samples.

Finally, as students' introductions were pre-corrected to remove grammar and spelling mistakes before being sent to teachers, potential shortcomings in these areas that might affect grading were not considered. We decided to pre-correct, as this is standard practice before submitting academic papers at university, and would make ChatGPT-generated text, which is usually free of grammatical errors, too obvious to pick out. Regardless of motivation, this still poses a significant limitation.

6.8. Implications and further directions

We theorized in the introduction to this paper that ChatGPT's main advantage might be its adaptability to individual needs. Rather than providing learners with generic text examples, ChatGPT can generate examples closely related to their current research questions, thereby increasing their interest and comprehension. We assumed that students benefited mainly from seeing how their task is performed *for* them by the AI with minimal own involvement. This experience differs significantly from using templates or best-practice examples, as it is based on the generation of task-tailored texts as learning materials (in the sense of worked examples).

Future studies are required to substantiate this assumption in greater detail, for instance, by comparing this method of providing task-tailored example texts generated by ChatGPT with providing more general, high-quality example texts that are, however, not specifically aligned to students' current writing task. As mentioned in the introduction, there are differences in students' learning styles (planners vs. revisers) and in how weaker students learn compared to stronger students [31–33]. Due to the timeframe of our study, students could barely outline their work,

so only the *reviser's* writing style was realistically achievable. Still, seeing how both weaker and stronger students improved, we could theorize that they used ChatGPT in ways that helped them and further refined its output until they were pleased with the content. One implication for in-class scenarios is that teachers should let students first determine their preferred writing style and level, and then tailor their ChatGPT experience to their individual needs. Planners would get a version of ChatGPT focused on helping them consider all aspects while outlining their ideas, whereas revisers would rely on ChatGPT more to improve their writing afterward. Similarly, weaker students need different input from ChatGPT than stronger students do, with stronger students potentially using it at different points in the writing process, such as after writing to further improve their text's quality.

Another distinction that future studies could address is whether students with higher language proficiency levels (e.g., native speakers) utilize ChatGPT differently than less proficient or novice learners. Here, studies could classify students by their English-language scores (e.g., TOEFL or IELTS) to assign them to different groups. Using these distinct groups, studies could highlight potential differences in usage patterns and changes in the progression of writing quality.

Teachers could train ChatGPT on course content to create a tutor-like system to determine whether custom GPTs can provide better results. Based on these positive results, we recommend that students practice writing both with and without AI assistance (as in our experimental setup) to develop the ability to produce high-quality texts and deepen their understanding of the subject.

Future research could provide insights into AI's role across other aspects of higher education to test ChatGPT's effects on other subjects. If effective across multiple classes, universities might consider creating specialized models tailored to their curricula and policies. Here, it is important to consider different levels of prior writing proficiency, as there may be differences between new learners using ChatGPT and more advanced students, such as postgraduate or doctoral students who learned to write without ChatGPT. Therefore, comparing the usage behavior and learning effects of inexperienced students using ChatGPT with those of experienced students could help determine the optimal point in the curriculum for integrating AI-assisted writing. Special attention needs to be paid to whether the extended use of AI by students impairs important aspects of the academic writing process, such as idea generation, feelings of ownership, and perceived workload.

Further studies should determine the optimal integration, potentially via a mixed approach in which students alternate between ChatGPT and independent writing to develop critical thinking skills. Students could write a chapter independently first, then use ChatGPT to generate multiple versions to compare with their versions, discussing the logic, content accuracy, highlighting differences, and identifying preferred aspects in each version.

7. Conclusion

Our study provides preliminary evidence of increased writing quality in our sample of first-year fashion design students in academic writing when using ChatGPT, with improvements evident immediately after use, suggesting that integrating AI into writing curricula could enhance comprehension, skill acquisition, and achievement. Increased perceived ownership after ChatGPT use and an improved writing experience during and after ChatGPT indicate that it could strengthen students' connection to their work, perceived skills, writing affinity, and in-class involvement. Students found ChatGPT both useful and easy to use, according to the TAM. Therefore, in-class applications need to be explored in future studies to determine how ChatGPT can be used most effectively to enhance student learning and provide lasting quality improvements during and after its use.

The large number of statistical tests increases the chance of obtaining statistically significant results by chance; therefore, our results should be considered preliminary. Further studies are needed that replicate these

results to determine whether the findings can be confirmed.

This study marks a promising step as AI tools become increasingly popular and their capabilities continue to grow. ChatGPT shows potential for academic writing and teaching, and its use could enhance learning while fostering critical thinking skills.

Ethical approval

The university's ethics committee approved the study. Informed consent was obtained from all participants, and their privacy rights were strictly observed. The data can be obtained by sending request emails to the corresponding author.

CRediT authorship contribution statement

Nicolas Bauer: Writing – review & editing, Writing – original draft, Resources, Methodology, Data curation, Conceptualization. **Tim Fütterer:** Writing – review & editing, Validation, Methodology, Conceptualization. **Birgit Brucker:** Writing – review & editing, Project administration, Methodology, Conceptualization. **Peter Gerjets:**

Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

We have no known conflicts of interest to disclose. Full reproducible code, the data, and supplementary material of this study can be accessed at <https://doi.org/10.17605/OSF.IO/9CAU2>. The authors are responsible for the contents of this publication.

Acknowledgments

This research was supported by the Foundation of German Business gGmbH (Stiftung der Deutschen Wirtschaft gGmbH; sdw) as part of the BMBF's scholarship program. This research was supported by the Post-doctoral Academy of Education Sciences and Psychology of the Hector Research Institute of Education Sciences and Psychology, Tübingen, funded by the Baden-Württemberg Ministry of Science, Research, and the Arts.

