Student experiences of ChatGPT as a feedback tool in higher education

Tommy Bruhn, University of Copenhagen
Franziska Marquart, University of Copenhagen

Abstract

Generativ kunstig intelligens giver både udfordringer og muligheder for videregående uddannelse. Få studier har indtil nu taget højde for studerendes erfaringer med formålstjenestelig brug af generativ AI. Denne artikel tager afsæt i en undersøgelse af to hold universitetsstuderende, der har brugt ChatGPT til at generere feedback på skriftlige opgaver. De studerendes holdninger er blevet indsamlet gennem en survey, reflektionsprotokoller og klassebaserede diskussioner. Analyserne viser, at studerende oplevede deres rolle som modtagere af feedback kvalitativt anderledes i feedbacksituationen med AI end med medstuderende eller underviseren, idet de med AI følte, at alt ansvar for den kritiske vurdering af prompts- og svar falder på dem selv. De studerende følte, at det var en mere frustrerende proces, men følelsesmæssigt lettere at bede ChatGPT om feedback end at bede deres medstuderende eller underviseren om feedback, hvilket peger på vigtige forskelle i sociale- og interaktive dynamikker mellem feedbackmodtagere og menneskelige versus AI-feedbackgivere.

English abstract

Generative artificial intelligence provides both challenges and opportunities for higher education. Few studies to date have accounted for student experiences of purposeful use of generative AI. This article reports on a mixed methods study of two university classes using ChatGPT to generate feedback on written assignments. Students' attitudes were collected through a survey, lab reports, and in-class discussions. The analyses show that students experienced their role as feedback receiver qualitatively different in the AI feedback situation compared to teacher- and peer feedback, because they themselves had to assume all the responsibility for the critical judgment of prompts and replies. Students felt that asking ChatGPT for feedback was more frustrating but emotionally easier than asking peers or teachers, which points to important differences in the dynamics of sociality and interaction between feedback receivers and human vs. AI feedback givers.

ISSN: 1903-248X

Introduction

With the public launch of Open AI's chatbot ChatGPT in November 2022, generative artificial intelligence (AI) technology became widely available, and within less than a year an increasingly complex AI market sprouted (Garg, 2023). A large part of the public discussions on the use of artificial intelligence in higher education has been revolving around the implications of the technology as a possible tool for students cheating on tests, assignments, and exams. Facing plagiarism and fraud cases, teachers claimed to be in "full-on crisis mode", with students copying entire answers from ChatGPT and similar bots (Zilber, 2023) while schools and universities struggle to protect academic integrity. However, a focus on the negative implications and threats of artificial intelligence is bound to leave educators always one step behind technical developments.

Many of the professions that current university students will enter after graduation may rely on AI in their daily workflow (Kelly, S. et al., 2023). It is therefore important that students learn the necessary skills that enable them to use AI in a responsible and critical manner. Generative AI can provide many opportunities for learning and knowledge generation; it is both necessary and valuable to explore these opportunities early on and in cooperation with students in order to foster digital literacy skills and open conversations about the contexts in which AI can and cannot be put to good use (Cotton et al., 2023).

Among other possible applications, AI can be employed for feedback generation (Mollick & Mollick, 2023), and students already use ChatGPT in this manner (Møgelvang et al., 2023). Since chatbots based on large language models (LLMs) rely on repositories of real-world written communication, they operate on both tacit and explicit norms of written language attained through their programming and training data. Despite important differences between AI and human feedback givers, the possible use of AI in generating feedback on academic text is a promising avenue of research. Whereas there is a growing body of research on the comparability and agreement of AI and human feedback givers, there is to date very little scholarship on how students understand and respond to the interaction with AI tools for generating feedback (Hasse & Bruun, 2023).

In this paper, we ask how students experience interacting with a conversational AI (CA; ChatGPT 3.5) as a supplementary feedback tool in a university context. We ask (RQ1) how students critically evaluate the feedback generated by the AI, (RQ2) how they compare it to peer- and teacher feedback, and (RQ3) what attitudes and emotions students express in

relation to the AI's feedback style, tone, and its usefulness. We developed classroom exercises carried out in two BA courses at the University of Copenhagen, where students used ChatGPT to generate feedback on assignments. Students' opinions about the exercise were assessed via qualitative lab reports, in-class discussion, and a survey. Our study contributes to the knowledge about how AI can be implemented in higher education through clarifying some of the particularities of AI feedback as opposed to other formats and highlighting important considerations for using ChatGPT as a complementary feedback tool.

We begin by discussing current research on AI in educational environments, forms and effects of feedback, and computer-mediated and AI-generated feedback. After a description of our mixed-methods approach, we present results which contextualize our analysis of how students perceived the AI feedback situation as a learning experience, and of their affective responses to the AI. We end the paper with discussing how our findings on interactional dynamics connect to and complement current research, including some limitations of our approach and potentials of using AI for feedback on academic writing.

Human-computer interaction in higher education

Prior research has shown that the adaptation and success of AI in an educational environment is connected to students' level of trust in the technology, its perceived usefulness and ease of use, and the communication style of the conversational agent (e.g., a chatbot). Bilquise and colleagues (2023) document that students are more willing to accept a chatbot for academic advising if they perceive the interaction as user-friendly/effortless and efficient (cf. Kim et al., 2020). Users appreciate conversing with a chatbot "that possesses some social-emotional intelligence and is not robotic" (Bilquise et al., 2023, p. 18). In other words, text-based CAs that appear more 'human' and even express sympathy are met with more favourable attitudes by users (Liu & Sundar, 2018; Rapp et al., 2021). Findings from Kim et al. (2021) further confirm that the communication style of an AI influences student perceptions, in the sense that a relational as opposed to functional instruction style positively affects attitudes towards AI instructor-based communication. The authors document that relationship-building communication fosters social presence, which in turn increases positive attitudes and the intention to further use the technology. However, particularly first- or second year students are often more comfortable when receiving guidance from a teacher (Bilquise et al., 2023). The social interaction and relational aspect of the teacher-student relationship can be crucial for academic success and may be difficult to replace with AI. Nevertheless, AI-generated feedback may support student learning if it supplements inputs from teachers and peers.

Forms and effects of feedback

Feedback spans a range of teaching and learning practices, where a student is provided information on their performance and understanding of the subject matter. Hattie and Timperley (2007) argue that feedback can focus on four levels: corrective feedback on the *task* itself, the *process* of understanding and performing tasks, the *self-regulation* of students in learning activity and the personal *self* (i.e. assessing student character). Common formats include teacher feedback, sometimes called formative assessment (e.g., Yorke, 2003), and peer-, self- and automated feedback. Ideas of feedback typically fall within one of two paradigms: a *cognitivist* paradigm that foregrounds input practices and transmission of information on goal attainment, or a *social constructivist* paradigm that focuses on interaction and student activity (Winstone & Carless, 2020). While our study focuses on the latter, the ideas of feedback implied by our study participants span both.

