



Intelligent Support Engages Writers Through Relevant Cognitive Processes

Andreas Göldi
andreas.goeldi@unisg.ch
University of St.Gallen
St.Gallen, Switzerland

Seyed Parsa Neshaei
seyed.neshaei@epfl.ch
EPFL
Lausanne, Switzerland

Thiemo Wambsganss
thiemo.wambsganss@bfh.ch
Bern University of Applied Sciences
Bern, Switzerland

Roman Rietsche
roman.rietsche@unisg.ch
University of St. Gallen, Institut Digital Technology
Management
Bern University of Applied Sciences
Bern, Switzerland

ABSTRACT

Student peer review writing is prevalent and important in education for fostering critical thinking and learning motivation. However, it often entails challenges such as high effort and writer's block. Leaving students unsupported may thus diminish the efficacy of the process. Large Language Models (LLMs) offer a potential remedy, but their utility hinges on user-centered design. Guided by design-determining constructs from the Cognitive Process Theory of Writing, we developed an intelligent writing support tool to alleviate these challenges, aiding 1) ideation and 2) evaluation. A randomized experiment (n=120) confirmed users were less inclined to utilize the tool's intelligent features when offered pre-supplied ideas or evaluations, validating our approach. Moreover, students engaged not less but more with their writing if support was available, indicating an enhanced experience. Our research illuminates design choices for enhancing LLM-based tools' usability and user experience, specifically optimizing intelligent writing support tools to facilitate student peer review.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; **Empirical studies in interaction design**.

KEYWORDS

Creativity Support, Education/Learning, Schools/Educational Setting, Artifact or System

ACM Reference Format:

Andreas Göldi, Thiemo Wambsganss, Seyed Parsa Neshaei, and Roman Rietsche. 2024. Intelligent Support Engages Writers Through Relevant Cognitive Processes. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3613904.3642549>



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '24, May 11–16, 2024, Honolulu, HI, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0330-0/24/05
<https://doi.org/10.1145/3613904.3642549>

1 INTRODUCTION

Student peer reviews are becoming more common in online education, especially in large massive open online courses (MOOCs) on platforms like Coursera, where 37% of courses now use this approach [36]. However, students often find participation tough, mainly because of its high effort and writer's block, which is a common issue [1]. Since MOOCs had 220 million students in 2021, there is a growing need for better tools to help students in the peer review process [42].

The latest developments in Large Language Models (LLMs) could be used to improve these tools. However, using them effectively requires careful planning and design [19, 49]. This is because the multitude of functions can overwhelm users and their benefits outweigh the disadvantages of increased cognitive load. Therefore, the LLM-based tools need to be designed in a user-centric way to support the user while generating and rewriting text [14, 27].

That is why we chose to base our design on the Cognitive Process Theory of Writing [16]. This theory enables a user-centered design by focusing on the interface's ease of use and helpfulness to ensure it benefits students during the peer review process [24]. In particular, using LLMs adds advanced features like targeted feedback to writing tools beyond just catching errors [14]. However, these features are not automatically useful, especially for student peer reviews [1, 42]. Leveraging constructs from the Cognitive Process Theory of Writing enables us to design intelligent writing support tools that align closely with the cognitive steps that users actually go through when writing. This helps limit the feature set to those essential for facilitating specific cognitive functions like *ideation* and *evaluation* [19]. Consequently, our approach does not just make these tools more usable but in particular, serves its intended purpose of genuinely aiding the writing process, fulfilling the user's primary needs and expectations [24].

To better target the specific challenges students face in peer review writing within online courses, we formulate the following research question:

How does intelligent writing support, guided by the Cognitive Process Theory of Writing, influence the cognitive processes of ideation and evaluation in student peer review writing?

In this study, we developed a system harnessing the capabilities of OpenAI's GPT-3.5-Turbo model to provide intelligent writing support targeting two specific cognitive processes: ideation and evaluation. Our choice of focus aligned with existing literature on feedback and creativity support tools [17, 35, 39, 50, 53]. We evaluated our tool through a fully randomized experiment involving 120 users. The results demonstrated that users were less inclined to engage with the intelligent features when pre-supplied ideas or evaluations were available, thereby validating our approach [16].

Our work contributes in three main ways. First, it corroborates the utility of design-determining constructs from the Cognitive Process Theory of Writing in shaping intelligent writing support tools [16, 24]. Second, it offers empirical evidence suggesting that the presence of intelligent writing support can positively impact time spent on writing tasks and user engagement [43]. Lastly, our research provides practical insights into the design considerations for enhancing usability and user experience, setting the stage for more nuanced studies in controlled settings [24].

2 BACKGROUND AND RELATED WORK

2.1 Writing Support Tools

Writing is a multifaceted skill, crucial for various activities such as learning, practicing, and teaching [21, 23]. Given the cognitive demands of writing, which include memory capacity [16], effective utilization of cognitive resources is key for sustained performance [15]. The need for effective writing practices has been emphasized by the pervasive issue of writer's block, i.e., the inability to come up with or implement new ideas for continuing writing [40].

Traditional writing support tools were limited in scope, primarily focusing on spelling, grammar, and style [14]. Their development has been influenced by a long history of research and a focus on a collaborative relationship between the tool and the user [3, 32]. These tools gradually evolved to include more complex tasks such as paraphrasing and offering feedback [2]. Despite the advancements, these tools were limited in their ability to address varied writing processes crucial for effective writing [22, 45].

2.2 Large Language Models in Writing Support Tools

The advent of LLMs marks a turning point in the development of intelligent writing support tools. Advances in model architectures like Transformers have dramatically improved the capabilities of LLMs [7, 8, 46]. This technological leap has broadened the scope of applications for writing support tools, making them even more "intelligent" [9].

LLMs like the GPT-3.5 Turbo model we use in our research [7] have been employed across various domains [31, 38], from creative writing to language learning [12, 26, 30, 33, 51]. The main advantage of tools incorporating LLMs over their predecessors is the added intelligence that enables them to assist in ideation, elaboration, and even the creative process [13, 26].

Modern LLMs offer new features that build upon traditional spell and grammar-checking capabilities. They have evolved to perform tasks like paraphrasing, continuation, and elaboration [2, 13]. Additionally, LLMs are capable of providing adaptive feedback [47],

offering creative support [17], and potentially equalizing outcomes for groups that have traditionally faced writing challenges [6].

However, there are challenges to surmount. Issues such as inconsistent style, unsatisfactory text generation, and the need for human intervention remain [26, 44]. However, Gero et al. [20] emphasize that while LLMs show promise, they inherit significant challenges such as generating misleading information and exhibiting difficulty in user-driven steering, underscoring the need for nuanced understanding and careful application in writing support tools. Furthermore, integrating intelligent features like ideation and elaboration into tools requires careful consideration of cognitive processes and existing literature [14, 18]. If we can find a way to overcome these challenges, incorporating LLMs in writing support tools promises a more sophisticated, adaptive, and intelligent future for assisting writers.

2.3 Application in Specific Context: Student Peer Review Writing

The principles underlying our intelligent writing support tool, guided by the Cognitive Process Theory of Writing, are universally applicable across various writing domains. However, the specific focus on student peer review writing in our study serves as a practical example to illustrate these principles in action. This context provides a unique setting to explore and address common writing challenges, such as high effort and writer's block, which are prevalent in peer review scenarios.

In emphasizing student peer review, we aim to demonstrate the effectiveness of our tool in a concrete scenario, while acknowledging its potential applicability in broader contexts. The focus on the writing component of peer review is particularly critical, as it often represents a significant barrier to effective peer review. By addressing this key aspect, our study not only contributes to the field of peer review writing but also sets a foundation for applying these insights to other writing domains, leveraging the general scope of the Cognitive Process Theory of Writing.