Appendix A. Contextual information regarding the academic writing course – from the course Syllabus

Course Information: Academic Writing and Research

After completing the module, students will be able to:

- L1: Name the individual steps of an empirical research process. (Competence level 1)
- L2: Explain the basics of scientific discovery. (Competence level 2)
- L3: Classify sources concerning their reliability based on quality criteria for scientific work. (Competence level 2)
- L4: Summarize the state of research concerning a research question systematically and according to scientific standards. (Competence level 2)
- L5: Carry out scientific literature research. (Competence level 3)
- L6: Critically evaluate one's own approach to literature review and topic focus in a broader context. (Competence level 5)
- L7: Examine possible research ethical challenges. (Competence level 5)
- L8: Reflect on the use of AI in the writing process. (Competence level 5)
- L9: Derive research questions and/or hypotheses from the current state of research. (Competence level 6)

Contents of the Module:

- Fundamentals of scientific epistemology
- Basics of scientific writing and publishing according to the rules of the American Psychological Association (APA)
- Methods for researching and organizing scientific literature, as well as rules for assessing the quality of scientific publications
- Basics of practical procedures for recording, organizing, and solving scientific problems
- Understanding of scientific terminology and quality indicators (e.g., deduction, induction, validity, reliability, representativeness); ability to derive research questions and hypotheses
- Application or case-by-case avoidance of AI in the research process
- Techniques of scientific argumentation and reasoning
- Methods for creating and presenting scientific posters
- General Objectives of all BA Programmes
- Understanding of the interrelationships and challenges of a diverse, volatile, and agile economy and society, including an international perspective. (Relevance: 5)
- Ability to grasp complex tasks and develop creative and sustainable solutions using suitable methods. (Relevance: 5)
- Self-reflection and judgment in order to grow personally in a dynamic, international, and discontinuous working environment and to make ethically responsible decisions in intercultural contexts. (Relevance: 3)

Objectives of the Specific Study Programme

Awareness of the social role of the media and critical reflection on the economic and political characteristics of the media and communications industry in the context of social change and digital transformation, based on international competencies and individual interests. (Relevance: 3)

Ability to prepare, critically reflect, and practically apply the current international state of research on media and communication management

and related topics from various disciplines. (Relevance: 5)

Entrepreneurial spirit that reflects individual interests, combined with an innovation orientation and a willingness to assume social responsibility as a media or communications manager. (Relevance: 6)

Structure of the lecture:

U1 Introduction: Didactic concept of this course, requirements for project work

U2 The Scientific Research Process: Steps, Quality Criteria, and Logic L1, L2, L7

U3 Researching and reading I: Introduction to library systems, types of scientific publications
L3, L5, L6

U4 Researching and reading II: Introduction to library systems, types of scientific publications
L3, L5, L6

U5 Organization, time management, structure & AI: How do I organize and manage a project?
L8

U6 Thinking and Writing I: How to Write a Good Text L4

U7 Quote, report, and design I: Citation rules according to APA L2

U8 Quote, report, and design II: Reports and design according to APA L2

U9 Thinking and Writing II: How to Write a Good Text (APA Guidelines) L4

U10 Presentations in a scientific context: posters and Co. L7, L9

U11 Limitations L6, L7

U12 Wrap-up

Structure of the accompanying campus session:

U1: Introduction and developing a Topic.

U2: Conducting Literature Research

U3: Developing a Research Question

U4: Structuring a paper (2)

U5: AI in academic writing (1)

U6: Citing according to APA

U7: Formatting the paper

U8: Proofreading and final improvements

Competencies:

K1: Academic knowledge (Role: 6)

K2: Knowledge in professional practice (Role: 2)

K3: Methodological skills (research) (Role: 5)

K4: Methodological skills (professional practice) (Role: 2)

K5: Social skills (Role: 1)

K6: Personal skills (e.g., reflection, organisation) (Role: 3)

Module Length: 1 Semester

ECTS Credits: 5

Prerequisites / Recommendations: None

Total Workload

125 h [SSZ 80 h / UE-V 15 h / UE-Ü 30 h]

Study Semester: 1

Lecturer information:

The experiment took place during a regular academic writing course session, following the lesson on AI use in academic writing. An experienced professor always delivers the academic writing lecture at our university, whereas either professors or external teachers conduct the campus sessions. In our experiment, the class was taught by the first author, an experienced academic writing teacher with several semesters of teaching this subject and experience in grading academic papers. In response to the introduction of ChatGPT into academic writing, the course structure and content were adapted to reflect these changes. The course materials on AI use in academic writing were developed by the course professor and the first author of this study.

Appendix B. The experimental surface and instructions

Dear participant

Welcome to our study "sciwi01"! Thank you for your interest in this study. Please read the following information carefully and then decide whether to participate in this study or not.

In this study, we are using GPT-3 to aid the writing process in the module academic writing.

For this purpose, you will use our system to create three academic texts, each with roughly one-page-long introduction sections, based on journal articles we will provide as references.

No side effects are to be expected.

Procedure of the study

The following study consists of three sections and takes place in the module academic writing as part of your scheduled classes. In total, it takes **about 180 min**, and your extra effort in answering some questionnaires is remunerated with 5€. A prerequisite for your participation is that you are **at least 16 years old and speak fluent English**.

Your task is ...

... in Section 1) to answer a preliminary questionnaire with various questions about yourself (demographic data, experience with AI-writing tools and writing in general, ability and value beliefs with regard to academic writing) in Section 2) write an introduction about the first of three topics using only your prior knowledge and skills, followed by a short questionnaire about your writing experience in Section 3) write a second introduction about a second topic, using the aid of the AI tool GPT-3, again followed by a short questionnaire about your writing experience in Section 4) write a third introduction about a third topic, this time again without the AI tool, again followed by a short questionnaire about your writing experience in Section 5) complete a survey about your experience of writing with and without artificial intelligence assistance.

The individual sections and different tasks of the examination will each be explained to you in detail before you begin.

Voluntariness, anonymity, confidentiality and handling of the data

Participation in the study is voluntary. You may terminate your participation in this study at any time without giving any reason and without incurring any disadvantages.

All data will be kept strictly confidential and used solely for research purposes. By confirming below, you allow us until withdrawal to evaluate your data for scientific purposes and, if necessary, to use it in the context of a publication in completely anonymized form and to make it publicly available to other scientists only in this form.

If you participate, you will have the opportunity to withdraw your consent to data analysis until the end of the course on 23.06.2023, using your personal code word that you create at the start of the study based on a predefined rule.

Advantages/benefits from participation

Beyond the compensation, there is no particular benefit from participating in the study. However, you will contribute to the scientific knowledge of how information is processed in different postures. Furthermore, practicing academic writing in this experiment might potentially be beneficial for you, as it is the topic of the course.

Personal data

The following personal data will be collected in the study:

- Demographic data: Age, gender, reading/writing disability, English proficiency.
- Questionnaires on: experience with AI-writing tools and writing in general, ability and value beliefs about academic writing, technology acceptance and usability, cognitive load, sense of ownership, sense of control, sense of leadership, situational interest, and intention to use.

