



The Use of ChatGPT in Source-Based Writing Tasks

Christian Tarchi¹ · Alessandra Zappoli¹ · Lidia Casado Ledesma¹ ·
Eva Wennås Brante²

Accepted: 8 June 2024 / Published online: 19 June 2024
© The Author(s) 2024

Abstract

ChatGPT, a chatbot based on a Generative Pre-trained Transformer model, can be used as a teaching tool in the educational setting, providing text in an interactive way. However, concerns point out risks and disadvantages, as possible incorrect or irrelevant answers, privacy concerns, and copyright issues. This study aims to categorize the strategies used by undergraduate students completing a source-based writing task (SBW, i.e., written production based on texts previously read) with the help of ChatGPT and their relation to the quality and content of students' written products. ChatGPT can be educationally useful in SBW tasks, which require the synthesis of information from a text in response to a prompt. SBW requires mastering writing conventions and an accurate understanding of source material. We collected 27 non-expert users of ChatGPT and writers ($M_{age} = 20.37$; $SD=2.17$). We administered a sociodemographic questionnaire, an academic writing motivation scale, and a measure of perceived prior knowledge. Participants were given a source-based writing task with access to ChatGPT as external aid. They performed a retrospective think-aloud interview on ChatGPT use. Data showed limited use of ChatGPT due to limited expertise and ethical concerns. The level of integration of conflicting information showed to not be associated with the interaction with ChatGPT. However, the use of ChatGPT showed a negative association with the amount of literal source-text information that students include in their written product.

Keywords Source-based Writing · Argumentative Writing · Artificial Intelligence · Higher Education

✉ Christian Tarchi
christian.tarchi@unifi.it

¹ University of Florence, IT Via Laura, 48, Florence, IT 50121, Italy

² Malmö University, Malmö, Sweden

Introduction

Soon after ChatGPT was released to the public in 2022, concerns were raised about its contribution to disinformation. On the one hand, AI can contribute to the spread of false or inaccurate information. But, on the other hand, AI can become an ally in the fight of disinformation, as it can support people's reflective writing competence. Indeed, disinformation is also caused by a difficulty by young adults to access and elaborate information about complex topics. As such, higher education students could benefit from AI in becoming more informed over relevant topics for effective decision-making and reflective citizenship. Overall, more efforts are made towards regulation of the use of AI (European Parliament et al., 2019), rather than on investigating how people interact with AI and whether AI can support people in critical thinking task.

In the present study we aimed at exploring how higher education students interact with AI-tools generating text when they are required to read and reflect about a complex socio-scientific issue. In specific, they were asked to interact with ChatGPT in a source-based writing task, a task in which students are asked to read sources on an issue, often offering different perspectives, with the purpose of writing an essay. Essays can vary in genre (argumentative, synthesis, persuasive, opinion, and the like). We were interested in identifying how they interact with the tool, whether their interactions are strategic, whether they use the text produced by ChatGPT and how these aspects are associated with the argumentative quality of their written essays. To this end, the study design combined exploration of source-based writing with AI-assisted writing with the addition of retrospective think-aloud protocols.

In the following paragraphs we will first provide a definition of artificial intelligence and describe how it has been traditionally studied in educational settings. Secondly, we turn our attention towards ChatGPT as a chatbot producing text and its potential implication in writing tasks. Thirdly, we introduce the target outcome of the study, namely source-based writing. Fourthly, we discuss the issue of collaborative writing, as interacting with ChatGPT could be considered as a form of human-AI collaboration in text production. Finally, we present the rationale and the research question driving the present study.

Regarding our methodological approach, in the study the participants were asked to read a text about a controversial topic (use of Glyphosate) and write an argumentative essay, while giving the possibility to interact with ChatGPT. The study was conducted with Italian university students, reading and writing in Italian as their L1. This study was conducted in 2023, an era when generative language models were synonymous with ChatGPT for the general public.

Artificial Intelligence in Higher Education

According to Baker and Smith (2019), “Artificial Intelligence (AI) comprises computers that execute cognitive functions typically associated with human minds, particularly those involving learning and problem solving” (p. 10). These authors assert that AI does not represent a singular technology but serves as an umbrella term that

encompasses a variety of technologies and methodologies, including machine learning, natural language processing, data mining, neural networks, and algorithms.