2.4 Cognitive Process Theory of Writing

We turn to writing theory to overcome the challenges in exploiting the powerful capabilities of LLMs. Only by doing so can we match the design of our support tools to the actual processes they are striving to assist in. Different from other influential theories on writing [23], Cognitive Process Theory [16] highlights not processes that influence the writer but processes in the writer's mind itself. The theory describes how writing is achieved. Any writing activity is based on three *priors*: 1. the writer's long-term memory, which includes information on the topic, audience, and reason to write; 2. the text produced so far; and 3. the rhetorical problem currently to be solved, i.e., what has to be achieved to address the reason to write.

These priors inform three distinct subprocesses: 1. *planning*, which includes the development of concrete writing goals, ideation on what writing content can address them, and how it should be organized; 2. *translating*, sometimes called transcribing, which is the act of finding formulations for the content and writing them down; 3. *reviewing*, which consists of evaluating and revising text

produced so far. The writing process cognitively consists of dynamically switching between these three subprocesses. An additional subprocess, namely monitoring, conducts switching.

These processes can be grouped into two distinct classes: 1) gathering new information, and 2) applying the information to produce observable results. In the context of the Cognitive Writing Process Theory, these classes are called *exploration* and *exploitation*, respectively [29]. An overview of the theory and how it is applied to the design reported in this paper can be seen in Figure 1.

Paying attention to these naturally occurring cognitive processes can help improve beneficial and preventing detrimental effects of writing support tools. For example, review writers sometimes conflate generated suggestions with the text written so far [27]. This can lead to the introduction of opinions not aligned with those of the writer and factual inaccuracies in the final text. Additionally, it is important to avoid distracting writers with suggestions at the wrong moment, as this may impact writing quality: Since pre-writing pauses are a good predictor for writing quality [4]. This is because, in these pauses, planning is allowed to occur. More generally, the theory has been applied in a recent review classifying writing tools based on the cognitive process they address [19].

Combining the Cognitive Process Theory of Writing with the new capabilities emerging from LLMs allows us to determine how intelligent writing support will likely impact specific writing processes. *Writing Priors*: Intelligent writing can be prompted using the three priors identified in the Cognitive Process Theory: 1. long-term memory information, i.e., why is what written for whom, and 2. the text as it exists so far. It may be possible to use intelligent writing to suggest which 3. rhetorical problem is currently to be solved. If the text to be written is very short, these three variables collapse into one: as the reason for writing and the current rhetorical problem become congruent, and there may be no text written so far, or, in experiments, it could be held constant, as we did.

Planning: Intelligent writing can be used to identify subgoals that can be implemented immediately. In the planning-subprocess of generation, it can help with ideation [11]. More advanced models can advise on the structure of the text [8]. If we again assume a very short text, goal setting collapses with the priors as long as the rhetorical problem is concrete enough to implement immediately. If someone gives instructions on what to write, content is given, and the generation process is removed. If the text is very short, reorganizing is not an applicable subprocess of planning because there is no room for maneuver.

Translating: LLMs can be used in the translating phase. An example of such a system is IntroAssist, which includes a checklist of best practices, highlighted text functionality and annotated examples to guide users in writing help requests [25]. Generally, they can be used to rephrase or elaborate on ideas, depending on the rhetorical problem and writing goal. This will include style. While it is feasible to pay no attention to depth and style, this does not remove the process. It simply carries it out poorly. Therefore, the translating variable must be kept even in a minimal case.

Reviewing: LLMs can evaluate text and suggest revisions (at least it should if prompted correctly). An example of such a system is AL, which analyzes the text the user provides to the system and identifies the level of argumentativeness and persuasiveness of the text while providing insights to the user to improve the content

further [47]. Reviewing is often omitted in a minimal case, such as brainstorming or chatting.

We therefore assumed the following a priori: Because intelligent writing support aids in exploration processes, there will be an impact on the time spent on writing tasks. Valid arguments can be made that this could be more or less time. For more time there can be several reasons. Namely, the stimulating nature of quasi-collaborative support may increase engagement with the writing task. Reduced opportunities for failure may lead to less satisfaction. For less time, we could take the argument that intelligent writing support can substitute cognitive processes, increasing time efficiency by making certain processes redundant. As the theoretical picture was unclear beforehand, we entertained both possibilities and assumed an undirected overall group difference.

Besides time, there is also the question of whether intelligent writing support is taken advantage of. The Cognitive Process Theory of Writing posits that individuals transition from one process to another, utilizing the outcomes of the previous process in the subsequent one. Based on this theory, we assumed that assistance for a specific process would likely be sought only when no existing results exist.

2.5 Hypotheses derived from the Cognitive Process Theory of Writing

- H1) Comparing intelligent support and static support for writing, the time people spend on ideation and translation differs significantly between the groups.
- H2) Comparing intelligent support and static support for writing, the time people spend on evaluation and revision differs significantly between the groups.
- H3) There is a decrease in the use of intelligent ideation support if static ideation support is present.
- H4) There is a decrease in the use of intelligent evaluation support if static evaluation support is present.

In the remainder of this paper, we will outline the operationalization of these hypotheses. We will report on our results and explanations for unexpected outcomes. Finally, we will discuss the implications of these results for using the Cognitive Process Theory of Writing in concert with intelligent writing support systems.

3 METHODS

To isolate essential hypotheses critical for theory falsification, we intentionally streamlined the variables, focusing on what is minimal or essential to test the theory-based constructs. This minimization aims to reduce the risk of confounding variables that could distort our findings. As has been shown, only the rhetorical problem and the processes of translation or revision are universally relevant, even in the minimal case of very short and short-lived text.

To gain insight into the inherently dynamic writing process as delineated in theory, however, a minimal interesting case needs to include at least one further process. That is to allow monitor activity, i.e., switching between processes. Given the current discussions emphasizing the role of existing ideas to translate in the Cognitive Process Theory of Writing [29], we opted to include 'ideation' in our first minimal case. In the second minimal case, we incorporated 'evaluation' to supplement the study's focus on the arch from

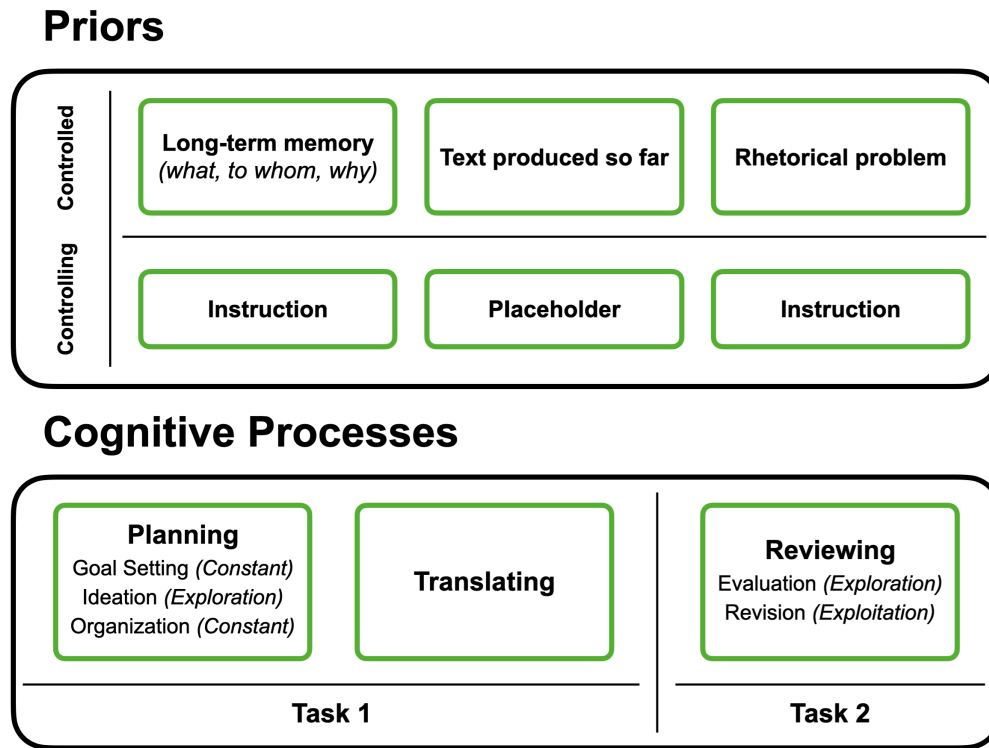


Figure 1: Overview of Cognitive Writing Processes highlighting the role of intelligent writing support. Exploration processes include goal setting, ideation, organization, and evaluation.

