Proposed Experiment for the Impact of AI on Writing as Thinking

Katy Ilonka Gero
katy@g.harvard.edu
Harvard University
Cambridge, MA, USA

Abstract

In this short paper, we propose an experiment design for understanding how generative models impact the thinking that occurs as part of the writing process. We first introduce the arguments and evidence for writing as a thinking tool. Then we propose investigating how generative models impact the thinking that occurs during the drafting process. The proposed experiment has participants bring an outline for a writing task and then either drafting the writing themselves or using a generative model to do so for them. The experiment would measure how participants' ideas change before and after the drafting process in the two conditions.

Keywords

writing assistants, critical thinking, generative AI tools

1 Introduction

Writing has long been associated with learning. Janet Emig made a compelling argument in her paper "Writing as a Mode of Learning" [2] in which she proposes that there is a unique correspondence between writing and learning. Drawing on theories from philosophy and psychology, she defines learning as a "re-organization ... of a cognitive scheme in light of an experience" and then argues that because the *process* of writing results in an immediate and available *product* (that is, the text), writing uniquely allows for the feedback necessary for a re-organization in light of an experience, i.e., learning. It is writing's unique relationship to learning that makes it a powerful tool for thinking (presuming, as we do, that learning requires thought).

In educational contexts, writing is often used as an evaluation of critical thinking. Starting in the 1980s in the United States, writing was also proposed and pushed as a learning activity to develop thinking skills, rather than just evaluate them [4]. This emphasis on writing as a teaching device highlights again its unique role as a mode of thinking.

To further this point, consider these three vignettes that demonstrate how writing, even writing in which the writer may presume they already have done all the thinking, can change our thoughts:

(1) You prepare send a quick email to schedule a meeting. You already have in mind when to schedule the meeting. As you write the proposed time and date, you realize you are actually busy at that time, and need to think of a new time to propose.
(2) You prepare to draft a grant proposal. You already have an outline for the proposal, having discussed and detailed the idea with collaborators extensively. As you write a particular section of the grant, you realize a flaw in the ideas, and come up with a way to correct it.
(3) You prepare to journal about your day. As you write down what happened in the morning, you realize that something that occurred at the start of the day had been causing you stress all day, although you had not realized, during the day, the source of your stress.

The rise of generative models as writing assistants in a variety of contexts thus poses critical questions about their impact on *writing as a mode of thinking*. In this paper, we consider the particular case of using a generative model to convert an outline or series of rough ideas into a first draft. This is an interesting use case of writing assistants as it presumes that the writer has already thought of the ideas or the content of the writing, and the assistant "merely" converts it into clean prose. However, as articulated above, the act of converting ideas into fluent text can result in new thoughts.

We propose two hypotheses on how writing assistants may change writers' thought processes when used in this way:

(a) The AI generated draft can introduce new and valuable ideas that the writer had not previously considered.
(b) The AI generated draft can result in the writer not thinking of ideas they would have come up with had they done the drafting themselves.

We do not consider these hypotheses to be mutually inclusive; in fact, we expect that each are true in certain cases and may even be true at the same time. If this is the case, then it is worth considering the trade-off between the new ideas writing assistants provide and the ideas "lost" by a writer offloading the drafting task.

In this paper, we propose an experimental design that would test these hypotheses and lead to initial investigations on how AI acts as a tool for thought in the writing process.