The extensive research literature on feedback suggests that it is generally beneficial for student learning and performance. A range of factors have been studied for their impact on learning through feedback (e.g., anonymity, timing, rubric- vs. in-text based feedback, etc.). Just how large their effects are is debated, and the varying results can at least in part be explained by the heterogeneity of feedback practices. Therefore, different feedback forms need to be studied both independently and comparatively (Wisniewski et al., 2020). Research suggests that while students generally find teacher feedback more useful than peer feedback, different feedback formats can play complementary roles. For example, perceived usefulness does not predict the amounts of revisions students make (Yu & Lee, 2016). In general, results are better when feedback is integrated into the everyday learning practices (Graham et al., 2015).

Students' experiences of and emotional responses to feedback present some important considerations for its implementation (Hadden & Frisby, 2019). These can influence student *feedback literacy*, that is, students' ability to make sense and use of comments on performance. Carless & Boud (2018) divide feedback literacy in four categories: appreciating the role and function of feedback; making sound judgments about the work; taking appropriate action in response to feedback and managing affect. These can be facilitated by the social study environment through feedback practices and reception, and vice versa. For peer feedback to work well, there needs to be mutuality and engagement, as well as guidance from teachers (Evans, 2015). Hamer and colleagues (2015) found that tutors and high performing students identified more points to comment on and were more specific than peers but were also more prone to give negative comments. While a negative comment could therefore be

taken as a sign of expertise, it might also signal hierarchy or disrespect which some students could respond negatively to (Taggart & Laughlin, 2017). Praise may also be counterproductive, as it operates on the *self* level and promotes a reputational lens for students to understand themselves (Hattie & Timperley, 2007). Cho and colleagues (2006) found that while peers tend to provide shorter feedback texts than experts (i.e., teachers), both types of comments were perceived as equally helpful by students. However, peer feedback contained more praise and was therefore highly valued by students. In contrast, Bader and colleagues (2019) found that students tended to value teacher feedback higher than peer feedback because it contained more constructive criticism, while peer feedback was often described as too positive and not providing direction for improvement. These seemingly diverging results highlight the important dynamic between positive and constructive criticism, which seems to differentiate peer- from teacher feedback.

As our analysis will show, the AI feedback format blurs the roles of feedback giver and -receiver in important ways. This provides perspective on Møgelvang et al.'s point that AI feedback is an individual activity, as opposed to the dialogical emphasis in much feedback research (2023, p. 55). Research results on the learning effects for feedback givers are inconsistent, with some studies suggesting that feedback giving is particularly beneficial for global aspects of writing such as content, organization, and cohesion (Yu & Lee, 2016). There is also reason to assume that habitual patterns of human feedback interaction, such as collaborative, expert/novice, dominant/passive (Storch, 2002) may be confused in interactions with an AI, especially by students with little experience of it.

Computer-mediated and AI-generated feedback

Studies on anonymized computer-mediated feedback show that it is experienced as a less threatening environment than in-person situations, alleviating anxiety as well as facilitating equality and participation because time- and presence constraints are removed (Savignon & Roithmeier, 2004). Anonymity has also been shown to encourage critical feedback (Panadero & Alqassab, 2019). Students may, however, perceive computer-mediated feedback as more challenging and less dialogical than face-to-face interaction. Other studies suggest that computer-mediated feedback can reduce stress, and lead to more comprehensive feedback-giving from peers than in-person feedback (Yu & Lee, 2016). These results are pertinent to our study, but importantly concern feedback that is given and received asynchronously. As ChatGPT mimics synchronous online interaction of web chatting, it could potentially lead to a hybrid of computer-mediated and face-to-face feedback experiences.

There is a small but growing body of work on using ChatGPT for feedback generation in an educational context. Comparing the quality of feedback provided by ChatGPT and human raters for school student essays, Steiss and colleagues (2023) found that humans significantly 'outperform' the AI on most measured quality indicators (e.g., accuracy, supportive tone, clear directions for improvement). However, ChatGPT offered better criteria-based feedback by explicitly referencing principles for source-based argumentative writing. ChatGPT also occasionally delivered inaccurate comments, which may demotivate or mislead students. These weaknesses notwithstanding, Steiss et al. (2023) point out that "the ChatGPT feedback was still of a relatively high quality" (p. 18) and similar to that provided by human raters, particularly considering that these were well trained and had sufficient time to formulate their feedback. For regular teachers and educators, time is often a scarce resource (Jacobsen & Weber, 2023), and under these circumstances AI-generated feedback may be a valuable addition in- and outside of the classroom.

Dai and colleagues (2023) also tested ChatGPT as a feedback giver. They compared comments on student projects provided by instructors and ChatGPT in terms of readability and assessed agreement between both on different criteria. Text generated by ChatGPT scored higher on general readability than feedback provided by instructors, and it included process-oriented feedback as well (see Hattie & Timperley, 2007). However, instructors and ChatGPT 'disagreed' considerably in their assessment of student performance, indicating that the chatbot is a less reliable measurement tool despite having been supplied with relevant performance indicators.

These studies provide useful comparisons of different feedback givers. However, neither Steiss et al. nor Dai and colleagues asked the respective students about their perceptions of the AI-generated feedback or assessed how they evaluate ChatGPT's output compared to what they expect (or experienced) from educators or peers. Such perceptions are crucial for the adaptation of an AI for feedback tasks, since students themselves will be the ones implementing it. It does not only matter whether educators and chatbot 'agree' about student performance or provide similar comments to submitted assignments. Students also need to be able to understand, critically evaluate, accept (or dismiss), and be willing to incorporate AI-generated feedback for it to be a useful supplement to other feedback types. Furthermore, it is important that students learn how to ask for feedback when conversing with an AI (Jacobsen & Weber, 2023): Which criteria are relevant to consider? How generic or specific should ChatGPT's feedback be in order to be useful?

Our study takes these questions into account: Rather than contrasting instructor- and AI-based feedback, we asked our students to construct their own prompts to acquire what they consider helpful feedback from ChatGPT. We assessed their experiences with ChatGPT in general and the feedback-specific task in particular and considered students' perceptions of useful prompting and related responses. This is important, since the interaction with the AI requires students to understand what feedback they need from the chatbot and to have the necessary skills to request it via prompts (Jacobsen & Weber, 2023). Effectively communicating with the AI to achieve one's goals can thus be understood as a form of feedback giving, where students, through the editing and refinement of requests, understand what to ask for in order to receive the most helpful comments in return (cf. Mollick & Mollick, 2023).