exploration to exploitation. The first case examines the transition between planning (ideation) and translating, while the second case delves into the transition between evaluation and revision (within reviewing).

3.1 Design and Procedure

Following recent calls for increased standardization of experimental tasks [19], we use student review writing in a 2x2 between-group design. The thrust of the argument is fixed to be in favor of providing feedback. The two binary factors are the presence or absence of a) relevant example ideas/feedback suggestions for improvement and b) intelligent writing support in a button. The button produces example ideas/feedback suggestions using API calls to the LLM GPT-3.5-turbo¹. We used prompts incorporating the text written so far. It was structured as a chat history, providing the model with example outputs to constrain generations².

¹<https://platform.openai.com/docs/models/gpt-3-5-turbo>

²For example, we used a system prompt "You provide one example idea per response. Give only the idea without any preamble or comment. Be as brief as possible.", and a chat history that specified "I need an example idea to include in a message. The message should convince my study group partners to seek feedback from our professor before submitting your assignment." for the user role, and gave examples in the form "Feedback develops writing skills for academic and professional success" for the assistant role. We then added the text produced so far to a second system message and prompted with a final user message: "Do the same but with a new idea."

Depending on random group assignments (uniform sampling without replacement), in the ideation phase, participants were supported in the argumentative essay task with content to use in their argument and/or intelligent support with ideation. In the reviewing phase, participants were supported with evaluations that suggested how to revise the text. Participants moved on from the first to the second phase after submitting their text by clicking on the Submit button, which was available after at least 250 words had been written.

The intelligent support is an implementation consisting of a button and an output field where generated suggestions are displayed. We kept the interface simple, to not accidentally introduce confounding influences on our measurements. Figure 2 shows Group 4 for the ideation task. The evaluation task was set up analogously. Group 3 in both tasks did not receive buttons. Group 2 did not receive the ideas or feedback suggestions on the right side of the tool, and Group 1 received neither.

3.2 Measures

Besides the independent grouping variable, there were two measures in the tool, which we treat as dependent variables: a) the time needed to complete the essay (250 words) and b) the number of uses of the intelligent support button. The descriptive statistics for these main variables can be seen in Table 1.

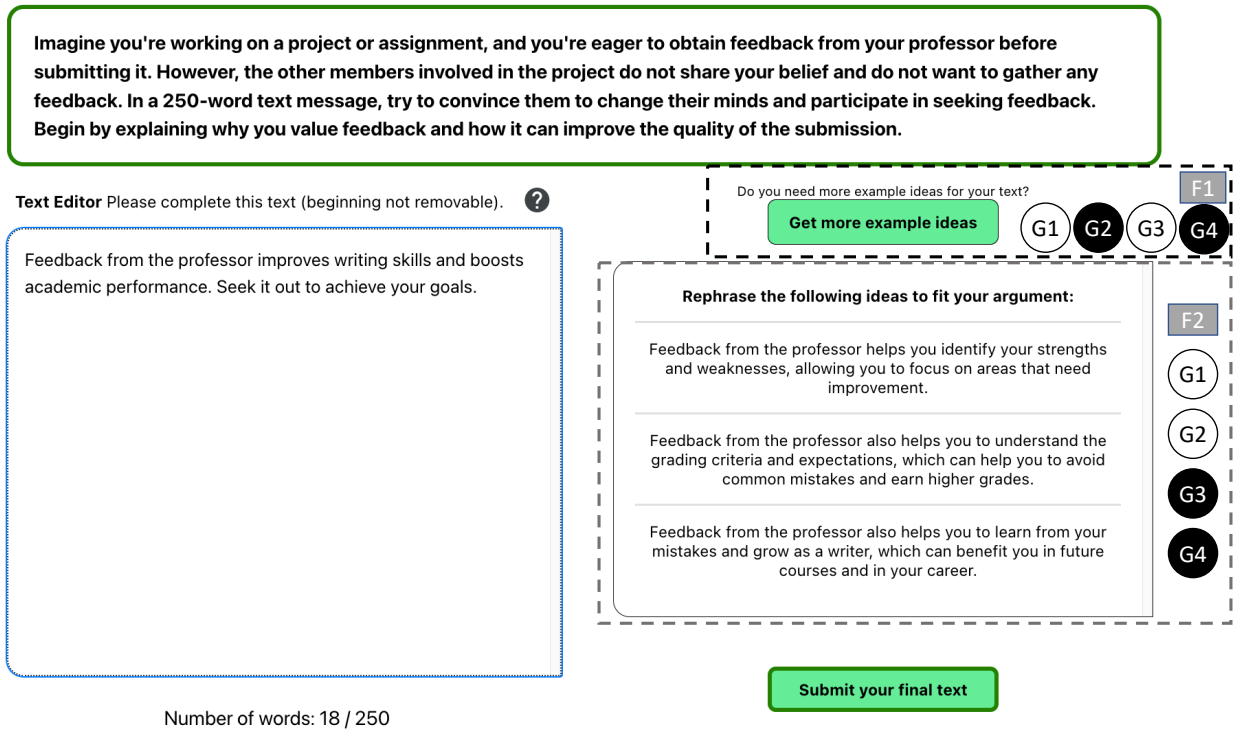


Figure 2: Full task and tool for ideation phase (task 1). Feature F1 and F2 for groups 1-4 are visible. Black background circle = feature present, white background circle = feature not present. G1-G4 indicates the groups.

Table 1: Descriptive statistics for each hypothesis by group. H1 was investigated in task 1 and H2 in task 2, H3 is related to ideation, and H4 to evaluation. Time spent is measured in seconds.

Hypothesis	Group	Variables	Mean	SD	Std. Mean	N
H1	1	time spent	817.33 s	404.2 s	-.195	27
H1	2	time spent	972.34 s	487.19 s	.134	32
H1	3	time spent	892.17 s	407.23 s	-.036	30
H1	4	time spent	941.93 s	567.96 s	.069	30
H2	1	time spent	296.88 s	168.34 s	-.059	25
H2	2	time spent	363.9 s	251.23 s	.266	31
H2	3	time spent	286.55 s	200.34 s	-.110	29
H2	4	time spent	285.19 s	186.35 s	-.116	31
H3	2	number button clicks	4.62	3.6	.233	32
H3	4	number button clicks	2.83	3.66	-.249	30
H4	2	number button clicks	2.71	2.62	.335	34
H4	4	number button clicks	1.12	1.52	-.356	32

In addition, we assessed potential covariates in the pre-and post-survey (See Table 2 for items originally developed for this study). We also assessed the number of cognitive process phases during the task, operationalized by defining exploitation (translating, revision) as the periods where typing was registered for 3 consecutive seconds or less and exploration (ideation, evaluation) where it was registered for more. We used typing as the indicator since it efficiently discriminates writing exploration from exploitation.