AI can be used to adapt instruction to the needs of different types of learners (Verdú et al., 2017), to provide customized prompt feedback (Dever et al., 2020), and to develop assessment strategies (Baykasoglu et al., 2018). AI can contribute to both teaching and learning processes at different educational levels, from primary education to university education. In a commentary on AI and education, Kasneci et al. (2023) foresaw for university students a few learning opportunities from the implementation of large language models (LLMs). LLMs can assist students in most reading and writing tasks and act as a support for critical thinking and problem-solving. For instance, LLMs can be used to generate concise summaries and structured outlines of texts, provide information about specific topics, or hint towards some aspects that students have not thought about. AI could be a valuable tool for enhancing research skills, as it provides students with comprehensive information and resources related to specific topics. It also offers hints about unexplored dimensions and current research trends, thereby enabling students to gain a deeper understanding of the material and conduct more insightful analyses. This is relevant to the present study, as it suggests a few ways through which students could interact with chatbots based on a Generative Pre-trained Transformer model when engaged in a writing task.

Eager and Brunton (2023) suggested how the efficacy of AI in education may depend on the ability to write effective prompts for use with conversation-style AI models. AI tools have been proven as beneficial in research endeavors, writing assignments, and the cultivation of critical thinking and problem-solving abilities (Eager & Brunton, 2023). This reflection is very relevant to the present study as it suggests analyzing whether students are effective in writing prompts for chatbots based on a Generative Pre-trained Transformer model when engaged in a writing task.

A recent literature review on Artificial intelligence in higher education (Crompton & Burke, 2023) found that most of the studies were limited to the areas of language learning, computer science and engineering. Moreover, they stressed that, as scholars from the Educational Sciences are assuming an increasing role in researching the use of AI in Higher Education, more effort should be made to share their research regarding the pedagogical affordances of AI so that this knowledge can be applied by faculty across disciplines.

The Role of ChatGPT in Writing Tasks

ChatGPT stands as a chatbot based on a large language model (GPT3.5 or GPT4). Its transformer model employs the concept of self-attention, effectively assigning varying levels of importance to different components of the input data. The transformer model undergoes a pre-training stage on an extensive corpus of text data, where it acquires an understanding of context and meaning by discerning connections within sequential data, like the words within a sentence. Following this pre-training phase, the transformer model becomes capable of crafting human-like responses to user inputs (Adamopoulou & Moussiades, 2020; Flordi, 2023; Kasneci et al., 2023; Su et al., 2023).

As research on the use of chatbots in writing is only emerging and there is a lack of studies on source-based writing, following we report a literature review of opinion papers published after the release of ChatGPT. ChatGPT can assume a valuable role in providing versatile assistance at different stage of writing (Sallam, 2023), spanning from generating ideas to finalizing proofreading and editing for written content. The potential of ChatGPT as an assistant to the writing process includes the following aspects (Imran & Almusharraf, 2023). Firstly, it enhances efficiency by significantly reducing the time and effort required for content creation, benefiting both students and educators (Lund et al., 2023; Yan, 2023). Secondly, it provides ideation support by suggesting new ideas and perspectives for writing assignments (Kasneci et al., 2023; Taecharungroj, 2023). Additionally, it offers invaluable language translation assistance, helping non-native language students ensure accuracy and grammatical correctness in their writing (Lametti, 2022; Lund & Wang, 2023; Stacey, 2022; Stock, 2023).

In summary, ChatGPT can contribute to writing processes in multiple ways. In fact, some studies have highlighted the positive results of ChatGPT in relation to the development of argumentative texts (e.g., Su et al., 2023), the development of writing practices linked to second language learning (e.g., Barrot, 2023; Yan, 2023), or their positive impact in academic writing (e.g., Kumar, 2023). Yan (2023) applied ChatGPT in a L2-writing practicum for higher education students. ChatGPT was considered an effective tool by students and allowed them to write with fewer grammatical errors and with a higher lexical diversity. However, ChatGPT was not effective in more demanding tasks, such as using it as the major information source. Interestingly, participants expressed more concern than satisfaction towards the use of ChatGPT in L2 writing, which is an interesting aspect to further analyze. Kumar (2023) has assessed the academic writing potential of ChatGPT on a series of queries (e.g., “Write a 5000-word report on Chocolate toxicity in dogs and cite references”). The responses were coherent with the topic, systematic, and original, yet with a few pitfalls: instructions were not well-followed, intext references were not always cited, references were overall inaccurate, lack of practical examples. As higher students’ ability to detect such deficiencies in ChatGPT output may be suboptimal, it is important to investigate how they interact with such a tool: whether text gets transferred from ChatGPT to their own written production or whether ChatGPT output is critically analyzed before being implemented in their own text.