Method

To generate a relevant set of qualitative data on student experiences of using ChatGPT as a feedback tool, we set up exercises on feedback generation in two second-semester Bachelor courses at the University of Copenhagen's Department of Communication (March 2023). We gathered students' experiences and reflections via a survey, qualitative lab reports, and class discussions in a mixed-methods design. Students were informed that the study was part of a research project by their two respective teachers, and participation in the exercise was voluntary. One class focused on "Rhetorical writing and stylistics" (Rhetoric program), the other on "Qualitative and quantitative research methods" (Film and Media Studies program; F&M). Students in both classes used assignments that they had previously submitted as part of the course requirements, and they had already received feedback on these assignments from their teacher and peers.

In the F&M course, students were tasked with using ChatGPT 3.5 to generate feedback for research questions and survey questionnaires that they had previously developed in study groups. The exercise took place in class and was prefaced by a short (10min) introduction on generative AI and ChatGPT. The instructions were kept deliberately open: Students were tasked with generating feedback on their written work and encouraged to be "clear and specific" in their questions to the chatbot, and to see what it could 'teach' them about improving their work. The students worked in their study group for approximately 20min and subsequently filled in the standardized survey.

Students in the rhetoric program were asked to use ChatGPT to generate feedback on their own 2-page literature report, written in pairs. They were encouraged to explore the usefulness of the

AI alone or in pairs as a voluntary homework assignment. The thematic introduction was a slightly modified version of the one given to F&M students. Rhetoric students were asked to try to get ChatGPT to give feedback on different aspects of their writing (e.g., structure, grammar, style, clarity). They were tasked with taking notes on the AI's responses, reflect on their impressions, and analyse the feedback they received with regard to the criteria used in the survey (e.g., level of generic vs. custom feedback, usefulness, etc.). The students were asked to fill out lab reports, using a template with four open questions (see Table 1). The lab report method was chosen to enable more immediate reflections than the survey, as the students completed the feedback exercises in their own time and the survey in class. We also expected that the lab reports would result in longer and more exemplified reflective texts than the open text survey questions. They provided more comprehensive material for qualitative analysis adding depth to statements that were only superficially mentioned in the survey. In total, 19 of these lab reports were submitted.

Table 1. Questions for lab reports, Rhetoric program.

1	What are your experiences of working with ChatGPT in this way?
2	What types of feedback queries did you pose to ChatGPT, and how were these handled in its replies?
3	What general observations would you make on using ChatGPT to generate feedback on writing assignments?
4	How does ChatGPT as a feedback tool compare to other feedback experiences you have from your education, such as teacher- peer- and/or group feedback?

After having completed the exercise, students in both courses answered a standardized online questionnaire to allow for comparability between the two classes. They also engaged in plenary and group discussions with their teachers taking notes. The instructions for feedback generation and the way in which students' impressions and experiences were gathered were more structured in the Rhetoric course, to account for the different text genres that students were working with in each class. The data set reflects this difference mainly in how long and

reflexive the free text answers were, albeit not so much in the sentiments and experiences reported which were consistent across both groups.

In total, 64 students filled out the online questionnaire, and 51 had participated in the feedback exercise (F&M *n* = 34, rhetoric *n* = 17); we focus on these 51 students in our reports below. They were between 19 and 33 years old (*M* = 22.37; *SD* = 2.41), and 66.7% identified as female. On average, students reported a high general interest for trying out new technologies (*M* = 5.15, *SD* = 1.48, scale: 1 = not at all interested to 7 = very interesting), and almost half of them (49%) had heard about ChatGPT but never tried it before it was discussed in the classroom.

Aside from closed questions with answer scales, the questionnaire also included open-ended questions (see Appendix). These asked, for example, for students' evaluation of the results that they received from the bot and inquired whether they thought that the AI was particularly good or bad at something. The survey questions were designed to assess students' experiences both quantitatively (to allow for comparisons between the two courses) and qualitatively to get in-depth insights into their experiences with the chatbot. While their phrasing was inspired by previous work on human-computer interaction in higher education (e.g., Kim et al., 2020, 2021), they were created for the purpose of this study, not least given the lack of prior work on the use of ChatGPT for feedback generation. We did not ask for or collect any of the actual prompts or ChatGPT replies, although a handful of students submitted them as examples in their lab reports (see also question 2, Table 1). Our results therefore reflect early impressions of familiarization with this new technology, and a higher degree of training and experience may likely have yielded different results (cf. Kelly, A. et al., 2023). However, the relatively "raw" experiential data generated by our method provide insights that would not have surfaced if ChatGPT use had been more normalized among the respondents.

The answers to the open-ended questions in the survey were combined with those from the lab reports and the teacher discussion notes using NVivo and analysed as one data set. The analysis proceeded using a substantive coding strategy based in grounded theory (Holton, 2007), meaning that all general categories were formulated through the open coding of student answers. Sentences were coded for their general thematic category and concretization (e.g. "Other uses" was concretized as "idea generation", "summarizing" etc.) and for sentiment (e.g. "AI politeness" as "critical" or "appreciative"). Any sentence in the material could be coded for more than one theme (e.g. "comparison human feedback" and "learning experience"). Two types of themes emerged from this strategy, namely sentence content themes (e.g. Danish proficiency, feedback), that is, what a sentence was talking *about*, and

pragmatic themes (e.g. emotional expression, trust statements, comparison) which reflected common kinds of speech acts in the data set. This coding disclosed the common themes, points, and observations the students raised in their discussions and written reflections, and how these intersected in their answers, which we analytically approached as explications of these themes.

Results

Students in both classes had limited prior experience with ChatGPT. Of the F&M students, 57.9% reported never having used the chatbot before, compared to 30.7% in Rhetoric. In contrast, 23.1% of rhetoric students and 10.5% of F&M students said they had previously used the tool more than six times. A third of the sample indicated that they had used ChatGPT or a similar technology in a school or university setting before – to a large degree (84%) on their own initiative, rather than motivated by a teacher or tutor. When asked about how they experienced working with ChatGPT in our class exercise, students' answers ranged between 4 and 7 on a 7-point Likert scale (1 = very negative; 7 = very positive), indicating a positive experience overall (M = 5.45; SD = 1.03). Students in the F&M program were significantly more positive in their evaluation (M = 5.71, SD = .97) than their Rhetoric colleagues (M = 4.94, SD = .97; F(1, 49) = [7.06], p = .011).