Instructions:

Your task in this section is to write an **introduction** to an academic paper based on the instructions on how to write introductions from our class. This should be around 250–350 words long (one page in Word using the standard formatting). In each writing session, you will receive three journal articles, which are the only sources you should use. These will be displayed within this window. You will have approximately 50 min for each task; if you do not finish, that is not an issue.

Additional instruction regarding ChatGPT:

This time, you will see an additional window that shows the version of ChatGPT you will work with today. Here you can interact with your assistant and ask it for help with the introduction.

Screenshot from the Study interface: Post-exercise questionnaire

Please rate the following statements:

	strongly disagree						strongly agree	
	1	2	3	4	5			
It was easy for me to understand the topic.	☐	☐	☐	☐	☐			
This topic was interesting for me.	☐	☐	☐	☐	☐			
I could relate to the topic personally.	☐	☐	☐	☐	☐			

Please rate the following statements:

	strongly disagree						strongly agree	
	1	2	3	4	5			
It was easy for me to complete this task.	☐	☐	☐	☐	☐			
This task was challenging to me.	☐	☐	☐	☐	☐			
I felt well prepared for the task.	☐	☐	☐	☐	☐			

Please rate the following statements:

	strongly disagree						strongly agree	
	1	2	3	4	5			
I am the main contributor to the content of this text.	☐	☐	☐	☐	☐			

Appendix C. Overview of individual variables of the readability scale and the cohesion scale

The readability scale contains the following indices from ARTE [77]: Automatic Readability Index (ARI), Coh-Metrix L2 Readability Index (CML2RI), New Dale-Chall Readability Formula (DC), Flesch Kincaid Grade Level (FKGL), Flesch Reading Ease (FRE), SMOG Grading, as well as the Text Score by Grammarly and the Reading Level by wordcounter.net.

The cohesion scale includes relevant values for within-paragraph cohesion, as defined by TAACO [81,71]. We selected indices based on both theoretical relevance and statistical suitability for our data. Because most introductions consisted of only a single paragraph, we focused primarily on local-level cohesion indices, as global or paragraph-level measures (e.g., synonym overlap across paragraphs) did not yield meaningful values for our short texts.

Following Crossley et al. [71], we included local-level indicators such as demonstratives, determiners, and noun-phrases, which are well-established predictors of cohesion in short academic texts. For lexical diversity, we used trigram-lemma TTR, which showed the strongest association with human rater scores in our dataset, while other TTR variants were collinear with cohesion measures already included in the model.

Although word2vec sentence similarity typically predicts essay quality at the global level, it showed sufficient variability and correlation with human ratings in our sample and was therefore retained. In contrast, LDA topic similarity was excluded because it displayed high collinearity with word2vec, making it redundant and reducing interpretability [81].

Variables that showed no significant variation, no meaningful correlations with quality, or were theoretically inappropriate for extremely short texts were not included, as they would not contribute to a coherent or interpretable model of writing quality.

The following items were used to calculate the cohesion scale:

- trigram lemma TTR
- Overlap of nouns into adjacent Sentences
- Synonym overlap (sentence, noun)
- Synonym overlap (sentence, verb)
- Lsa cosine similarity (adjacent sentences)
- Lsa cosine similarity (two adjacent sentences)
- Word2vec similarity (adjacent sentences)
- Word2vec similarity (two adjacent sentences)
- Number of determiners
- Number of demonstratives
- Number of demonstratives followed by a noun phrase
- Number of demonstratives functioning as a noun phrase

Appendix D. Qualitative Examples of Students' ChatGPT Usage

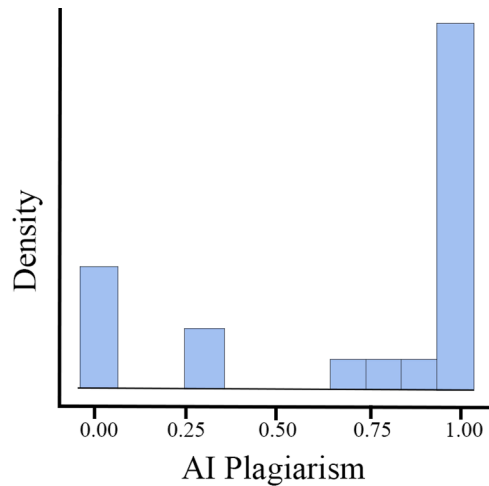
Explanation	Student Question	ChatGPT Answer
Improving Self-written text	Rewrite this introduction for a scientific paper introduction in a formal way: E-commerce as a definition for the sale of products and services online is a constantly growing and expanding, which is confirmed by the growing number of online sales to total shares. Specifically in the fashion industry, e-commerce is a widely spread and a popular way to purchase items, which also leads to about 70% of returns due to fit-related issues. In addition, almost 60% of consumers reported the lack of fitting and physical measure abilities while online shopping. With the introduction of VFR (Virtual Fitting Rooms) there have been various alternatives developed to provide the conventional fitting through the virtual interaction of the client. The key point of VFR is to try on the garments on a generated model based on our measurements which can visualise the clothes in terms of size, colour, fit, shape etc. VFR were introduced in late 2010s, mainly focusing on the performance while overlooking the affective aspects. Many of them being its accuracy, attractiveness and the extent of interaction between the software and the user. In this paper, the potential of different types of Virtual Reality on the Fashion Industry will be discussed along with the future of Augmented Realities and how current brands are implementing it.	The present paper aims to explore the implications and potential of Virtual Reality and Augmented Reality on the Fashion Industry. Specifically, the focus will be on the use of Virtual Fitting Rooms (VFR), a technology that has been developed to allow consumers to virtually interact with garments, based on their own measurements, and visualize them in terms of fit, size, color, and shape. Although VFR have been introduced in the late 2010s, their efficiency, attractiveness, and the extent of interaction with users have been subject to debate. Despite their limitations, VFR have been recognized as a viable alternative to conventional fitting processes, which often result in high rates of returns due to fit-related issues. In this context, the paper will provide an overview of the current state of the art in VFR, with a particular focus on their affective aspects. Finally, the paper will discuss the potential of Augmented Reality to further enhance the fashion retail experience, and how leading fashion brands are currently implementing this technology.
Generating Text Without Input	Write an introduction about the Potentials of Virtual Reality on the Fashion Industry	Virtual reality (VR) technology has been making waves in the fashion industry in recent years. From virtual fashion shows to virtual shopping experiences, VR has opened up a world of opportunities for designers, retailers and consumers alike. The potential for VR in the fashion industry is vast, offering immersive experiences that could revolutionize the way we design, produce and consume fashion. In this essay, we will explore the exciting potentials of virtual reality on the fashion industry and how it can contribute to the development of a more efficient, sustainable and engaging fashion ecosystem.