2 Related Work

Multiple scholars have suggested that writing can be a mode of thinking. Notably, Emig (1977) draws on multiple sources to argue that writing is a unique mode of learning for a variety of reasons including that writing is multi-representational (requiring the hand, eye, and brain), that writing provides both short- and long-term feedback, that the rapid feedback loop between transcription and reading is key to re-organizing thoughts, as well as the fact that writing is active, engaged, and personal (i.e., self-rhythmed) [2]. Research on the development of writing skills has provided evidence for this idea, for instance Scardamalia and Bereiter's (1987) research on the development of writing in skills in children [8]. This research shows how immature writers move from "'knowledge telling" (e.g., listing associated concepts or events) to the more mature writing of "knowledge transforming" in which the "understanding of what they are trying to say grows and changes in the course of writing", allowing for more sophisticated writing structures (i.e., beyond listing associated concepts or events). There has been more recent

experimental evidence that writing activities improve critical thinking skills, such as Quitadamo and Kurtz's (2007) semester-long experiment with biology students, in which they compared writing activities to quizzes, and found a significant increase in critical thinking skills in the writing group [7].

With the advent of generative AI, discussion and experiments about how such tools impact thought more generally are starting to be disseminated. Lee et al. (2025) surveyed 319 knowledge workers, finding that the use of generative AI "shifts the nature of critical thinking toward information verification, response integration, and task stewardship" [5]. Levine et al. (2024) study how high schoolers make use of ChatGPT in group writing assignments, finding that students were critical of generated outputs, and "participating students treated ChatGPT as a stand-in for a teacher, peer, or range of existing online sources and digital applications" [6]. In contrast, Krupp et al (2024) found that nearly half of students with access to ChatGPT mistakenly assumed its answers were correct, even in their own field of study (physics) [3]. The impact of generative AI on thinking certainly can vary depending on user and task characteristics.

3 Proposed Experiment Design

In this work, we propose an experiment design in which participants are asked to either a) write a first draft based on their own outline or notes, or b) generate (with a generative AI tool) a first draft based on their own outline or notes. The goal of this design is to isolate the impact of a generative model on the thinking involved in the drafting process.

There are two ways to run this experiment. In one, participants bring their own writing task or tasks; in the other, participants are assigned a writing task. The benefit of the first version is that participants are more invested in their own writing tasks than toy tasks provided by an experimenter; the detriment is that the writing task becomes a large confounder in the experiment. The benefit of the second version is that it is easier to control across or between participants; however participants may not engage deeply with toy tasks, making their actions and thinking poor representations of their actual workflow or activities.

3.1 Potential Tasks

Potential writing tasks that participants could bring or be assigned include:

- an analytical essay
- a research proposal
- a cover letter
- a reported article
- an opinion piece

Given time constraints, writing tasks need to be scoped to approximately one-page tasks such that tasks can be completed in 30-60 minutes.

Participants would either bring an outline for a self-selected writing task, e.g., one they are actively working on outside of the experimental context, or they would be provided with a prompt for a writing task and develop an outline for the writing task as part of the experiment procedure. These represent two different versions of the study that could be run.

3.2 Procedure

Participants would first fill out a pre-task survey on demographic information, as well as relevant attributes such as experience with and trust in AI, self-efficacy in writing, and need for cognition [1].

Participants would then produce an outline for the writing task (either previously developed or developed during the experiment). This outline could consist of notes, bullet points, or other text that they have written that represents their plan for the writing, but does not contain verbatim language expected to be used in the draft.

Participants would then answer the following question:

> *What are the core ideas you have developed (including ones you may have discarded) in preparation for this writing task? Please list at least two, but feel free to list as many as you feel are relevant. Ideas should be short sentences that represent a kernel of a thought or idea.*

Participants then produce a first draft in response to the writing task. In the baseline condition, they draft the essay themselves without AI support or intervention. In the intervention condition, they prompt a language model to draft the essay based on their notes. In both conditions, participants are allowed to write and revise (including, in the intervention condition, revising prompts or otherwise interacting with the language model) as they see fit. In the intervention condition, participants are required to generate at least one draft and make use of it as the 'starting point' for their writing.

After a set period of time, or once the participant feels they have an adequate first draft, participants are asked the following question:

> *Given the core ideas you listed previously, how have these ideas changed, if at all? Please add to the list any new ideas that have been developed (if new ideas have been developed) or how an idea may have changed or been discarded.*

Finally, participants are given a post-task survey asking questions about effort, engagement, enjoyment, ownership, and authorship.