The students' responses reflect their arguably low level of familiarity with ChatGPT, and several reported feeling surprised at its capabilities. While many described positive experiences, they were also critical, sceptical, or hesitant about the tool. Virtually every student filling out the questionnaire or lab report mentions, at some point or in some capacity, that the chatbot results need to be considered critically. This could be attributed to the novelty of the exercise and to the contemporary media discourse about AI chatbots (e.g., regarding 'hallucinations'). But some students report specific examples from their trial use that motivate a cautious attitude (see further below).

In our survey, we asked students to rate the output they received from ChatGPT regarding different criteria (Table 2). We see some differences between the two program groups, likely related to the type of text they requested feedback on and the details with which students designed their prompts.

Many of the students perceived the feedback as basic, general, or generic (see also Table 2). Some report that ChatGPT drifted away from the text for feedback used in the prompt into giving

general writing advice. Such feedback was rarely judged as outright useless, but most students seemed to put some effort into generating more concrete feedback on specific arguments, stylistic choices, etc. Some students submitted verbatim examples of this general writing advice, for example "[prompt]'How can I work with the text to make it more clear?' to which the chat replied [...] that I should thoroughly analyse the situation and consider its purpose". While certainly generic, this is an example of feedback on the processual level. As most responses and reported prompts suggest, students approached the AI in quite an instrumental way, focusing on the task level of feedback (Hattie & Timperley, 2007). In general, their prompts typically used verbiage focusing on "this assignment", "this text" and similar. One student describes his/her process as follows:

> "I started by asking it this: How can I work on making the text clearer?: (insert text) [...] Thereafter, I wrote: Give feedback on this academic assignment: (insert text) [...] Thereafter, I asked it to correct language errors. [...] Finally, I asked it to correct punctuation errors."

One student wrote that "It couldn't answer whether our survey was fitting for our research question", and another reported trying to have it base its answer on the assignment description. Another pointed out that "When I asked, 'can you give criticism and feedback on the following assignment with special focus on ...' it gave me good answers". A small number of students report posing prompts inviting feedback on other levels, e.g.: "How can I sound more academically correct?" or "I think the authors are being very presumptuous, do you agree?". These queries ask for feedback on the level of self – which ChatGPT reportedly declined to answer.

Table 2. Students' perceptions of ChatGPT feedback.

	Rhetoric (n = 17)	Film & Media (n = 34)	Total (n = 51)
Not helpful at all/Very helpful	5.06	4.91	4.96
Not trustworthy at all/Very trustworthy	3.88	4.29	4.16
Difficult to understand/Easy to understand	6.00	5.82	5.88

Too general, generic/Too specific, detailed	3.12	3.62	3.45
Strict, bound to rules/Creative	3.88	3.65	3.73
Oriented towards assessment/ Oriented towards improvement	4.00	4.59	4.39
Not explained enough/Very well explained	4.71	4.74	4.73

Note. Numbers in the cells represent mean values in each group. Question wording: "On a scale of 1-7, how would you rate the results that you received from ChatGPT in this exercise with regard to the following dimensions?" Answer options (e.g., 1 = 'Not helpful at all'; 7 = 'Very helpful') were presented on a semantic differential scale.

Many students described a trial-and-error learning process in how to prompt the chatbot. The coding theme *learning experience*, where any sentence that explicitly mentioned "learning" or substantially concerned skill or knowledge acquisition was included, typically unfolded as narratives about the students' interactions with the AI. While they reported that precision in prompts gave better results, many found this process frustrating. When just asking for 'feedback', ChatGPT did not provide results perceived as particularly useful. Several students mentioned that finding out how to pose questions was the biggest challenge (cf. Jacobsen & Weber, 2023). A few noted that the feedback provided by the bot was helpful for considering how to edit their own text, with one writing that "as long as I use it to think about how to change the text myself, instead of just having it do the assignment for me, I feel like I've learned something". One student wrote that "It forces you to critically assess its replies, since they're sometimes super good, and sometimes totally wrong". The results from the survey point in a similar direction: When asked about how they would rate the bot's suggestions for revisions, the majority (60.8%) of the students said it was possible to use them with some minor changes, and an additional 33.3% said the revisions could be used but required major changes. These results suggest that the chatbot interaction forced the students to draw on their feedback literacy in a more direct manner, specifically their judgment in assessing the replies.

One content theme that emerged from the data set was *other uses for ChatGPT*: 'Idea generation' and 'text summary generation' were most frequently discussed as better uses for the tool than feedback on writing. Students from both subjects reflected on how they thought

it could be useful for finding out what to do and be inspired in coursework, and some students explicitly delimited it to the first stage of a project - for example when beginning to formulate a research question. One student wrote “Normally, I have only used ChatGPT for ‘discussion’ - I ask it about pros and drawbacks of some subjects, that I can’t make up my mind about, to get input and a wider perspective on the subject”. Another student appreciatively wrote that it “summarizes what your assignment is about and delivers analytical observations”, and another remarked that “the AI is very helpful in terms of getting an overview and being pointed in the right direction”. A rhetoric student said in plenary that it was “good for navigating your own stream of thoughts; I asked it if a text I wrote is meaningful”. This would be to use the summarizing and paraphrasing capabilities of ChatGPT to distance oneself and gain perspective on a text one has written.

These uses could be included under feedback as an umbrella term, like Møgelvang et al. (2023) do. Notably, they may operate on the process and self-regulation levels of feedback. It is however important to note that our students did not perceive them as feedback. In their vocabulary, “feedback” seemed to be associated with later stages of a writing process and the task level. While this may be attributable to the format of and limited introductions to the exercise, it is indicative of how the students perceived the interaction for the specific task. Some students mentioned that it was a time-consuming method of getting feedback that would come faster and more easily from a teacher or fellow student. As we will show below, the different and unfamiliar role of feedback-receiver, assuming some responsibilities that are typically a part of feedback giving, lies at the heart of many of our students’ reflections on the feedback situation.

Critical evaluation of generated feedback and comparisons to peer- and teacher feedback

One of our key findings is that the critical labour in generating AI feedback is differently distributed and of a different kind than would be the case in human feedback situations. Prompting by the feedback receiver takes over the role of informed reflection from a feedback giver. Most of the students pointed out that replies by the chatbot had to be critically evaluated. *Trust* emerged as a very prominent theme in the material, defined as any sentence that concerned the trustworthiness of the AI, or any sentence that expressed trust or distrust of the AI. Likewise, among the codes for the theme *emotional expression*, scepticism (understood as any expression of hesitancy about the quality of result or process) stood out as an often-expressed attitude. While this was to be somewhat expected given the public discourse on AI

and falsity, and given that we mentioned possible "hallucinations" (i.e. generated answers with no basis in fact) in our brief introductions which could have had a priming effect, it is clear from our material that students found issues of trust and confidence to permeate the experience. Also, among sentences coded as concerning *demands placed on the user*, only a handful concerned issues of the interface and the learning curve of generative AI, and reflections on demands on the *faculties of users* were more prominent. The more detailed reflections show that our students were also cautious about their own prompting, indicating that they experience an unfamiliar demand on their own critical labour in feedback situations. In a way these reflections indicate that students were confronted with their own feedback literacy in an unusual manner. The students' replies hint at a sense of themselves being the judge of disciplinarity in the AI feedback situation because they had to assume responsibility to both request and evaluate feedback. In contrast, peer feedback situations were typically described as dialogical and teacher feedback as authoritative. As one student put it: "With teacher feedback, I know it is right".