(continued on next page)

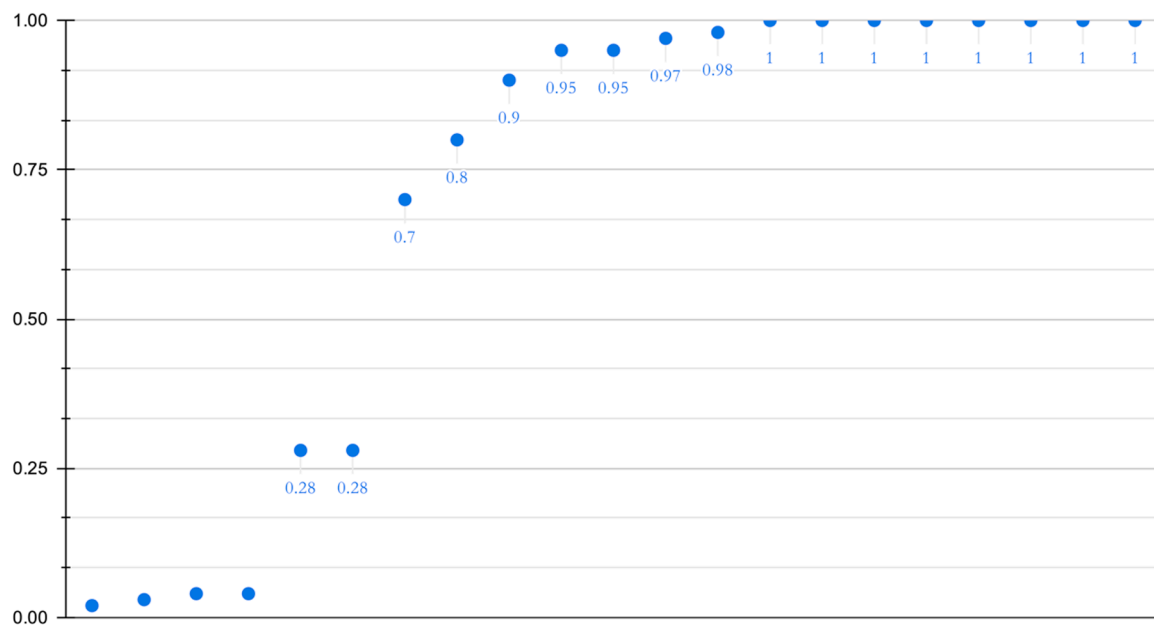
(continued)

Explanation	Student Question	ChatGPT Answer
Generating Text Without Input	Write a 250 to 350 word introduction about self-image in virtual avatars	In today's digital age, virtual environments are becoming increasingly popular. With the development of virtual reality and online gaming, individuals are able to create their own virtual avatars that serve as representations of themselves in these immersive digital spaces. A virtual avatar is an animated 3D figure that can be customized to reflect an individual's physical features, personality traits, and personal preferences. While creating a virtual avatar is often seen as a fun and entertaining activity, it also plays a significant role in shaping one's self-image. In virtual environments, individuals have the ability to construct a version of themselves that they may view as ideal or aspirational. This virtual self-image can have a profound impact on an individual's self-esteem, confidence, and overall sense of self-worth. For example, studies have shown that individuals who are dissatisfied with their physical appearance in real life may feel more confident and comfortable in their virtual avatars. Additionally, individuals may use their virtual avatars to explore different aspects of their identity, such as gender or sexuality, without fear of judgment or discrimination. However, creating a virtual avatar that accurately represents one's self-image can also be challenging. As virtual avatars are highly customizable, individuals must make a series of choices about features such as skin tone, hair style, and body shape. These decisions may reflect an individual's personal preferences but can also be influenced by societal beauty standards and biases. Furthermore, there is often pressure to conform to certain social norms and expectations within virtual environments, such as gender roles or body type ideals. Moreover, virtual avatars can have an impact on how individuals interact with others in virtual environments. For example, individuals may be judged based on the appearance or behavior of their avatars rather than their actual character. This can lead to an increased focus on superficial qualities, such as physical appearance, rather than on meaningful characteristics, such as intelligence or empathy. Overall, virtual avatars have both positive and negative implications for an individual's self-image. As technology continues to advance, it is important to consider the impact of these digital representations on our perceptions of ourselves and others in virtual environments
Questions and Answers – a thought process:	QUESTION 1: What is an Avatar	ANSWER 1: An avatar is a digital representation of a user in a virtual world or online platform. It can be a graphic image or a 3D model that represents a person or a character created by the user. Avatars are typically used in online games, virtual reality environments, social media, and messaging apps. Users can customize their avatars to look like themselves or create unique characters with different personalities, outfits, and abilities. In some cases, avatars can also be used for online identity and anonymity purposes.
	QUESTION 2: self image in virtual characters	ANSWER 2: Self-image in virtual characters, also known as avatars, refers to the way individuals perceive and project themselves through their digital representations or digital identities in virtual environments. When people create avatars, they are able to design them to represent themselves in ways that depart from their physical appearance, gender, or social status in the real world. Avatars offer a way for people to experiment with different parts of their identity, explore new or different personalities, and interact with others in a variety of social contexts. In many virtual environments, self-image can be further shaped and influenced through the acquisition of virtual assets, such as clothing, accessories, or special abilities. The concept of self-image in virtual characters is particularly important when considering the potential impacts of these virtual environments on individuals, both positive and negative. For some people, avatars may provide a space for self-expression and self-discovery, while for others, they may exacerbate issues related to body image, social anxiety, or addictive behavior.
	QUESTION 3: write an introduction about the self image in virtual avatars	ANSWER 3: Virtual avatars have become increasingly popular in recent years and have revolutionized the way we interact with technology. With the help of advanced technology, we are able to create avatars that resemble real-life individuals or characters, giving an entirely new dimension to our virtual experiences. The creation of virtual avatars has not only brought about exciting new possibilities for gaming and entertainment but has also sparked interesting discussions regarding our self-image in the virtual world. Our avatars represent the image we want to project to others in the virtual world, and it is crucial to understand how these representations affect our perception of ourselves. In this paper, we will explore the concept of self-image and how it is shaped in the virtual world through the creation of avatars.