3.3 Hypotheses Testing

In an exploratory data analysis, we determined whether assumptions for parametric tests of our hypotheses were given. They were for hypotheses 1 and 2; for 3 and 4, they were not because of the distribution of the dependent variables. We therefore used non-parametric equivalents for them:

- H1) As time and residuals are approximately normally distributed, we used analysis of variance (ANOVA). We test overall group differences in the time spent on the task in seconds.
- H2) As time and residuals are approximately normally distributed, we again used ANOVA. We test overall group differences in the time spent on the task in seconds.
- H3) As the number of ideation button clicks is not normally distributed, we used a Wilcoxon rank sum test. We test a directed group difference between groups 2 and 4, the number of times the support button was clicked. Group 2 was predicted to use the button more due to the absence of cognitive process results.
- H4) As the number of evaluation button clicks is not normally distributed, we used a Wilcoxon rank sum test. We test a directed group difference between groups 2 and 4, the number of times the support button was clicked. Group 2 was predicted to use the button more due to the absence of cognitive process results.

4 RESULTS

4.1 Participants

We performed our field experiment over Prolific³. This is a crowdsourcing platform for experiments, and we selected it since previous studies on behavioral research platforms found that Prolific had the highest response quality and sample variety [37], crucial criteria for evaluating crowdsourcing platforms [10, 41, 48]. We recruited 120 participants with age: $m=33.27$, $SD=10.28$; gender: 28% female, 72% male; 66.7% indicated at least part-time employment and 28% were students. The selection criterion to be included in the study was fluency in English. Participants were compensated with standard rates if attention checks were fulfilled (which was the case for 4 participants; they were replaced).

4.2 Measures

Time spent in task 1 and task 2, respectively, were approximately normally distributed, as were the residuals of the linear models with the respective groups. *Use of ideation button and evaluation button* were Poisson-distributed because these are count data. We, therefore, had to use non-parametric tests for hypotheses 3 and 4. As a quality check for our tool, we assessed technology acceptance variables using Likert scales anchored at 1 (strongly disagree) to 7 (strongly agree), and a middle anchor (neither disagree nor agree), namely intention to use ($\alpha=.92$, $m=5.45$, $SD=1.34$), perceived usefulness ($\alpha=.94$, $m=5.47$, $SD=1.36$), and perceived ease of use ($\alpha=.75$, $m=5.68$, $SD=1.16$). We also analyzed the text submissions. Matching them with the suggestions, we found clear evidence that about 77% of ideas and 34% of evaluation suggestions were incorporated into the submissions; the discrepancy here is likely due to the higher difficulty in detecting the implementation of evaluation suggestions versus ideas, and should not be interpreted as conclusive evidence that ideas are implemented more likely. The submissions, furthermore, did not significantly differ in quality as measured by Text

Coherence, defined as cosine similarity between consecutive sentences [5, 34], and did only differ between groups 1 and 3 in task 1 by Fleischman Reading Ease score (See Table 6). Furthermore, we asked participants how important they felt ideas ($\alpha=.89$, $m=5.69$, $SD=0.93$) and evaluations ($\alpha=.93$, $m=5.47$, $SD=1.16$) were in the writing task and how well the tools supported them (ideation: $\alpha=.95$, $m=5.05$, $SD=1.53$; evaluation: $\alpha=.90$, $m=5.13$, $SD=1.28$). There were no group differences for these variables in task 2; however, in task 1, the technology acceptance variables and the variables indicating whether ideation/evaluation was important and supported did differ (see Table 3). Namely, group differences were pronounced between the presence and absence of intelligent writing support, with higher values in the supported groups (See Table 4). Interestingly, this is true for the importance of ideas/evaluations, which were influenced by the experimental variation. Another result was that these differences pertain even to variables that were, on the surface, more relevant for groups in task 2. This may be because participants spent more time on task 1 than task 2, rendering the impact of this grouping more powerful than the grouping for task 2. The cognitive phases were $m=34.38$ ($SD=22.34$) for task 1 and $m=13.32$ ($SD=8.75$) for task 2.

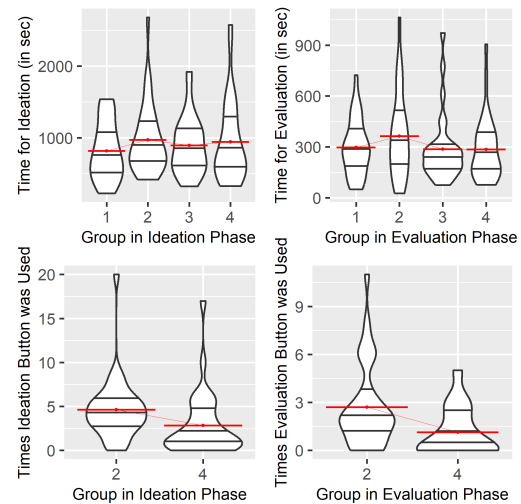


Figure 3: Violin density plots for the hypotheses. H1: upper left, H2: upper right, H3: lower left, H4: lower right. The red line connects the group means.

4.3 Results of Hypotheses Testing

Figure 3 and Table 1 show group differences relevant to the hypotheses. In terms of hypothesis testing, we can report the following findings:

- H1) is upheld with ($p=3.66e-11$, $F=21.55$). See the group differences in Table 7. Groups 1 and 3 are not significantly different, while the difference between groups 2 and 4 is the smallest significant difference. This indicates that group differences result from the presence or absence of intelligent writing support. Namely, the presence of intelligent writing support increases time spent with the tool.

³<https://www.prolific.com>; average compensation was advertised as 3£ for 20 minutes but turned out to be 7.19£ per hour on average, with median time spent 31 minutes 48 seconds

Table 2: Original items of the four variables.

1) Subjective Ideation Support	2) Subjective Evaluation Support
The tool helped with generating ideas for my writing task.	The tool helped me identify areas for improvement in my writing task.
The tool supported brainstorming for my writing task.	The tool supported my content evaluation and revision process in the task.
The tool aided in developing concepts for my writing task.	The tool assisted me in finding areas to refine in my writing task.
3) Importance of Ideas	4) Importance of Evaluation
Good ideas were essential for improving my writing.	Good feedback suggestions were essential for improving my writing.
Generating good ideas was key for enhancing my writing.	Incorporating good feedback suggestions was key for enhancing my writing.
Having good ideas was crucial for elevating the quality of my writing.	Having access to good feedback suggestions was crucial for elevating the quality of my writing.