One common expression of this unfamiliar division of critical labour relating to prompting is captured here:

> "One needs to be quite specific in their queries. This is something we, of course, also need to consider when providing feedback in pairs/groups. [If I am not specific enough, it prompts me to clarify my question.] The good thing about this is that it helps me become better at specifying my questions and writing clearly [...]. The downside is that I spend a lot of time trying to find concrete and precise formulations, time that I am already lacking."

While this student does mention that specificity in queries is desirable in peer feedback, he/she attributes it specifically as an aspect of prompting. 21 students in the survey made some point about how precision in queries was necessary. In general, they found some frustration in the demand it placed on them, perceived it as a laborious task compared to peer feedback, and attributed this to the chatbot lacking the contextual knowledge of a peer or teacher. One student remarked that "I think it depends a lot on how good I am at asking, which becomes problematic in a feedback situation, as I have to keep everything I want in mind myself". This expresses an implicit notion that feedback is partially supposed to find out what one does not know yet, that is, gaining new perspectives on one's own ideas. One student wrote, on a more positive note, that "I believe it makes you reflect on and practice asking good, precise, and concrete questions, which you can use when engaging in a feedback situation with people." We would suggest that the students found themselves in a sort of ambiguous critically

evaluative position: Not only did the situation warrant a critical evaluation of the replies, and how these reflect the text students prompted the chatbot for feedback on, but it also forces a critical evaluative perspective on the students' own queries. One student remarked that "Can I trust the answer? Have I posed a correct question that it understands?".

The artificial nature of the AI played a substantive role in the critical evaluation of its responses. Many sentences coded as trust-related coincided with codes for *comparison to human feedback givers*. Despite several students commending its mimicry of a human interlocutor, many of them found it lacking regarding collaborative thinking-together in feedback. Students contrast it to the "teamplay in feedback face to face" and argue that "I can't spar with it in the same way, so that we together can find out what is meaningful to look at". In this regard, the fidelity of its mimicry of human interaction seemed to be a key issue, specifically regarding creative interaction. A handful of students point to interactions with ChatGPT that built trust in the quality of the results. For example, they reported that the bot provided similar feedback to what they had already received on the assignment from their teacher, concluding that "well, that made me trust it more". Another student wrote that "It resembled the feedback we had already received - or [...] gave us thoughts we had already considered ourselves. Provided a few new comments we haven't heard before" indicating a more negative evaluation of the same dynamic. The evaluations of the feedback therefore seem based on whether the chatbot would confirm things about the students' assignments, or the students' prior knowledge of the assignment/field, and implies a cognitivist focus on the content and delivery of feedback. But there was also a clear wish for novelty in feedback that the students found lacking in the AI.

One key point for the students' critical evaluation was that ChatGPT lacks the contextual knowledge that peers and teachers would draw on in feedback. Students pointed out how both the assignment and the entirety of the course and program as context were unknown to the AI. One student wrote that "everything that cannot be found in the 'input' it has been given is often imprecise or inaccurate". Another wrote that "When I get feedback from real people (teachers, buddies, groups) all of them know the text, the assignment, and possible limitations/criteria. It makes it, everything else equal, easier to ask for feedback on (specific) areas, without being overly specific in your questions". Another student pointed to this dynamic in terms of trust in the results: "It was interesting to see how I could get help evaluating my assignment - although I do not have confidence that the same points of criticism would be highlighted by my instructor." Recognizing this lack in the AI allowed the students to critically evaluate its replies. The students read the AI feedback within a contextual framework, the course and assignment, and used this knowledge to judge the usefulness of the results. This also explains why many of

the students found ChatGPT's feedback inefficient as it requires explication of established and perhaps tacit knowledge that can be expected in other feedback situations.

Overall, these replies show that the chatbot was not deemed a suitable replacement for other feedback forms. In line with this, students in both groups reported that the results they received from ChatGPT could not live up to their teacher's or peer feedback, although we found some interesting differences in the survey: Students in the F&M class rated the AI's output significantly higher (M = 3.91, SD = .996) in comparison to peer feedback than their colleagues in the Rhetoric program (M = 3.12, SD = 1.111; F(1, 49) = [6.671], p = .013), though still below the middle scale point (range 1 = much worse than peer feedback to 7 = much better than peer feedback). When asked to compare ChatGPT's revisions to their respective teacher's feedback, both groups reported similarly low values (Rhetoric M = 2.41, SD = .939; F&M M = 2.82, SD = 1.24; range 1 = much worse than teacher feedback to 7 = much better than teacher feedback).

Attitudes and emotions towards ChatGPT as feedback giver

The interactional behaviour of ChatGPT played a significant role in the students' reported experiences, and its politeness was often mentioned as a key concretization of the theme *AI characteristics*, i.e. any sentence that concerned a description and/or judgment of interactional behaviour of ChatGPT. One student connected this to norms for feedback: "Its tone of feedback has been exemplary, and it has attained all the demands of clear, concrete and kindly feedback". These terms (Da: klar, konkret & kærlig) are often used in Danish teaching as hallmarks of good feedback giving, and several students used them as labels in their responses. One wrote that "When I asked GPT to provide feedback with the 3 Ks in mind, it knew what to do as well." Other students mentioned that its formal and correct "tone", including its linguistically correct and clear responses, give an impression of competence. In so far, the students seemed to acknowledge that ChatGPT can provide feedback in a way that formally mimics a "good" feedback giver.

However, this politeness is also related to a sense of distrust. A handful of students reported trying to get it to give "critical" feedback, only succeeding after several attempts and revised prompting. Students reported having to ask it to be "meaner" to get more constructive and useful feedback. One student wrote that "it is good at giving positive feedback, but in a form where it just mentions something that you have written, quote it indirectly and says that that was good. For example, you could write something utterly basic, and it would say that it was

good that it was basic". Others noted that "It's purely objective and very soft, it won't give any interesting answer but only strict, straight and precise answers", or that "I'm left with the impression that ChatGPT is a bit of a cheerleader when it comes to feedback. Either that, or I have written a genius assignment, which is hardly the case." These students indicate that critical feedback, or lack thereof, is related to their trust in and the perceived quality of the feedback they receive (cf. Bader et al., 2019).