Appendix E. Distribution of the amount of AI plagiarism



Distribution of the Amount of AI Plagiarism



References

- ## References
- [1] Gurung RK. Why academic writing? Pursuits a J Eng Stud 2022;6(1):75–82. <https://doi.org/10.3126/pursuits.v6i1.46853>.
 - [2] Lea MR, Street BV. Student writing in higher education: An academic literacies approach. Stud High Educ 1998;23(2):157–72. <https://doi.org/10.1080/03075079812331380364>.
 - [3] Elkins K, Chun J. Can GPT-3 pass a writer's Turing test? J Cult Anal 2020;5(2): 1–16. <https://doi.org/10.22148/001c.17212>.
 - [4] Abbas MS. ChatGPT in academic writing: a boon or a bane? J Electron Inf Syst 2024;6(2):16–24. <https://doi.org/10.30564/jeis.v6i2.8491>.
 - [5] Lendvai GF. ChatGPT in academic writing: a scientometric analysis of literature published between 2022 and 2023. J Emp Res Hum Res Eth 2025;20(3):131–48. <https://doi.org/10.1177/15562646251350203>.
 - [6] Dergaa I, Chamari K, Zmijewski P, Saad HB. From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. Biol Sport 2023;40(2):615–22. <https://doi.org/10.5114/biolsport.2023.125623>.
 - [7] Lund B, Ting W, Mannuru NR, Nie B, Shimray S, Wang Z. ChatGPT and a new academic reality: AI-written research papers and the ethics of the large language models in scholarly publishing. SSRN Electron J 2023. <https://doi.org/10.2139/ssrn.4389887>.
 - [8] Mondal H, Mondal S. ChatGPT in academic writing: Maximizing its benefits and minimizing the risks. Indian J Ophthalmol 2023;71(12):3600–6. https://doi.org/10.4103/ijo.ijo_718_23.
 - [9] Alsaedi N. ChatGPT and EFL/ESL writing: a systematic review of advantages and challenges. Eng Lang Teach 2024;17(5):41. <https://doi.org/10.5539/elt.v17n5p41>.
 - [10] Demsky D, Yang D, Yeager DS, Bryan CJ, Clapper M, Chandhok S, Eichstaedt JC, Hecht C, Jamieson J, Johnson M, Jones M, Krettek-Cobb D, Lai L, JonesMitchell N, Ong DC, Dweck CS, Gross JJ, Pennebaker JW. Using large language models in psychology. Nat Rev Psychol 2023;2(1):688–701. <https://doi.org/10.1038/s44159-023-00241-5>.
 - [11] Fütterer T, Fischer C, Alekseeva A, Chen X, Tate T, Warschauer M, Gerjets P. ChatGPT in education: Global reactions to AI innovations. Sci Rep 2023;13(1). <https://doi.org/10.1038/s41598-023-42227-6>. Article 15310.
 - [12] OpenAI. (n.d.). ChatGPT. ChatGPT. <https://chat.openai.com/>.
 - [13] Putra FW, Rangka IB, Aminah S, Aditama MHR. ChatGPT in the higher education environment: Perspectives from the theory of high order thinking skills. J Public