3.3 Analysis

There are two central measures in this experiment. The first is how ideas differ from producing an outline (i.e., the initial list of ideas) and from producing a draft (i.e., how participants reported those ideas changing). Quantitative measures could include the change in total number of ideas, the number of ideas that changed, and the number of ideas that were discarded. Diversity of ideas can also be measured across participants if participants respond to the same writing prompts; e.g., are final ideas more diverse in one of the conditions. Finally, a qualitative analysis of how ideas changed (like thematic analysis or grounded theory) can reveal other aspects of how ideas or thinking change.

The second is how these measures differ between the conditions. In this case, the measure is the *difference* in change of ideas from pre- to post-drafting.

3.4 Open Questions

There are several confounding factors to consider. The first is experience with, confidence in, and sentiment towards using generative models to draft writing. For instance, it could be that participants who are experienced with generative models for writing are able to produce higher quality drafts which may modulate how the generative model changes their thinking process. Participants who feel negatively towards this use case of generative models may not engage with the generated draft.

The prompt asking for participant ideas pre- and post-drafting may be unclear. Participants may need to do a small example task or be given explicit examples in order to respond to the prompt consistently, or the prompt may need revision in order to elicit the desired response.

Measuring quantity and quality of ideas is, obviously, difficult. Measures like "the number of ideas" may be difficult to align across participants. Finally, participants' self-report of how ideas changed may not be accurate represents of either how their ideas changed or of their thinking process. In the first case, actually looking at the draft and how it differs from the outline may be a more fruitful way to investigate these questions. In the second case, retrospective think aloud protocols may be a better initial methodology for exploring these questions.

References

[1] John T. Cacioppo and Richard E. Petty. 1982. The Need for Cognition. *Journal of Personality and Social Psychology* 42, 1 (Jan. 1982), 116–131. doi:10.1037/0022-3514.42.1.116

[2] Janet Emig. 1977. Writing as a Mode of Learning. *College Composition and Communication* 28, 2 (May 1977), 122–128. doi:10.58680/ccc197716382

[3] Lars Krupp, Steffen Steinert, Maximilian Kiefer-Emmanouilidis, Karina E. Avila, Paul Lukowicz, Jochen Kuhn, Stefan Küchemann, and Jakob Karolus. 2024. Unreflected Acceptance – Investigating the Negative Consequences of ChatGPT-Assisted Problem Solving in Physics Education. In *Frontiers in Artificial Intelligence and Applications*, Fabian Lorig, Jason Tucker, Adam Dahlgren Lindström, Frank Dignum, Pradeep Murukannaiah, Andreas Theodorou, and Pınar Yolum (Eds.). IOS Press. doi:10.3233/FAIA240195

[4] Judith A. Langer and Arthur N. Applebee. 1987. *How Writing Shapes Thinking: A Study of Teaching and Learning.* Number no. 22 in NCTE Research Report. National Council of Teachers of English, Urbana, Ill.

[5] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. (2025).

[6] Sarah Levine, Sarah W. Beck, Chris Mah, Lena Phalen, and Jaylen PIttman. 2024. How Do Students Use ChatGPT as a Writing Support? *Journal of Adolescent & Adult Literacy* (July 2024), jaal.1373. doi:10.1002/jaal.1373

[7] Ian J. Quitadamo and Martha J. Kurtz. 2007. Learning to Improve: Using Writing to Increase Critical Thinking Performance in General Education Biology. *CBE—Life Sciences Education* 6, 2 (June 2007), 140–154. doi:10.1187/cbe.06-11-0203

[8] Marlene Scardamalia and Carl Bereiter. 1987. Knowledge Telling and Knowledge Transforming in Written Composition. In *Advances in Applied Psycholinguistics*. Vol. 2. Cambridge University Press, 142–175.