ChatGPT's feedback style seemed to divert from expectations about how feedback should be phrased to the extent that it was not being taken seriously by some of the students. They mentioned the formal and correct tone as cause for suspicion, especially in relation to banal or incorrect replies. One student reflected on this:

> "It's impressive how well-articulated the AI actually is, but at the same time, it quickly reveals its limitations as a feedback provider. For instance, it is very reluctant to give criticism, making its feedback somewhat toothless, bordering on useless. Additionally, it 'lies'. I didn't provide any criteria for the assignment, but it still claimed that the task was a satisfactory answer 'based on the given criteria' or something similar. Hardly an objective feedback provider."

In general, the students reported feeling quite ambivalent about the kind and competent tone of the chatbot. While it was often appreciated, it was also cause for suspicion or unease. Across the students' replies, kindness was most often featured as surprising and worthy of mention, regardless of evaluation. If students had expected that the AI would respond as a feedback giver in an impersonal, "robotic" and correct manner, the kind and "personable" tone would confuse these expectations.

A handful of students expressed that they found ChatGPT to mitigate some of the negative feelings typically associated with other feedback formats. One student wrote that "Above all else, I do know that the bot has ACTUALLY read and analysed my text thoroughly", indicating a certain distrust in feedback preparations among teachers and peers. A similar thought was expressed by a student reflecting on how asking for feedback from stressed peers who are struggling with their own assignments feels like being a burden, unlike using ChatGPT. Three students reported that the mechanical nature of ChatGPT helps them to separate themselves from the text. They specifically pointed to not comparing themselves to the feedback giver and mention that they do not take the AI feedback personally. One of them wrote that:

> "It is more mechanical, as it can see the text as a produced piece of writing and not as an extension of you. At the same time, it also makes it easier for the producer of the text not to take it 'as

> an attack' when feedback is given to it because I know that it is not an attack on me but feedback on the task."

These reflections are a reminder of how feedback situations can cause anxiety. The psychological attachment to written products is a pedagogical problem discussed as early as roman antiquity: "Of course, we love all our own productions when they are newly born; if we did not, they would not even be written down" (Quintilian, 2002, X.3.7). Students point to an interesting dynamic where detachment from the text becomes easier and reasonable due to the artificial and detached feedback-giver – which bypasses the social constituents of affect. In relation to affect management as a dimension of feedback literacy, this is not necessarily a means of facilitation that is transferrable to other feedback situations. Nevertheless, this arguably beneficial affective response to AI feedback indicates a possible supplementary role in higher education, for example as a low affect contrastive situation to peer and teacher feedback.

Discussion and concluding remarks

While all students participating in our exercise reported on it as a positive experience, they were decidedly critical about using ChatGPT as a feedback tool. They were generally suspicious of the feedback it generated, judged it significantly lower than peer- and teacher feedback, found it time-consuming and inefficient, and indicated that it placed the burden of disciplinary judgment solely on the user, as opposed to the shared judgment in peer feedback. This is in part attributable to the fact that task-level feedback guided their interactions. While it is possible that this focus is partially a product of the study design, framing, and the students' prior understandings of feedback, the affordances of ChatGPT should not be disregarded either. The interface of ChatGPT seems to invite task-oriented prompting, and its lack of social knowledge of the student and access to study environment context would seem to render any self-, process- or self-regulatory level feedback notably artificial or generic. In terms of Hattie and Timperley's three feedback questions (2007), ChatGPT seems to respond well to "how am I going"-questions (in relation to general norms in its training data), but unable to relate to "where am I going"-questions. It is particularly hard to imagine the AI giving a well-grounded response to "where to next?", given how this requires knowledge of both the student and the discipline.

The chatbot interface also gave rise to emotionally ambivalent experiences: While its mimicry of a human interlocutor was commended as good, it was found lacking in expressing realistic

affective responses to a text and limited in its capacity for creative brainstorming. Some students also found that the polite and positive tone created distrust through the absence of "sharp" criticism. These findings confirm and add some explanation to the previous research of Kim et al. (2020) and Bilquise et al. (2023). We would assume that these negative assessments would be mitigated by training and habituation in using AI chatbots. A short introduction like the one we gave in our courses is insufficient for a full contextualization, especially for students without prior experience in AI use. It is important that both teachers and students are aware of the shortcomings of artificial intelligence and understand its limitations (Steiss et al., 2023), and training should at least comprise didactical prompting strategies ('promptology', see e.g., Dobson, 2023). Nevertheless, the interactional dynamics we have uncovered are important since they may be present but obfuscated in situations where AI is more naturalized.

We find that a key difference in perceptions of AI- and human feedback lies in a different role distribution. Human feedback was viewed as originating in the givers' reception, knowledge, and response to the text. In contrast, AI feedback shifts the dynamic of the process toward the receiver, as it is an individual activity (Møgelvang et al., 2023). Our students experienced a decidedly higher demand on their own judgment, both in making fruitful prompts and critically evaluating the results according to disciplinarity and course content. This different role distribution was cause for both critical evaluation, doubt, and frustration. We argue that this different dynamic is key for how to implement AI as a complement to peer- and teacher feedback: First, it demands careful consideration of when and for what types of texts it should be used. Second, it points to the importance of aiding students in attaining knowledge of what criteria and themes are crucial in a specific task in order to help them improve their own prompting. In other words, students need to know what to look and ask for in the feedback situation – otherwise, the results they receive from ChatGPT are perceived as too general and not particularly useful.

In line with this, some students reflected on the study participation itself as a learning experience: They emphasized how it was a practice in posing good, correct questions and what to look for in an assignment, evaluating replies, and pointed out how this was transferrable to other learning activities. The students noted that the quality of replies is connected to both abstract knowledge of and a metalanguage for the elements of a particular type of academic text. In so far, AI-based feedback has a potential use for training student metaknowledge about the constituents of specific academic tasks and genres. If framed in the correct way, it could be a means of actively training feedback literacy, specifically in the dimension of judgment,

appreciation and acting. That is, it can be positive for facilitating reflectivity in learning, which has been noted as one principle of good feedback practice (Nicol & Macfarlane-Dick, 2004).

Keeping in mind the importance of timeliness and time constraints in feedback (Steiss et al., 2023), applying AI feedback at suitable points in a learning process may complement peer- and teacher feedback and allow for a more expedient use of time. Our students reported considering ChatGPT to be a better sparring partner for early stages of a task, such as when trying out ideas or finding the key elements in early drafts. They also found the chatbot useful for concrete language feedback, mentioning grammar, syntax, and diction. The direct response of ChatGPT's interface seems to invite a task level focus in prompting. AI could therefore potentially be used to discover and improve students' language habits. Taking over this part of text feedback from teachers and peers can free up time to focus on other, more disciplinary aspects of the assignment or other levels of feedback. We would here add that such an exercise should be framed precisely as feedback on writing: If the AI is used for automated corrections, it could be detrimental to acquiring language skills – however, similar reservations can be made for all auto-correcting programs, including Microsoft Word.