- Health (Bangkok) 2023;45(4):e840–1. <https://doi.org/10.1093/pubmed/fgad120>.
- [14] Chen M, Rau PP, Ma L. LLM asks, you write: enhancing human-AI collaborative writing experience through flipped interaction. *AHFE Int* 2025;195. <https://doi.org/10.54941/ahfe1006247>.
- [15] Draxler F, Werner A, Lehmann F, Hoppe M, Schmidt A, Buschek D, Welsch R. The AI ghostwriter effect: Users do not perceive ownership of AI-generated text but self-declare as authors. *ACM Trans Comp-Hum Inter* 2023;31(2):1–40. <https://doi.org/10.48550/arxiv.2303.03283>.
- [16] Xu X, Wang X, Zhang Y, Zheng R. Applying ChatGPT to tackle the side effects of personal learning environments from learner and learning perspective: An interview of experts in higher education. *PLoS One* 2024;19(1):e0295646. <https://doi.org/10.1371/journal.pone.0295646>.
- [17] Yildiz T, A. The impact of ChatGPT on language learners' motivation. *J Teach Edu Lifelong Learn* 2023;5(2):582–97. <https://doi.org/10.51535/tell.1314355>.
- [18] Chitrakar N, Nisanth PM. Frustration and its influences on student motivation and academic performance. *Int J Sci Res Mod Sci Tech* 2023;2(11):1–9. <https://doi.org/10.59828/ijrsmst.v2i11.158>.
- [19] Wang J, Liao Y, Liu S, Zhang D, Wang N, Shu J, Wang R. The impact of using ChatGPT on academic writing among medical undergraduates. *Ann Med* 2024;56(1). <https://doi.org/10.1080/07853890.2024.2426760>.
- [20] Yan D. Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Educ Inf Technol (Dordr)* 2023;28(11):13943–67. <https://doi.org/10.1007/s10639-023-11742-4>.
- [21] Boudouaia A, Mouas S, Kouider B. A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *J Edu Comp Res* 2024;62(6):1509–37. <https://doi.org/10.1177/07356331241247465>.
- [22] Maghamil M, Sieras S. Impact of CHATGPT on the academic writing quality of senior high school students. *J Eng Lang Teach Appl Linguist* 2024;6(2):115–28. <https://doi.org/10.32996/jeltal.2024.6.2.14>.
- [23] Mahapatra S. Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learn Environ* 2024;11(1). <https://doi.org/10.1186/s40561-024-00295-9>.
- [24] Grabe W, Kaplan RB. *Theory and practice of writing: an applied linguistic perspective*. Routledge; 1996.
- [25] Kellogg RT. Writing performance. *Writ Commun* 1987;4(3):269–98. <https://doi.org/10.1177/0741088387004003003>.
- [26] Witte SP, Faigley L. Coherence, cohesion, and writing quality. *Coll Compos Commun* 1981;32(2):189. <https://doi.org/10.2307/356693>.
- [27] Ngoc CY, Hoa CTH. Explicit lexical collocation instruction in online teaching. *Advances in wireless technologies and telecommunication book series*. 2023. p. 133–54. <https://doi.org/10.4018/978-1-6684-7034-3.ch008>.
- [28] Zhao H, Dang TNY, Finlayson N. Education lecturers' expectations about writing proficiency: Insights from corpus analysis of teacher feedback on academic writing. *J Second Lang Writ* 2024;67:101173. <https://doi.org/10.1016/j.jslw.2024.101173>.
- [29] Ayadi M. Lexical richness and syntactic complexity as predictors of academic writing performance. *J Eng Stud Arabia Felix* 2023;2(1):23–33. <https://doi.org/10.56540/jesaf.v2i1.43>.
- [30] Li C, Wang X, Qian L. Exploring syntactic complexity and text readability in an ELT textbook series for Chinese English majors. *Sage Open* 2025;15(1). <https://doi.org/10.1177/21582440251323619>.
- [31] Van Der Loo J, Krahmer E, Van Amelsvoort M. Learning how to write an academic text: a comparison of observational learning with learning by doing. *J Writ Res* 2018;9(3):233–58. <https://doi.org/10.17239/jowr-2018.09.03.01>.
- [32] Braaksma M, Rijlaarsdam G, Van Den Bergh H. Observational learning and the effects of model-observer similarity. *J Educ Psychol* 2002;94(2):405–15. <https://doi.org/10.1037/0022-0663.94.2.405>.
- [33] Kim S, Shim J, Shim J. A study on the utilization of OpenAI ChatGPT as a second language learning tool. *J Multimed Inform Syst* 2023;10(1):79–88. <https://doi.org/10.33851/jmis.2023.10.1.79>.
- [34] Boscolo P, Arfé B, Quarisa M. Improving the quality of students' academic writing: an intervention study. *Stud High Educ* 2007;32(4):419–38. <https://doi.org/10.1080/03075070701476092>.
- [35] Millin T, Millin M. English academic writing convergence for academically weaker senior secondary school students: Possibility or pipe-dream? *J Engl Acad Purp* 2018;31(1):1–17. <https://doi.org/10.1016/j.jeap.2017.12.002>.
- [36] Mullen MG, Thompson J, Murphy A, Malenczak D, Giacobbe G, Karyczak S, Holloway C, Twamley EW, Silverstein SM, Gill KJ. Evaluation of a cognitive remediation intervention for college students with psychiatric conditions. *Psychiatr Rehabil J* 2017;40(1):103–7. <https://doi.org/10.1037/prj0000254>.
- [37] Yasuda R. Fluency development through freewriting and transfer to other more structured tasks. *Lang Teach Res* 2022;(0):0. <https://doi.org/10.1177/13621688221084899>.
- [38] Bruton DL, Kirby D. Research in the classroom: written fluency: Didn't we do that last year? *Eng J* 1987;76(7):89–92. <https://doi.org/10.2307/818661>.
- [39] Roose K. The brilliance and weirdness of ChatGPT. December 6. *The New York Times*; 2022. <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>.
- [40] Stevenson C, Smal I, Baas M, Grasman R, van der Maas H. Putting gpt -3's creativity to the (alternative uses) test. 2022. arXiv, <https://arxiv.org/abs/2206.08932>.