Source-Based Writing

Education professionals agree on the significance of being able to manage tasks that involve writing from various sources. Grabe and Zhang (2013, p. 9) defined writing from sources as follows:

Learning to write from textual sources (e.g., integrating complementary sources of information, interpreting conceptually difficult information) is a challenging skill that even native speaking students have to work hard to master ... Tasks that require reading/writing integration, such as summarizing, synthesizing

information, critically responding to text input, or writing a research paper, require a great deal of practice.

Source-based writing activities are highly demanded in the university context (Beach et al., 2016). Students often deal with academic assignments that require reading one or more sources and evaluating their reliability to produce a written essay or report. Synthesis writing activities are an example of source-based writing tasks and require writers to comprehend the sources, judiciously extracting pertinent information, structuring the chosen data coherently, and generating an original text that integrates information across within and across sources (Klein & Boscolo, 2016; Vandermeulen et al., 2023). Spivey (1997) defined synthesis writing as a ‘hybrid’ activity that cannot be effectively accomplished by a simple sequence of reading followed by writing. Indeed, they demand a recursive approach wherein three interconnected processes contribute to the ultimate quality of the synthesis text: the act of selecting, the art of organizing, and the skill of connecting.

Synthesis writing tasks can be interpreted as a variant of argumentative writing. This stems from the need to thoroughly examine both perspectives (Mateos et al., 2018). Consequently, several strategies can be employed to interrelate arguments and counterarguments linked to different points of view (Nussbaum & Schraw, 2007). When the arguments supporting a position are considered erroneous or insufficiently substantiated, a counter-argumentation strategy should be employed. Conversely, an alternative approach might involve endorsing one of the perspectives following a thorough assessment and thoughtful contemplation of the arguments aligned with each stance (weighing strategy). The third and ultimate approach, as delineated by these scholars, entails a synthesis strategy, wherein the writer endeavors to propose a harmonious resolution that combines the favorable elements of the opposing viewpoints. It is worth noting that, while Nussbaum and Schraw (2007) encompass all these strategies within the realm of synthesis, they exclusively employ the term “synthesis” to denote the specific strategy of formulating a “conciliatory solution” to the addressed issue.

Collaborative Writing

Collaborative writing has been defined by Storch (2013, 2019) as an activity in which both authors are active throughout the whole process. When writing collaboratively, authors try to find the accurate phrasing by negotiating wording and grammar in ‘collaborative dialogues’ (Swain, 2000), thus expanding their language skills. Such deliberations are often called ‘language related episodes’ (Swain, 1998). Other concepts frequently used to describe the collaborative writing process are equality and mutuality (Damon & Phelps, 1989). Storch (2002) has operationalized these concepts by analyzing students’ discussions during joint writing tasks. “Equality” referred to an equal distribution of turn-taking in conversation, an equal contribution to writing, and an equal level of control over the task’s direction. “Mutuality” involved engagement with each other’s contributions, reflected in language functions such as confirming statements, elaborating on what someone had said, or seeking an expla-

nation if something was not understood. Based on her conversation analyses, Storch identified four different patterns of pair interactions:

- Collaborative (high equality and high mutuality)
- Expert-novice (low equality and high mutuality)
- Dominant-Dominant (high equality and low mutuality)
- Dominant-Passive (low equality and low mutuality)

Collaborative writing has been especially of interest in L2-research, where language growth is an issue. In a longitudinal study with EFL learners in Saudi Arabia by Storch and Aldosari (2013) the authors found that the language proficiency level of your partner matters. Pairs with similar language proficiency were more likely to collaborate, than pairs with mixed proficiency. The interaction between the authors was higher when their language skills matched.

With computers as a more or less standard tool in schools, face-to-face collaborations have decreased in frequency in favor of computer-supported collaborative writing. With that, the interaction patterns changed, and students focused more on using edit tools than discussing the organization of texts (Oskoz & Elola, 2014). When Strobl (2014) compared individual and collaborative writing in a synthesis task performed by proficient German L2 learners, she found no significant differences in accuracy or complexity in the final product, but the collaborative writing demonstrated better synthesis. The collaborative writing mode also involved intertwined writing and reviewing processes, in which students often discussed language-related issues. Interestingly, following-up interviews revealed mixed feelings among students, with most preferring individual writing.