A handful of students expressed that the AI feedback situation mitigated negative feelings like social anxiety and helped them treat the text as an object rather than an extension of themselves. Students have been shown to respond emotionally negatively to feedback reflecting hierarchy and perceived disrespect (Taggart & Laughlin, 2017), and ChatGPT does seem to provide some of the same benefits as anonymized computer mediated feedback (Savignon & Roithmeier, 2004; Yu & Lee, 2016). But there are significant differences in the orientation to feedback among students (Yang et al., 2023). Some of our students mentioned the lack of negative comments as a reason for distrust, which could indicate that they expected similarly negative comments from the AI as from a tutor or high performing peer (Hamer et al., 2015). Getting feedback from an AI could render any perception of hierarchy or disrespect negligible for the feedback receiver: it is literally a machine providing feedback. Our results further indicate that the human-machine interaction logic could also redirect a students' feedback orientation in a more instrumental and active way, for example focusing on posing purposeful questions about the text. Further study is needed to understand the affective difference between human- and AI feedback situations, but our results indicate that these orientational dynamics may be relevant to AI feedback implementation. One way of construing our results is that the AI made the social constructivist learning aspects of feedback interaction more salient than the specific inputs that the students sought. On the one hand, the *social* aspects were so artificial that some of the potentially negative effects of student sociality were

mitigated, such as social anxiety. On the other hand, the students missed some positive effects such as creative sparring and disciplinary dialogue.

AI chatbots come with several challenges for use in higher education, and it is critical that educators and policy makers consider their environmental impact and company's questionable business models as well (cf. Driessens & Pischetola, 2024). While we show that chatbots have the potential to supplement human feedback in productive ways, the AI feedback situation differs notably from its human counterpart in terms of interactional dynamics, affect and implications for students' feedback orientation. These differences need to be accounted for in how AI is implemented and introduced to students to ensure that it is used in fruitful and ethical ways.

References

Bader, M., Burner, T., Iversen, S. H., & Varga, Z. (2019). Student perspectives on formative feedback as part of writing portfolios. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2018.1564811

Bilquise, G., Ibrahim, S., & Salhieh, S. E. M. (2023). Investigating student acceptance of an academic advising chatbot in higher education institutions. *Education and Information Technologies*, 1-26. https://doi.org/10.1007/s10639-023-12076-x

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*(3), 260-294. https://doi.org/10.1177/0741088306289261

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2190148

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y., Gasevic, D., & Chen, G. (2023, April 13). *Can large language models provide feedback to students? A case study on ChatGPT*. Working paper. Available at https://doi.org/10.35542/osf.io/hcgzj

Dobson, J. E. (2023). On reading and interpreting black box deep neural networks. *International Journal of Digital Humanities*. https://doi.org/10.1007/s42803-023-00075-w

Driessens, O., & Pischetola, M. (2024). Danish university policies on generative AI: Problems, assumptions and sustainability blind spots. *MedieKultur: Journal of media and communication research*, *40*(76), 31-52. https://doi.org/10.7146/mk.v40i76.143595

Evans, C. (2015). Students' perspectives on the role of peer feedback in supporting learning. *Journal of Cognitive Education and Psychology, 14*(1), 110-125. http://dx.doi.org/10.1891/1945-8959.14.1.110

Garg, S. (2023, August 20). 30 Best ChatGPT alternatives in 2023 (Free & Paid). *Writesonic*. Available at https://writesonic.com/blog/chatgpt-alternatives

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523-547. https://doi.org/10.1086/681947

Hadden, A. A., & Frisby, B. N. (2019). Face threat mitigation in feedback: An examination of student feedback anxiety, self-efficacy, and perceived emotional support. *Communication Quarterly, 67*(1), 60-75. https://doi.org/10.1080/01463373.2018.1531043

Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education, 40*(1), 151-164. https://doi.org/10.1080/02602938.2014.893418

Hasse, C., & Bruun, M. H. (2023). ChatGPT minder os om det gode ved uvidenhed. *Videnskap.dk*. https://videnskab.dk/teknologi/chatgpt-minder-os-om-det-gode-ved-uvidenhed/

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Holton, J. A. (2007). The coding process and its challenges. In A. Bryant & K. Charmaz (Eds.), *The SAGE Handbook of Grounded Theory* (pp. 265–289). SAGE Publications. https://doi.org/10.4135/9781848607941.n13

Jacobsen, L. J., & Weber, K. E. (2023, September 29). *The promises and pitfalls of ChatGPT as a feedback provider in higher education: an exploratory study of prompt engineering and the quality of AI-driven feedback*. Working paper. Available at https://doi.org/10.31219/osf.io/cr257

Kelly, A., Sullivan, M., & Strampel, K. (2023). Generative Artificial Intelligence: university student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching and Learning Practice, 20*(6), 12. https://doi.org/10.53761/1.20.6.12

Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics, 77*. https://doi.org/10.1016/j.tele.2022.101925

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2020). My teacher is a machine: Understanding students' perceptions of AI teaching assistants in online education. *International Journal of Human–Computer Interaction, 36*(20), 1902-1911. https://doi.org/10.1080/10447318.2020.1801227

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2021). I like my relational machine teacher: An AI instructor's communication styles and social presence in online education. *International Journal of Human–Computer Interaction, 37*(18), 1760-1770. https://doi.org/10.1080/10447318.2021.1908671

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625-636. https://doi.org/10.1089/cyber.2018.0110

Mollick, E. R., & Mollick, L. (2023). *Assigning AI: seven approaches for students, with prompts*. Available at SSRN: https://ssrn.com/abstract=4475995

Møgelvang, A., Ludvigsen, K., Bjelland, C., & Schei, O. M. (2023). HVL-studenters bruk og oppfatninger av KI-chatboter i utdanning. *HVL-Rapport*. Available at https://hvlopen.brage.unit.no/hvlopen-xmlui/handle/11250/3100624

Nicol, D., & Macfarlane-Dick, D. (2004). Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice. In C. Juwah, D. Macfarlane-Dick, B. Matthew, D. Nicol, D. & Smith, B. (Eds). *Enhancing student learning though effective formative feedback*. The Higher Education Academy.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253-1278. https://doi.org/10.1080/02602938.2019.1600186

Quintilianus. (2002). Institutio oratoria (D. A. Russell, Trans.). In *Loeb Classical Library 127*. Harvard University Press.