- [41] Huang Y. Reflection on whether ChatGPT should be banned by academia from the perspective of education and teaching. *Front Psychol* 2023;14(1):1181712. <https://doi.org/10.3389/fpsyg.2023.1181712>.
- [42] Stephens BM. University of Cambridge will allow students to use ChatGPT. *The Telegraph* 2023. <https://www.telegraph.co.uk/news/2023/03/02/university-cambridge-will-allow-students-use-chatgpt/>.
- [43] Ghimire A, Edwards J. From guidelines to governance: a study of ai policies in education. *Cornell University*; 2024. <https://doi.org/10.48550/arxiv.2403.15601>. arXiv.
- [44] Dong L. Brave New World" or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of ChatGPT, behaviors of using ChatGPT, and writing proficiency. *Curr Psychol* 2024;43(21):19481–95. <https://doi.org/10.1007/s12144-024-05728-9>.
- [45] Sudrajat W, Fikri MH, Putra RPB. Help me ChatGPT! What ways does ChatGPT influence students' productivity and creativity in English academic writing? *EDUCATUM Sci J Edu* 2024;2(2):48–57. <https://doi.org/10.59165/educatum.v2i2.65>.
- [46] Dale R. GPT-3: What's it good for? *Nat Lang Eng* 2020;27(1):113–8. <https://doi.org/10.1017/s1351324920000601>.
- [47] Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. *Minds Mach (Dordr)* 2020;30(4):681–94. <https://doi.org/10.1007/s11023-020-09548-1>.
- [48] Krzanowski R. Preliminary results of experiments with autoregressive language model GPT-3 and philosophical question. *HAL-03640357* 2022. <https://hal.science/hal-03640357>.
- [49] Woo DJ, Wang Y, Susanto H. Student-AI creative writing: pedagogical strategies for applying natural language generation in schools. 2022. arXiv, <https://arxiv.org/abs/2207.01484>.
- [50] Ebyary KE. Native and non-native speakers' reaction to Grammarly-flagged errors: Implications for L2 teaching of writing. *Electron J For Lang Teach* 2022;19(2):113–29. <https://doi.org/10.56040/ebya1921>.
- [51] Dong, L. (2025). Development and validation of a multidimensional ChatGPT feedback engagement scale in second language writing. *System*, 135, 103875. <https://doi.org/10.1016/j.system.2025.103875>.
- [52] Song C, Song Y. Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Front Psychol* 2023;14. <https://doi.org/10.3389/fpsyg.2023.1260843>.
- [53] Aguilera-Hermida AP. College students' use and acceptance of emergency online learning due to COVID-19. *Int J Educ Res Open* 2020;1:100011. <https://doi.org/10.1016/j.ijedro.2020.100011>.
- [54] Li M. Non-native English-speaking (NNES) students' English academic writing experiences in higher education: a meta-ethnographic qualitative synthesis. *J Engl Acad Purp* 2024;71:101430. <https://doi.org/10.1016/j.jeap.2024.101430>.
- [55] Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv Psychol* 1988;52(1):139–83. [https://doi.org/10.1016/s0166-4115\(08\)62386-9](https://doi.org/10.1016/s0166-4115(08)62386-9).
- [56] Silva BB, Kutylowska K, Otwinowska A. Learning academic words through writing sentences and compositions: Any signs of an increase in cognitive load? *Lang Teach Res* 2021;28(3):1143–75. <https://doi.org/10.1177/13621688211020421>.
- [57] Sweller J. Cognitive load theory and educational technology. *Edu Tech Res Dev* 2019;68(1):1–16. <https://doi.org/10.1007/s11423-019-09701-3>.
- [58] Bai L, Liu X, Su J. ChatGPT: The cognitive effects on learning and memory. *Brain-X* 2023;1(3). <https://doi.org/10.1002/brx2.30>. Article e30.
- [59] Ali JKM, Shamsan MAA, Hezam TA, Mohammed AAQ. Impact of ChatGPT on learning motivation: Teachers and students' voices. *J Eng Stud Arabia Felix* 2023;2(1):41–9. <https://doi.org/10.56540/jesaf.v2i1.51>.
- [60] Caratiquit K, Caratiquit LJC. ChatGPT as an academic support tool on the academic performance among students: The mediating role of learning motivation. *J Soc Human Edu* 2023;4(1):21–33. <https://doi.org/10.35912/jshe.v4i1.1558>.
- [61] Muñoz SaS, Gayoso GG, Huambo AC, Tapia RDC, Incaluque JL, Aguila OEP, Cajamarca JCR, Acevedo JER, Rivera HVH, Arias-González JL. Examining the impacts of ChatGPT on student motivation and engagement. *Soc Space J* 2023;23(1):1–27.
- [62] Yilmaz FGK, Yilmaz R. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Comp Edu Artif Intell* 2023;4(1). <https://doi.org/10.1016/j.caeai.2023.100147>. Article 100147.
- [63] Artino AR. Think, feel, act: Motivational and emotional influences on military students' online academic success. *J Comput High Educ* 2009;21(2):146–66. <https://doi.org/10.1007/s12528-009-9020-9>.
- [64] Pierce JL, Kostova T, Dirks KT. The state of psychological ownership: integrating and extending a century of research. *Rev Gen Psychol* 2003;7(1):84–107. <https://doi.org/10.1037/1089-2680.7.1.84>.
- [65] Mikalayeva L. Motivation, ownership, and the role of the instructor in active learning. *Int Stud Perspect* 2015;17(2):214–29. <https://doi.org/10.1093/isp/ekv001>.
- [66] Li Z, Liang C, Peng J, Yin M. The value, benefits, and concerns of generative AI-powered assistance in writing. In: *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*; 2024. p. 1–25. <https://doi.org/10.1145/3613904.3642625>.
- [67] Tavakol M, O'Brien D. The importance of crafting a good introduction to scholarly research: strategies for creating an effective and impactful opening statement. *Int J Med Educ* 2023;14:84–7. <https://doi.org/10.5116/ijme.6499.82af>.
- [68] Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. *Educ Sci (Basel)* 2023;13(4). <https://doi.org/10.3390/educsci13040410>. Article 410.