Table 3: Survey variable group difference probabilities based on Kruskal-Wallis. itu/pu/peou: intention to use, perceived usefulness, and ease of use; subji/subje: Subjective ideation/evaluation support, impi/impe: Importance of ideas/evaluations for writing. *//** indicate significance at .05, .01, and .001 levels**

Phase	ITU	PU	PEOU	SUBJI	SUBJE	IMPI	IMPE
Task 1 <i>p</i>	.0225	.0001***	.0014*	.0000***	.0000***	.0007**	.0035*
Task 2 <i>p</i>	.9889	.9822	.4374	.4434	.1912	.7313	.6939

Table 4: Dunn post-hoc test results for variables with significant overall differences in Table 3. *//** indicate significance at .05, .01, and .001 levels.**

Compared Groups	Perceived usefulness	Ease of use	Ideation support	Evaluation support	Ideation importance	Evaluation importance
1-2	-1.34 (p=0.36)	-1.29 (p=0.59)	-3.43*** (p=0.00)	-2.77* (p=0.02)	-1.81 (p=0.21)	-2.28 (p=0.09)
1-3	2.05 (p=0.12)	2.18 (p=0.12)	0.03 (p=0.98)	1.02 (p=0.61)	1.03 (p=0.60)	0.58 (p=1.00)
1-4	-2.21 (p=0.11)	-1.05 (p=0.59)	-3.20*** (p=0.00)	-2.87* (p=0.02)	-2.62* (p=0.04)	-2.23 (p=0.08)
2-3	3.48*** (p=0.00)	3.57*** (p=0.00)	3.52*** (p=0.00)	3.88*** (p=0.00)	2.91* (p=0.02)	2.92* (p=0.02)
2-4	-0.92 (p=0.35)	0.24 (p=0.81)	0.19 (p=1.00)	-0.14 (p=0.89)	-0.86 (p=0.39)	0.02 (p=0.98)
3-4	-4.33*** (p=0.00)	-3.28*** (p=0.00)	-3.29*** (p=0.00)	-3.96*** (p=0.00)	-3.71*** (p=0.00)	-2.85* (p=0.02)

H2) is rejected with (p=.387, F=1.019). We can explain this by including the interaction of the groups with the number of cognitive process phases (see Table 5). This indicates that the time spent on the task only increased if the presence of intelligent writing support led to more phase changes.

H3) is upheld with (p=.003, W = 674 ; group 2: m=4.62, SD=3.60, group 4: m=2.83, SD=3.66).

H4) is upheld with (p=.002, W = 767; d=.74; group 2: m=2.71, SD=2.62, group 4: m=1.12, SD=1.52).

Overall, these results indicate a difference between having and not having access to intelligent writing support. Furthermore, it indicates a difference within the groups that received writing support, namely that it was used much more if no product of the relevant cognitive process for the instructed task was present beforehand.

5 USER FEEDBACK ON THE USER INTERFACE DESIGN

We asked users of the writing support tool, "What could be improved in our tool to make your writing more comfortable and effective?" to which they responded with 139 unique answers (22 participants provided two, 4 three answers). In terms of effectiveness, users pointed to four broad themes: specificity of suggestions, diversity of suggestions, an addition of grammar and spelling assistance, and real-time interaction.

Users called for more concrete and exact suggestions. One user noted the need to "Not give such general examples like 'improve academic performance' but instead more concrete anecdotal ideas," and another desired the tool to "Be more precise and specific on feedback." Diversity emerged as a second theme, and refers to the call for a broader range of suggestions and perspectives, with users expressing a need for "More suggestions, more points of view," and a

Table 5: Explanatory model for H2. Only group 3 is marginally different from the others. However, interactions are all significant. R-squared=.596, adjusted=.571. */*** indicate significance at .05, .01, and .001 levels**

	Estimate	Std. Error	t-value	Pr(> t)	Std. Coefficient
(Intercept)	135.9715	55.6450	2.44	0.0162*	NA
Group2	-117.3643	71.4422	-1.64	0.1033	-0.2535
Group3	-125.2662	73.9588	-1.69	0.0932	-0.2648
Group4	-50.6494	80.3521	-0.63	0.5298	-0.1094
Group1: Number of phases (typing or pausing)	12.1532	3.6765	3.31	0.0013**	0.3809
Group2: Number of phases (typing or pausing)	20.1585	2.2010	9.16	0.0000***	0.9324
Group3: Number of phases (typing or pausing)	20.4592	3.0999	6.60	0.0000***	0.7089
Group4: Number of phases (typing or pausing)	17.9076	4.7187	3.79	0.0002***	0.4901

Table 6: Means (Standard Deviations) of text quality measures for Task 1. Only Fleischman Reading Ease between groups 1 and 3 is significantly different, with a Dunn test statistic of -2.71 (p=0.0398). For task 2, there are no differences.

Groups Task 1	Fleischman Reading Ease (SD)	First Order Coherence (SD)
1	4.445367 (0.366916)	0.740306 (0.107329)
2	4.596607 (0.219107)	0.727243 (0.173445)
3	4.625036 (0.275010)	0.738566 (0.138479)
4	4.568336 (0.288384)	0.738019 (0.158015)

Table 7: Results of Tukey HSD post-hoc test for H1 (group differences for time spent on task 1). Only groups 1 and 3 are not significantly different. This indicates that there was no difference in time spent on the ideation task if there was no intelligent writing support. */*** indicate significance at .05, .01, and .001 levels**

Groups	Diff Time Spent in Seconds	Confidence Intervals [lwr, upr]	p
2-1	4.625	[2.844, 6.406]	0.000***
3-1	0.200	[-1.608, 2.008]	0.992
4-1	2.833	[1.025, 4.642]	0.000***
3-2	-4.425	[-6.174, -2.676]	0.000***
4-2	-1.792	[-3.540, -0.043]	0.042*
4-3	2.633	[0.857, 4.410]	0.001**

desire to "Have more points that can be included. A variety of points, so that I could choose the points I wanted to help structure and make my written piece flow." More specific yet diverse suggestions are a difficult and potentially diametric requirement, especially using traditional methods of making suggestions more specific, such as training with a more restricted dataset. Intelligent writing support incorporating LLM technology could be most suited to addressing this double requirement.

An additional theme was to incorporate real-time suggestions. One user stated, "It could come up with suggestions automatically as we type our writing," another wanted the tool to "Offer suggestions while typing rather than having to click on the tool for improvement ideas." Separately, a requirement for spell and grammar checking emerged, with users calling for features like "Spell checking" and "Automated grammar and spelling checks." Besides improving the suggestions, the mode of interacting with the suggestions (real-time vs. elicited) and additional features incorporating established writing support mechanisms based on grammar could improve the effectiveness of writing support tools.

Regarding comfort, users pointed to two broad themes: improved user interface, and experience and customization features. Users expressed the need for an interface that is both intuitive and visually appealing. One user specifically highlighted the importance of "integrating with AI where it could give you examples for its suggestions would be pretty nice," while another suggested the tool should add "Maybe even more readable user interface. Besides that I think there's room for improvement in buttons design," or more specifically "Be conversational, be able to ask and get answers, as a chat." There was also a suggestion for more intuitive control over the writing space, as one user expressed the need for the writing box to "go up and down when controlling it." These statements reflect requirements for a more comfortable and inviting user experience. Customization emerged as a second theme, referring to the desire for personalized settings and features. Users expressed wishes like "Maybe trying to make it more customizable" and "Maybe add a keyboard with special characters, like bullet points." Both improvement in the user interface and customization can increase comfort in using intelligent writing support tools.