Savignon, S. J., & Roithmeier, W. (2004). Computer-mediated communication: Texts and strategies. *Calico Journal*, 265-290. https://www.jstor.org/stable/24149395

Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., ... & Warschauer, M. (2023, September 5). Comparing the quality of human and ChatGPT feedback on students' writing. *Working paper*. Available at https://doi.org/10.35542/osf.io/ty3em

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language learning, 52*(1), 119-158. https://doi.org/10.1111/1467-9922.00179

Taggart, A. R., & Laughlin, M. (2017). Affect matters: when writing feedback leads to negative feeling. *International Journal for the Scholarship of Teaching and Learning, 11*(2), n2. https://doi.org/10.20429/ijsotl.2017.110213

Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review, 9*, 1-15. https://doi.org/10.1016/j.edurev.2012.11.004

Winstone, N. E., & Carless, D. (2020). *Designing effective feedback processes in higher education: a learning-focused approach* (1st ed.). Routledge.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yang, L., Wu, Y., Liang, Y., & Yang, M. (2023). Unpacking the complexities of emotional responses to external feedback, internal feedback orientation and emotion regulation in higher education: a qualitative exploration. *Systems, 11*(6), 315. https://doi.org/10.3390/systems11060315

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*, 477-501. https://link.springer.com/article/10.1023/A:1023967026413

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching, 49*(4), 461-493. https://doi.org/10.1017/S0261444816000161

Zilber, A. (2023, August 14). College profs to use paper exams to thwart cheating with ChatGPT: 'We're in full-on crisis mode'. *New York Post*. Available at https://nypost.com/2023/08/14/college-profs-to-use-paper-exams-to-prevent-cheating-with-chatgpt/

Authors

Tommy Bruhn

Assistant Professor

Section of Rhetoric, Department of Communication, University of Copenhagen

https://orcid.org/0000-0003-3131-0856

Franziska Marquart

Assistant Professor

Section for Media Studies, Department of Communication, University of Copenhagen

https://orcid.org/0000-0002-8800-9303

Student experiences of ChatGPT as a complementary feedback tool in higher education

Appendix

A1 – Survey questionnaire

Dear all,

In this survey, we ask you about your experience with using ChatGPT inside and outside of the classroom. Please take a few minutes to answer the questions on the next pages, irrespective of whether you participated in the specific class exercise about *feedback* or not. Please also pay attention to the open-ended questions and share your thoughts and reflections with us. We developed this exercise as part of a project in our *Universitetspædagogikum* at KU, and your responses will help us in better understanding how generative AI tools can be used in higher education in the Humanities.

There are no 'right' or 'wrong' answers to the questions; instead, we are interested in your opinions and experiences. Your answers will not be graded; they are anonymous and cannot be connected to you personally, and we will aggregate them before sharing them with other scholars.

If you have additional comments or suggestions, please let us know!

---

1. Did you participate in the ChatGPT exercise and use the chatbot to generate feedback for your own work?
   [Yes/No]

2. On a scale of 1-7, how helpful was your teacher's introduction to this exercise?
   [1 Not helpful at all – 7 Very Helpful]

3. How much experience did you have with ChatGPT before we talked about it in the classroom?
   [1 - Had never heard of it; 2 - Had heard of it, but never tried it; 3 - Tried it once or twice; 4 - Used it a few times (3-6 times); 5 - Used it often (more than 6 times)]

4. Have you ever used ChatGPT or a similar technology for school/university before (e.g., for assignments or exercises, both inside and outside of the classroom)?
   [1 - No, I have not; 2 - Yes, I have]

   4a) You said that you used ChatGPT or a similar technology for school or university before. Was that your own initiative or introduced by a teacher/tutor?
   [1 - My own initiative; 2 - Initiated by a teacher or tutor; 3 - Both my own and a teacher's/tutor's initiative (on different occasions)]

   4b) If you have used another technology that is similar to ChatGPT, can you tell us what it was?
   [Open ended]

5. On a scale of 1-7, how did you experience working with ChatGPT in the exercise this week?
   [1 As very negative – 4 As neither negative nor positive – 7 As very positive]

6. On a scale of 1-7, how would you rate the results that you received from ChatGPT in this exercise with regard to the following dimensions?
   [1 Not helpful at all – 7 Very helpful]
   [1 Not trustworthy at all – Very trustworthy]
   [1 Difficult to understand – 7 Easy to understand]
   [1 Too general/generic – 7 Too specific/detailed]
   [1 Strict/bound to rules – 7 Creative]
   [1 Oriented towards assessment – 7 Oriented towards improvement]
   [1 Not explained enough – 7 Very well explained]
   [1 Much worse than peer feedback – Much better than peer feedback]
   [1 Much worse than teacher feedback – 7 Much better than teacher feedback]

7. If you have additional comments on the results that you received from ChatGPT, please add them here.
   [Open ended]

8. Did you think that the bot was particularly good and/or bad at some types of queries/feedback?
   [Open ended]

9. If you asked the bot to provide suggestions for revisions, how would you rate them?
   [1 – Possible to use as is, by copying them into my document; 2 – Possible to use with some minor changes; 3 – Possible to use, but with some major changes; 4 – Not useful]

10. On a scale of 1-7, how would you rate ChatGPT's Danish language proficiency?
    [1 Very poor – 4 Neither poor nor very good – 7 Very good]

11. How did you personally feel about using ChatGPT in this way?
    [Open ended]

12. To what extent do you agree or disagree with the following statement: Interacting with ChatGPT for feedback was a valuable learning experience for my own academic writing skills.
    [1 Completely disagree – 4 Neither disagree nor agree – 7 Completely agree]

13. If you have comments about the interaction as a learning experience, please add them here.
    [Open ended]

14. How likely is it that you would use ChatGPT for the following purposes in the future:
    [1 – Very unlikely – 4 Neither unlikely nor likely – 7 Very likely]

    a. Feedback on your writing
    b. Other study-related tasks
    c. Other tasks outside of the study context

15. University students may use ChatGPT for all kinds of tasks. What suggestions or tips do you have for others who use the AI to get feedback?
    [Open ended]

Lastly, we would like to know a few more things about yourself so that we can better understand how useful classroom discussions and exercises on artificial intelligence are for you and other students. Remember that none of the information that we collect can be connected to you individually, and that your answers remain completely anonymous.

16. What gender do you identify with?
    [Woman; Man; Transgender; Non-binary/non-conforming; Other; Prefer not to answer]

17. Generally speaking, how interested would you say you are in trying new technologies?
    [1 Completely uninterested – 7 Very interested]

18. How old are you?
    [Open ended]

19. Who is your teacher in this course?