- [69] Caulfield J, March 27. Writing a research paper introduction | step-by-step guide. Scribbr; 2023. <https://www.scribbr.com/research-paper/research-paper-introduction/>.
- [70] Chiang S. The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System* 2003;31(4):471–84. <https://doi.org/10.1016/j.system.2003.02.002>.
- [71] Crossley SA, Kyle K, McNamara DS. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behav Res Methods* 2016;48(4):1227–37. <https://doi.org/10.3758/s13428-015-0651-7>.
- [72] McNamara DS, Crossley SA, Roscoe R. Natural language processing in an intelligent writing strategy tutoring system. *Behav Res Methods* 2012;45(2):499–515. <https://doi.org/10.3758/s13428-012-0258-1>.
- [73] Chan KKY, Bond T, Yan Z. Application of an automated essay scoring engine to English writing assessment using many-Facet Rasch measurement. *Lang Test* 2022;40(1):61–85. <https://doi.org/10.1177/02655322211076025>.
- [74] Kundu A, Barbosa D. Are large language models good essay graders? arXiv (Cornell University; 2024. <https://doi.org/10.48550/arxiv.2409.13120>.
- [75] Welsby P, Cheung BMY. ChatGPT. *Postgrad Med J* 2023;99(1176):1047–8. <https://doi.org/10.1093/postmj/qgad056>.
- [76] Graham S, Harris KR, Hebert M. It is more than just the message: presentation effects in scoring writing. *Focus Except Child* 2011;44(4):1–2. <https://doi.org/10.17161/foec.v44i4.6687>.
- [77] Choi JS, Crossley SA. Advances in readability research: a new readability web app for English. In: 2022 International Conference on Advanced Learning Technologies (ICALT). IEEE; 2022. p. 1–5.
- [78] De Haan P, Van Esch C. Assessing the development of foreign language writing skills: Syntactic and lexical features. In: Fitzpatrick E, editor. *Corpus linguistics beyond the word*. BRILL; 2007. p. 185–202. https://doi.org/10.1163/9789401203845_012.
- [79] Kol S, Scholnick M, Spector-Cohen E. Google Translate in academic writing courses? *EUROCALL Rev* 2018;26(2):50–7. <https://doi.org/10.4995/eurocall.2018.10140>.
- [80] Loxterman JA, Beck IL, McKeown MG. The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Read Res Q* 1994;29(4):352–67. <https://doi.org/10.2307/747784>.
- [81] Crossley SA, Kyle K, Dascalu M. The tool for the automatic analysis of cohesion 2.0: Integrating semantic similarity and text overlap. *Behav Res Methods* 2019;51(1):14–27. <https://doi.org/10.3758/s13428-018-1142-4>.
- [82] Bergström J, Knibbe J, Pohl H, Hornbæk K. Sense of agency and user experience: Is there a link? *ACM Trans Comp-Hum Inter* 2022;29(4):1–22. <https://doi.org/10.1145/3490493>.
- [83] Tápal A, Oren E, Dar R, Eitam B. The Sense of Agency Scale: a measure of consciously perceived control over one's mind, body, and the immediate environment. *Front Psychol* 2017;8(1):1552. <https://doi.org/10.3389/fpsyg.2017.01552>.
- [84] Eccles JS, Adler TF, Futterman R, Goff SB, Kaczala CM, Meece JL, Midgley C. Expectancies, values and academic behaviors. In: Spence JT, Freeman WH, editors. *Achievement and achievement motives*; 1983. p. 74–146.
- [85] Gaspard H. Promoting value beliefs in mathematics: a multidimensional perspective and the role of gender. University of Tübingen]. University of Tübingen; 2015 [Doctoral dissertation].
- [86] Harackiewicz JM, Tacketts Y, Canning E, Hyde JS. Harnessing values to promote motivation in education. *Adv Motiv Achieve* 2014;71–105. <https://doi.org/10.1108/s0749-742320140000018002>.
- [87] Leaper C. More Similarities than Differences in contemporary Theories of social development? *Adv Child Dev Behav* 2011;40:337–78. <https://doi.org/10.1016/b978-0-12-386491-8.00009-8>.
- [88] Bauer CA, Job V, Hannover B. Who gets to see themselves as talented? Biased self-concepts contribute to first-generation students' disadvantage in talent-focused environments. *J Exp Soc Psychol* 2023;108(1). <https://doi.org/10.1016/j.jesp.2023.104501>. Article 104501.
- [89] Lefly DL, Pennington BF. Reliability and validity of the adult reading history questionnaire. *J Learn Disabil* 2000;33(3):286–96. <https://doi.org/10.1177/002221940003300306>.
- [90] Aljarrah E, Elrehail H, Aababneh B. E-voting in Jordan: Assessing readiness and developing a system. *Comput Human Behav* 2016;63:860–7. <https://doi.org/10.1016/j.chb.2016.05.076>.
- [91] Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability evaluation in industry*. CRC Press; 1996. p. 189–94.
- [92] Kirakowski J, Corbett M. SUMI: the software usability measurement inventory. *Br J Edu Tech* 1993;24(3):210–2.
- [93] Kirakowski J. The software usability measurement inventory: Background and usage. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability evaluation in industry*. Taylor & Francis; 1996.
- [94] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Manage Inf Syst Quart* 1989;13(3):319–40. <https://doi.org/10.2307/249008>.
- [95] Lewis JR. Comparison of four TAM item formats: Effect of response option labels and order. *J Usab Stud* 2019;14(4):224–36.
- [96] Lewalter D, Willems A. Die Bedeutung des motivationsrelevanten Erlebens und des individuellen Fachinteresses für das situationale Interesse im Mathematikunterricht. *Psychologie in Erziehung Und Unterricht* 2009;56(4):243–57. <https://psycnet.apa.org/record/2009-22902-001>.
- [97] The Jamovi Project. jamovi (Version 2.3) [Computer Software, <https://www.jamovi.org>]; 2023.
- [98] Athanassopoulos S, Manoli P, Gouvi M, Lavidas K, Komis V. The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Adv Mob Learn Edu Res* 2023;3(2):818–24. <https://doi.org/10.25082/amler.2023.02.009>.
- [99] Toktosunova A, Fallah H, Toutouchi Z, Tick A. The impact of ChatGPT on learning in higher education – Results of a pilot study. *Pract Theo Syst Edu* 2023;18(1):31–50.
- [100] Wu R, Yu Z. Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *Br J Edu Tech* 2023;55(1):10–33. <https://doi.org/10.1111/bjjet.13334>.
- [101] AlZaabi A, AlAmri A, Albalushi H, Aljabri R, AalAbdulsalam AR. ChatGPT applications in academic research: a review of benefits, concerns, and recommendations. *BioRxiv* 2023. <https://doi.org/10.1101/2023.08.17.553688>.
- [102] Imran M, Almusharraf N. Analyzing the role of ChatGPT as a writing assistant at higher education level: a systematic review of the literature. *Contemp Educ Technol* 2023;15(4). <https://doi.org/10.30935/cedtech/13605>. Article ep464.
- [103] Hawanti S, Zubayduloevna KM. AI chatbot-based learning: Alleviating students' anxiety in English writing classroom. *Bull Soc Inform Theo Appl* 2023;7(2):182–92. <https://doi.org/10.31763/businta.v7i2.659>.
- [104] Alneyadi S, Wardat Y. ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemp Educ Technol* 2023;15(4):ep448. <https://doi.org/10.30935/cedtech/13417>.
- [105] Perkins M. Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *J Univ Teach Learn Pract* 2023;20(2). <https://doi.org/10.53761/1.20.02.07>. Article 07.
- [106] Wasi AT, Islam MR, Islam R. LLMs as writing assistants: exploring perspectives on sense of ownership and reasoning. 2024. arXiv, <https://arxiv.org/abs/2403.12345>.
- [107] Selenko E, Bankins S, Shoss M, Warburton J, Restubog SLD. Artificial intelligence and the future of work: a functional-identity perspective. *Curr Dir Psychol Sci* 2022;31(3):272–9. <https://doi.org/10.1177/09637214221091823>.
- [108] Kobiella C, López YSF, Waltenberger F, Draxler F, Schmidt A. If the machine is as good as me, then what use am I? – How the use of ChatGPT changes young professionals' perception of productivity and accomplishment. *CHI '24 Proc CHI Conf Hum Fact Comp Syst* 2024;1(1). Article 1018.
- [109] Weber CJ, Burgkart S, Rothe S. wr-AI-ter: Enhancing ownership perception in AI-driven script writing. In: Proceedings of the 2024 ACM International Conference on Interactive Media Experiences. Association for Computing Machinery; 2024. p. 145–56.
- [110] Zhai X. ChatGPT user experience: Implications for education. *Social science research network*. 2022. <https://doi.org/10.2139/ssrn.4312418>.
- [111] Mahyoob M, Algaraady J, Alblwi A. Proposed framework for human-like language processing of ChatGPT in academic writing. *Int J Emerg Tech Learn/Int J Emerg Tech Learn* 2023;18(14):282–93. <https://doi.org/10.3991/ijet.v18i14.41725>.
- [112] Harnad S. Language writ large: ILMs, ChatGPT, meaning, and understanding. *Front Artif Intell* 2025;7. <https://doi.org/10.3389/frai.2024.1490698>.
- [113] Martínez G, Hernández JA, Conde J, Reviriego P, Merino-Gómez E. Beware of words: evaluating the lexical diversity of conversational LLMs using ChatGPT as case study. *ACM Trans Intell Syst Technol* 2024. <https://doi.org/10.1145/3696459>.
- [114] Sok S, Heng K. Opportunities, challenges, and strategies for using ChatGPT in higher education: a literature review. *J Digit Edu Tech* 2023;4(1):ep2401. <https://doi.org/10.30935/jdet/14027>.
- [115] Zhu F, Zou W. The role of generative AI in human creative processes: Experimental evidence. *Soc Sci Res Netw* 2024. <https://doi.org/10.2139/ssrn.4676053>.
- [116] Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175–91. <https://doi.org/10.3758/bf03193146>.

Further reading

- [117] Şahin M., Aybek E.C. Jamovi: An easy to use statistical software for the social scientists. *Int J Assess Tool Edu* 2020;6(4):670–692. [doi:10.21449/ijate.661803](https://doi.org/10.21449/ijate.661803).