Table 8: User Requirements and Design Implications for Intelligent Writing Support Tools

Themes of Comments	Example Quotations	Design Implications
Specificity of Suggestions	"Not give such general examples..."	Use LLM technology for more specific suggestions
Diversity of Suggestions	"More suggestions, more points of view..."	Provide diverse suggestions for users to choose from
Real-Time Interaction	"It could come up with suggestions automatically..."	Implement real-time suggestion mechanism
Grammar and Spelling Assistance	"Spell checking", "Automated grammar and spelling checks."	Add grammar and spell-check features
Improvement in UI	"integrating with AI where it could give you examples..."	Enhance user interface and button design
Customization Features	"Maybe trying to make it more customizable..."	Add customization options, special characters
Speed and Performance	"It can work a little bit faster.", "Speed."	Optimize for speed
Accessibility for Non-Native Speakers	"It could have a grammar checker..."	Include features for non-native speakers
Concerns of Plagiarism	"It was hard not to plagiarise directly..."	Address issues of plagiarism in suggestions
Comfort from Assistance	"I felt comfortable because the suggestions gave me ideas..."	Focus on user-friendly features and clear instructions
Discomfort Factors	"Having to write 250 words, seemed too many..."	Address usability issues and specific word count concerns

Furthermore, three additional points emerged. Firstly, speed and performance. Users emphasized the desire to have a fast tool. Some users were explicit in their demands, stating, "It can work a little bit faster.", or just "Speed." Secondly, accessibility for non-native speakers emerged as a theme. Users expressed concerns over issues like "It could have a grammar checker, very useful for users that are not native speakers of some language." This would add to the already raised point about implications for designing for effectiveness. Thirdly, it was remarked that "The original writing suggestions were quite specific and it was hard not to plagiarise directly. I spent more energy rephrasing than coming up with my own ideas". This last point may especially become important in positioning writing support tools in broader society, as it points to a shift in the relative importance of cognitive writing processes.

We asked our participants more specifically what made them comfortable and uncomfortable using the tool (139 unique answers for what made them comfortable, with 24 providing two, and 2 providing three answers; 92 unique answers, with 8 users providing two for what made them uncomfortable). In analyzing users' responses to the question "What made you feel comfortable?" several key themes emerged. Many respondents found comfort in the tool's assistance, suggestions, and guidance, with comments like "I felt comfortable because the suggestions gave me ideas that I haven't thought of" and appreciation for the "ideas generation tool." The ease of use, highlighted by remarks such as "user-friendly and simple to use" and "the simpleness of the platform" played a vital role in enhancing comfort. Some participants also emphasized the freedom and lack of pressure, illustrated by the statement, "I didn't feel like I had to rush and took my time to gather my thoughts." Others attributed comfort to personal confidence, enjoyment, or familiarity with the topic, reflecting sentiments like "Writing tasks are something I enjoy doing." Clear instructions and guidance were also valued, as in responses such as "The instructions were simple and clear." Conversely, discomfort was associated with specific word

count concerns, tool usability issues, pressure, and uncertainty. For instance, the remarks "Having to write 250 words, seemed too many for the task required," and "I wanted to copy a sentence and paste it, [...] but the program would not let me do that" reveal areas of user dissatisfaction or discomfort. Interestingly, a significant portion of users reported a lack of discomfort, indicating a generally positive experience for many participants; numerically, our 120 participants indicated on an analog scale of 1-101 that they felt on average, $m=69.9$ ($SD=22.7$) comfortable.

6 DISCUSSION

Large Language Models (LLMs) [7, 8] have enabled various new avenues for intelligent writing support [19]. Crafting valuable interactions with such AI models is challenging due to uncertainty about and complexities around them [49]. Such design challenges are a recurring theme in Human-Computer Interaction (HCI) research, e.g., when engaging AI for ideation that incorporates user context [28] or considering biased productions [27]. Previously, writing support systems were mainly rooted in enhancing grammar or style [14]. The rise of LLMs goes beyond mere syntax or grammar correction. These models now allow for enhanced planning and ideation, bridging the gap between conventional writing tools and those designed specifically for creativity, such as brainstorming software or concept mapping tools [17]. Our research looked into the effect of intelligent writing support for two important cognitive processes in writing [16], namely evaluation and ideation processes.

6.1 Effects of Intelligent Writing Support on Cognitive Writing Processes

Our results indicate that participants were writing for longer time periods when intelligent writing support was present — a mean difference of 100 seconds for ideation and 30 seconds for evaluation. However, this was only the case during the evaluation phase when

no predefined evaluations were shown. This suggests that tool engagement depended on intelligent writing support and the absence of pre-displayed evaluations. Participants used intelligent writing support less when ideas and evaluations were already displayed, showing a 39% and 59% decrease in usage for ideation and evaluation respectively. This supports the Cognitive Process Theory of Writing [16], implying reduced reliance on intelligent writing support when the results of cognitive processes of exploration are substituted with external information.

Our findings shed light on research for the support of the cognitive writing process in general and on HCI research on usability and user experience with intelligent writing support tools specifically. We believe this supports the notion that the Cognitive Process Theory of Writing provides design-determining constructs, namely what cognitive processes are used can therefore be supported during writing, for HCI in the domain of writing support [24].

Furthermore, we observed that intelligent writing support may play a role in increasing writing engagement. Namely, more time was voluntarily spent on the tasks when intelligent writing support was present, possibly indicating higher intrinsic motivation to submit high-quality writing. Hence, HCI researchers and practitioners can build on our research to study how different writing phases (planning, translating, and reviewing) can be supported in different writing domains (professional writing, educational writing, or creative writing).

6.2 Interplay of External Inputs and Cognitive Processes

In the ideation and evaluation tasks, measurements varied by group. Specifically, group 4 of the evaluation task, with evaluations and a generation button, had fewer processing phases ($m=10.81$, $SD=5.5$) than the overall mean ($m=13.32$, $SD=8.75$). This suggests the evaluations provided might have sufficed for task completion. Post-survey variable differences were found only in task 1 (ideation), indicating greater ease of use for groups with intelligent writing support. We speculate that task duration might impact covariates more than the experimental variations.

We also expand on previous research on the impact of writing support on the rhythms of writing [43]. Namely, the introduction of intelligent writing support impacts time and the number of distinct process phases. Future studies may extend this paradigm to more than two processes per task until the monitor activity, i.e., the switching between processes in natural writing tasks, is fully accounted for. For this, we call for investigations into the operationalization of all particular processes; these operationalizations should extend our measuring, focusing on whether typing occurred. This approach is feasible in very controlled circumstances when only two processes (exploration and exploitation) are expected. This controlled setting helps in isolating the effects of the writing support tool, offering clearer insights into its direct impact on the writing process.

Basing the study of new phenomena on previous insight can help overcome the uncertainty caused by them. We used an established theory on how writing works on the cognitive side to help us understand how the cognitive automation driven by intelligent writing support can be understood in the context of writing. Using

such a theoretical approach, we could integrate existing knowledge with this new phenomenon, which facilitated studying its application for writing support. It is a use case important to those looking to aid underperforming demographics and those looking to reinvent writing as a practice. Writing has been reinvented by new technology several times; only recently, handwriting was largely replaced by digital writing. This time, the changes may seem more paramount; however, by using theory to inform it, we can possibly steer the practice much better than ever before.

6.3 User Feedback and Practical Design Considerations

Our research may help users better understand the impact of collaborating with intelligent writing support. For designers, our research may help guide the configuration of writing support systems, as we show that cognitive writing processes ought to be included when considering how systems will be used. We can emphasize this for the role of time spent on writing and the actual use of the system.

Our users' feedback can be used to improve future intelligent writing support systems for both research and real-world applications. Users emphasized the need for increased effectiveness in guidance during the writing process and expressed the desire for a more comfortable interface. By integrating features that promote clear and concise suggestions while maintaining user-friendly navigation, future designs can better align with the practical requirements of writers. Balancing these factors can lead to a more engaging and productive writing experience, which supports the observed positive impact on time spent on writing and actual system usage in our study.

6.4 Study Limitations, Ethical Concerns, and Future Avenues

Some of the feedback of the experiment participants pertained to study specifics that may have influenced our results, namely the 250-word minimum requirement that caused discomfort for some participants. This constraint may have affected the natural flow of the writing process and potentially altered the way users interacted with the intelligent writing support tool. Another limitation of this study is that we did neither explicitly ask for prior experience with similar tools, nor run the study longitudinally, leaving the possibility open that part of the tool's effects were related to novelty. Understanding these study-specific limitations can inform future research designs, allowing for a more authentic assessment of how writers engage with intelligent support systems in unconstrained writing scenarios.

Our study has further limitations suggesting avenues for future research. Firstly, the scope is confined to the specific domain and participants studied. Despite pre-testing our tool and implementing attention checks, potential data invalidity may arise from Prolific participants, which furthermore are students, which is why our results may not generalize to professional writers. Secondly, we introduced an intelligent writing support system using GPT-3.5-turbo, which may produce biased or erroneous results due to inherent model constraints. However, future improvements in these models could mitigate such limitations. Thirdly, while we recognize the

ethical concerns of intelligent writing support, especially in academic writing [52] and unintentional plagiarism [26], they are not the focal point of this paper. We stress the need for future studies to explore biases, expand to diverse populations, and delve into the ethical dimensions of intelligent writing support.

We also want to point out that our users' feedback extended beyond the core focus of our research, uncovering general interest topics related to intelligent writing support. These include concerns about the potential for over-reliance on automated suggestions [27], curiosity about how AI can foster creativity [17], and interest in ethical considerations [52]. While these areas were not the primary focus of our investigation, they open avenues for future research in the domain of intelligent writing support.

7 CONCLUSION

Our study, rooted in the Cognitive Process Theory of Writing as a source of constructs for determining interaction design, investigated the complex relationship between human cognition and Large Language Models (LLMs) within the context of intelligent writing support tools. We developed a specialized tool for our investigation, focusing on how the introduction of intelligent support influences cognitive writing processes. Our findings revealed that when intelligent writing support was incorporated, users spent more time engaged with the tool.

Our findings bring a new dimension to the Cognitive Process Theory of Writing by demonstrating its applicability as a source of design-determining constructs for interaction design, particularly in intelligent writing support systems. This extension is particularly highlighted by increased user engagement and enhanced usability when intelligent support, powered by Large Language Models, is incorporated. The altered user experience further proves that this theoretical framework can be instrumental in shaping the interaction design of emerging, intelligent systems.

In this context, our research acts as a bridge between traditional writing practices and the evolving landscape of AI-powered support tools. Using LLMs as intelligent support changes the dynamics of user engagement, emphasizing the importance of theoretically informed design for higher usability and improved user experience. This focus on the Cognitive Process Theory of Writing offers beneficial insights not just for future Human-Computer Interaction initiatives but also for interdisciplinary approaches seeking to understand the influence of emerging technologies on interaction design and user behavior.

ACKNOWLEDGMENTS

We used a Large Language Model to improve the clarity and style of the paper.

REFERENCES

- [1] Ecenaz Alemdag and Zahide Yildirim. 2022. Effectiveness of online regulation scaffolds on peer feedback provision and uptake: A mixed methods study. *Computers & Education* 188 (2022), 104574. <https://doi.org/10.1016/j.compedu.2022.104574>
- [2] Kenneth C. Arnold, April M. Volzer, and Noah G. Madrid. 2021. Generative Models can Help Writers without Writing for Them. In *IUI Workshops*. RWTH Aachen University, 1–8. <https://ceur-ws.org/Vol-2903/IUI21WS-HAIGEN-1.pdf>
- [3] Tamara Babaian, Barbara J. Grosz, and Stuart M. Shieber. 2002. A writer's collaborative assistant. In *Proceedings of the 7th international conference on Intelligent user interfaces*, Kristian Hammond, Yolanda Gil, and David Leake (Eds.). ACM, New York, NY, USA, 7–14. <https://doi.org/10.1145/502716.502722>
- [4] Caroline Beauvais, Thierry Olive, and Jean-Michel Passerault. 2011. Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology* 103, 2 (2011), 415–428. <https://doi.org/10.1037/a0022545>
- [5] Gillinder Bedi, Facundo Carrillo, Guillermo A. Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B. Mota, Sidarta Ribeiro, Daniel C. Javitt, Mauro Copelli, and Cheryl M. Corcoran. 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ schizophrenia* 1 (2015), 15030. <https://doi.org/10.1038/npschz.2015.30>
- [6] Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). 2019. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. <https://doi.org/10.1145/3290605>
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc, 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc94967418bfb8ac142f64a-Paper.pdf>
- [8] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. [n. d.]. Sparks of Artificial General Intelligence: Early experiments with GPT-4. <https://doi.org/10.48550/arXiv.2303.12712>
- [9] Minsuk Chang, John Joon Young Chung, Katy Ilonka Gero, Ting-Hao Kenneth Huang, Dongyeop Kang, Mina Lee, Vipul Raheja, and Thiemo Wambansgans. 2023. The Second Workshop on Intelligent and Interactive Writing Assistants. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, and Anicia Peters (Eds.). ACM, New York, NY, USA, 1–5. <https://doi.org/10.1145/3544549.3573826>
- [10] Peng Cheng, Xiang Lian, Zhao Chen, Rui Fu, Lei Chen, Jinsong Han, and Jizhong Zhao. [n. d.]. Reliable diversity-based spatial crowdsourcing by moving workers.
- [11] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *CHI Conference on Human Factors in Computing Systems*, Simone Barbosa, Cliff Lampe, Caroline Appert, David A. Shamma, Steven Drucker, Julie Williamson, and Koji Yatani (Eds.). ACM, New York, NY, USA, 1–19. <https://doi.org/10.1145/3491102.3501819>
- [12] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop. In *23rd International Conference on Intelligent User Interfaces*, Shlomo Berkovsky, Yoshinori Hijikata, Jun Rekimoto, Margaret Burnett, Mark Billinghurst, and Aaron Quigley (Eds.). ACM, New York, NY, USA, 329–340. <https://doi.org/10.1145/3172944.3172983>
- [13] Andy Coenen, Luke Davis, Daphne Ippolito, Emily Reif, and Ann Yuan. [n. d.]. Wordcraft: a Human-AI Collaborative Editor for Story Writing. <http://arxiv.org/pdf/2107.07430v1>
- [14] Robert Dale and Jette Viethen. 2021. The automated writing assistance landscape in 2021. *Natural Language Engineering* 27, 4 (2021), 511–518. <https://doi.org/10.1017/S1351324921000164>
- [15] Ralph P. Ferretti and Yue Fan. 2016. Argumentative writing. In *Handbook of Writing Research*. Guilford Press, 301–315. <https://link.springer.com/article/10.1007/s11145-019-09950-x>
- [16] Linda Flower and John R. Hayes. 1981. A Cognitive Process Theory of Writing. *College Composition and Communication* 32, 4 (1981), 365. <https://doi.org/10.2307/356600>
- [17] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–18. <https://doi.org/10.1145/3290605.3300619>
- [18] Takayuki Fujimoto, Muhammad Dzulqarnain Muhammad Nasir, and Tokuro Matsuo. 2008. A Design on Collaborative-Cooperative Document Edit System Based on Cognitive Analyses. In *Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008)*. IEEE, 661–666. <https://doi.org/10.1109/ICIS.2008.81>
- [19] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for Science Writing using Language Models. In *Designing Interactive Systems Conference*, Florian 'Floyd' Mueller, Stefan Greuter, Rohit Ashok Khot, Penny Sweetser, and Marianna Obrist (Eds.). ACM, New York, NY, USA, 1002–1019. <https://doi.org/10.1145/3532106.3533533>
- [20] Katy Ilonka Gero, Tao Long, and Lydia B. Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal,

- Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–15. <https://doi.org/10.1145/3544548.3580782>
- [21] Steve Graham. 2019. Changing How Writing Is Taught. *Review of Research in Education* 43, 1 (2019), 277–303. <https://doi.org/10.3102/0091732X18821125>
- [22] Nick Greer, Jaime Teevan, and Shamsi T. Iqbal. [n. d.]. An Introduction to Technological Support for Writing. <https://www.microsoft.com/en-us/research/publication/an-introduction-to-technological-support-for-writing/>
- [23] Tracey S. Hodges. 2017. Theoretically Speaking: An Examination of Four Theories and How They Support Writing in the Classroom. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 90, 4 (2017), 139–146. <https://doi.org/10.1080/00098655.2017.1326228>
- [24] Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Gloria Mark, Susan Fussell, Cliff Lampe, m.c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, New York, NY, USA, 5040–5052. <https://doi.org/10.1145/3025453.3025765>
- [25] Julie S. Hui, Darren Gergle, and Elizabeth M. Gerber. 2018. IntroAssist. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Regan Mandryk, Mark Hancock, Mark Perry, and Anna Cox (Eds.). ACM, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3173596>
- [26] Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. [n. d.]. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. <https://doi.org/10.48550/arXiv.2211.05030>
- [27] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–15. <https://doi.org/10.1145/3544548.3581196>
- [28] Janin Koch, Andrés Lucero, Lena Hegemann, and Antti Oulasvirta. 2019. May AI?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300863>
- [29] Donald Ruggiero Lo Sardo, Pietro Gravino, Christine Cuskley, and Vittorio Loreto. [n. d.]. Exploitation and exploration in text evolution. Quantifying planning and translation flows during writing. <https://doi.org/10.48550/arXiv.2302.03645>
- [30] Xichu Ma, Ye Wang, Min-Yen Kan, and Wee Sun Lee. 2021. AI-Lyricist. In *Proceedings of the 29th ACM International Conference on Multimedia*, Heng Tao Shen, Yueting Zhuang, John R. Smith, Yang Yang, Pablo Cesar, Florian Metzger, and Balakrishnan Prabhakaran (Eds.). ACM, New York, NY, USA, 1002–1011. <https://doi.org/10.1145/3474085.3475502>
- [31] Shakked Noy and Whitney Zhang. [n. d.]. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. <https://doi.org/10.2139/ssrn.4375283>
- [32] R. L. Oakman. 1994. The evolution of intelligent writing assistants: trends and future prospects. In *Proceedings Sixth International Conference on Tools with Artificial Intelligence. TAI 94*. IEEE Comput. Soc. Press, 233–234. <https://doi.org/10.1109/TAI.1994.346488>
- [33] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. BunCho: AI Supported Story Co-Creation via Unsupervised Multitask Learning to Increase Writers' Creativity in Japanese. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, and Takeo Igarashi (Eds.). ACM, New York, NY, USA, 1–10. <https://doi.org/10.1145/3411763.3450391>
- [34] Alberto Parola, Jessica Mary Lin, Arndis Simonsen, Vibeke Bliksted, Yuan Zhou, Huiling Wang, Lana Inoue, Katja Koelkebeck, and Riccardo Fusaroli. 2022. Speech disturbances in schizophrenia: Assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophrenia research in press* (2022). <https://doi.org/10.1016/j.schres.2022.07.002>
- [35] Melissa M. Patchan, Christian D. Schunn, and Russell J. Clark. 2018. Accountability in peer assessment: examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education* 43, 12 (2018), 2263–2278. <https://doi.org/10.1080/03075079.2017.1320374>
- [36] Suparn Patra and Manoel Cortes Mendez. 2022. 37% of Coursera's 6400 Courses Have Peer Reviews: Here Are the Best. <https://www.classcentral.com/report/courses-with-peer-reviews/>
- [37] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153–163.
- [38] Junaid Qadir. [n. d.]. Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. https://www.techrxiv.org/articles/preprint/Engineering_Education_in_the_Era_of_ChatGPT_Promise_and_Pitfalls_of_Generative_AI_for_Education/21789434
- [39] Eva Ritz, Roman Rietsche, and Jan Marco Leimeister. 2023. How to Support Students' Self-Regulated Learning in Times of Crisis: An Embedded Technology-Based Intervention in Blended Learning Pedagogies. *Academy of Management Learning & Education* 22, 3 (2023), 357–382. <https://doi.org/10.5465/amle.2022.0188>
- [40] Mike Rose. 1980. Rigid Rules, Inflexible Plans, and the Stifling of Language: A Cognitivist Analysis of Writer's Block. *College Composition and Communication* 31, 4 (1980), 389–401. <http://www.jstor.org/stable/356589>
- [41] Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 2863–2872. <https://doi.org/10.1145/1753846.1753873>
- [42] Dhawal Shah. 2021. By The Numbers: MOOCs in 2021. <https://www.classcentral.com/report/mooc-stats-2021/>
- [43] Mike Sharples. 1994. Computer support for the rhythms of writing. *Computers and Composition* 11, 3 (1994), 217–226. [https://doi.org/10.1016/8755-4615\(94\)2894\(2\)2990014-0](https://doi.org/10.1016/8755-4615(94)2894(2)2990014-0)
- [44] Henrik Kohler Simonsen. 2022. AI Text Generators and Text Producers. In *2022 International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 218–220. <https://doi.org/10.1109/ICALT55010.2022.00071>
- [45] Carola Strobl, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Proske, and Christian Rapp. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education* 131 (2019), 33–48. <https://doi.org/10.1016/j.compedu.2018.12.005>
- [46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. [n. d.]. Attention Is All You Need. <https://doi.org/10.48550/arXiv.1706.03762>
- [47] Thiemo Wambsganss, Matthias Söllner, and Jan Marco Leimeister. 2020. Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills. In *International Conference on Information Systems (ICIS)*. AIS Electronic Library (AISeL), Hyderabad, India.
- [48] Vanessa Williamson. 2016. On the ethics of crowdsourced research. *PS: Political Science & Politics* 49, 1 (2016), 77–81.
- [49] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjorn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376301>
- [50] Jingwen Zhang, Yoo Jung Oh, Patrick Lange, Zhou Yu, and Yoshimi Fukuoka. 2020. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. *Journal of medical Internet research* 22, 9 (2020), e22845. <https://doi.org/10.2196/22845>
- [51] Xin Zhao. 2022. Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal* (2022), 003368822210940. <https://doi.org/10.1177/00336882221094089>
- [52] Hazem Zohny, John McMillan, and Mike King. 2023. Ethics of generative AI. *Journal of medical ethics* 49, 2 (2023), 79–80. <https://doi.org/10.1136/jme-2023-108909>
- [53] Zheng Zong, Christian Schunn, and Yanqing Wang. 2022. What makes students contribute more peer feedback? The role of within-course experience with peer feedback. *Assessment & Evaluation in Higher Education* 47, 6 (2022), 972–983. <https://doi.org/10.1080/02602938.2021.1968792>

A STATIC IDEAS AND FEEDBACK

The static ideas that we presented participants with were, in task 1 (ideation): "Feedback develops writing skills for academic and professional success", "Feedback tailors assignments to meet professor's expectations for better grades", "Feedback improves critical thinking skills and leads to better decision-making". In task 2 (evaluation), we used: "Avoid using informal language such as 'you guys' in academic or professional writing", "Avoid repetition, as there is here with defensiveness; instead consolidate similar points for greater clarity and conciseness.", "Include specific examples or anecdotes to highlight the importance of accepting feedback from professors, making your statement more relatable and impactful." On average, the cosine similarity, a measure of semantic similarity, for the static and the generated suggestions was .56 with the ideas and .28 with the evaluation suggestions (feedback).