A further step from computer-supported collaborative writing is collaboration with artificial intelligence (AI) of different types, although not relying on the use AI as LLM-based chatbots. Ouyang and Jiao (2021) present three paradigms of how AI is used in education. They define them as either AI-directed, in which the learner is the recipient, or AI-supported with the learner as a collaborator or AI-empowered in which the learner takes the position as the leader. However, Ouyang and Jiao (2021) lift a warning finger; AI in education is not equivalent to positive educational results and a high quality of learning. Rather, as suggested by Bower (2019), there is a need that both researchers and teachers think about technology in education from a holistic and integrated perspective, and not simplify it to ‘technology enhances learning’. Thus, a weak connection to theoretical pedagogical perspectives can be avoided (Zawacki-Richter et al., 2019).

The specific conversational mode of ChatGPT makes it a resource in writing, according to Beck and Levine (2023). For example, the AI does not suffer from writer’s block, it is not judgmental, and it can elaborate on the human’s writing input. For learners in higher education, it is a useful tool to for example generate summaries and outlines of texts (Kasneci et al., 2023) and with little effort retrieve a coherent paper (Zhai, 2022). Moreover, Wenzlaff and Spaeth (2022) found that ChatGPT is basically equivalent with human beings in composing explanatory answers to posed questions. Thus, it is interesting to explore whether collaborative writing human/AI can be con-

sidered as what Storch (2013) found to be the most fruitful pattern in collaborative writing, namely both sides showing/experiencing high equality and high mutuality.

The Present Study

In line with the current literature on the pedagogical potential of ChatGPT and according to the absence of previous experimental studies analyzing the impact of this artificial intelligence tool in addressing source-based writing tasks, we conducted the present exploratory study. The general objective of the study was to analyze the strategies employed by university students when approaching a source-based writing activity on a controversial topic with the support of ChatGPT.

We aimed at investigating whether participants had known and interacted with ChatGPT before the experiment and what kind of prompt they would spontaneously use in a source-based writing task. We were interested in both the strategic nature of the prompt itself and the reflection behind it. In other words, students may be strategic writers but not skilled prompt creators, or they may not even be aware of how ChatGPT could be strategically used. Therefore, we analyzed the actual prompts as well as students' reasons, captured through retrospective think-alouds. We also were interested in whether strategic interaction with ChatGPT would have led to written products of better quality. Finally, we were interested in the use students made of the text produced by ChatGPT, by analyzing the degree of overlap between the output produced by ChatGPT, after the student's collaboration with the tool, and the argumentative text generated.

The present study addressed the following research questions:

1. How do undergraduate students interact with ChatGPT in a source-based writing task?
2. Is a more strategic use of ChatGPT associated with a better quality of the texts produced by students?
3. How was the text produced by ChatGPT integrated in participants' essays?

As generative artificial intelligence has just recently started to be used on a large scale, research on this area is lacking, therefore we could not suggest specific hypotheses. Nevertheless, from what we know from the literature on source-based writing in higher education, we expected undergraduate students to interact with ChatGPT in a non-strategic way, merely interacting with it as they would interact with an online search engine. This expectation is based on studies suggesting how students struggle in source-based writing tasks (Tarchi & Villalón, 2021; Mateos et al., 2018). More specifically, the research hypotheses are:

- R1. Students apply only shallow strategies when interacting with ChatGPT.
R2. A strategic use of ChatGPT should be associated with a better quality of the texts produced by students.
R3. We expect students to include the text produced by ChatGPT in their essays without properly integrating it with their own essays and with the information they derive from the assigned source.

Method

Participants

Twenty-seven students participated in the study (1 male; age=26, and 26 females; age mean= 20.15 ± 1.95). All participants were first year psychology undergraduate students. They were Italian and spoke Italian as the primary language. No students reported to have a special educational need. The data collection procedure was conducted anonymously through the assignment of a code to each student. The study was approved by the Ethics Committee of the University of Florence (Italy).

Procedure

All research materials, instructions and texts were provided in Italian. Participants filled out a preliminary survey on the Qualtrics Platform, generating a personal ID granting them anonymous participation. All participants answered preliminary questions investigating their prior knowledge of the topic presented in the source, i.e., the use of Glyphosate, and their knowledge and frequency of use of the Artificial Intelligence-based conversational agent, ChatGPT. They also provided information about their frequency of use of ChatGPT and their level of agreement with the use of Glyphosate. In the preliminary phase, they completed an online questionnaire on writing motivation (the Academic Writing Motivation Questionnaire; Payne, 2012). In the second phase, they attended the experimental session, in the presence of a researcher. Participants were given a computer where they had access to the task instructions. They were asked to read the source text provided on Glyphosate, with the subsequent aim of writing a text for an informative scientific journal expressing their opinion on the use of this herbicide. Specifically, the students were given the following instructions:

“Imagine you are a journalist commissioned to write an informative article for a popular scientific journal based on the text you are about to read about glyphosate. The article you submit to the magazine should be approximately one A4 page in length in Microsoft Word, Calibri font size 12, with standard margins, no line spacing, double-spaced lines and justified formatting. In the article, you should inform the reader about the main topic, express your opinion about it and support your opinion with information from the text provided. While reading the text provided and writing your article, you will have access to the ChatGPT application with which you can interact freely. ChatGPT is a language model created by OpenAI. A language model is a type of artificial intelligence that can understand natural language and generate text similar to that produced by humans. ChatGPT has been trained on a vast corpus of text with data up to 2021 and can answer a wide range of questions and topics. It can be used for chatbots, virtual assistants and many other applications where natural language understanding is required. You can interact with ChatGPT as you wish, asking questions and using the text you receive in response to perform your typing task. You can only interact with ChatGPT. You cannot use other search engines or online applications” [courtesy translation from Italian provided by the authors].

Before conducting the reading-writing task, the students were allowed to practice the interaction with ChatGPT with a couple of task-unrelated questions to get acquainted with the tool. Participants were given one hour of time; the researcher supervised the session and informed participants about the time remaining every 15 min. The entire experimental session at the computer, comprehending the reading, writing, and AI interaction was recorded on a screen recording software. After the experimental session, participants attended a retrospective think-aloud session with the researcher, who observed the recorded interaction with ChatGPT and asked for the strategies and motivations behind the interaction with the AI tool. The think-aloud session was recorded and transcribed and linked to the ID of the participant for further analysis.

Materials and Measures

Source Text

For the reading-writing task, we used a source on a controversial topic, namely the use of Glyphosate as a chemical herbicide. The text was 804 words long and included arguments for and against its use. The text was created by the experimenters based on Internet sources from trustworthy associations (AIRC, *en. tr. Italian Association of Research on Cancer*), European institutions (EFSA, European Food Safety Authority), scientific publications (Journals: Pest management science and Weed science).

The arguments could refer to the following thematic categories: (1) general information about glyphosate, (2) the expansion of its use worldwide and in Italy, (3) available data and concerns about the possible danger to people's health from its use, (4) refutation of the data and concerns about its danger to health; (5) data and concerns about its effect on the environment; (6) refutation of the data about its danger to the environment; (7) policies regulating its use in Italy and the USA.

Perceived Prior Knowledge

Each participant was asked on a scale of 1 (no competence) to 5 (high competence) how they would rate their prior knowledge of Glyphosate.

Prior Topic Beliefs

Prior beliefs about Glyphosate were assessed using a Likert-type scale (1-fully disagree; 5-fully agree).

Familiarity with ChatGPT

Familiarity with ChatGPT was assessed by considering both prior knowledge of the tool and frequency of use. Knowledge was explored through an open-ended question, which was descriptively analyzed to investigate for what purposes ChatGPT had been used (e.g., never been used, used as a search engine, used to generate text, used

to revise text, and the like). To evaluate the frequency of ChatGPT use, a Likert-type scale (1- no use, 5- daily use) was employed.

Academic Writing Motivation

In order to assess students' motivation towards academic writing we used the Academic Writing Motivation Questionnaire (Payne, 2012). The AWMQ, in its completed version, consists of 37 items presented in a Likert-style format. Each item includes a statement designed to elicit participants' agreement levels, spanning from zero to four, with corresponding values as follows: 0=Strongly Disagree, 1=Disagree, 2=Uncertain, 3=Agree, and 4=Strongly Agree. Examples of items would be "*I complete a writing assignment even when it is difficult*" or "*It is easy for me to write good essays.*"

Interaction with ChatGPT

The interactions with ChatGPT were registered using a screen recording software, which made it possible to record all the actions conducted on the computer during the experimental task. The questions asked by the students to ChatGPT were coded following a deductive-inductive procedure. First, we included those categories aligned with the scientific literature on self-regulated writing processes and linked to source-based writing tasks (e.g., Graham et al., 2017; Nussbaum & Schraw, 2007; Weston-Sementelli et al., 2018). We then examined the prompts posed to ChatGPT and included those categories consistent with student behavior. The questions to ChatGPT were finally organized and marked according to six main categories: (1) Content Knowledge (targeting Glyphosate); (2) Argumentative Writing Strategies; (3) Task Dependent Questions (e.g. information on how to structure an argumentative text); (4) Questions about the correct use of vocabulary; (5) Metacognitive strategies on writing process; (6) Full output (requests to ChatGPT to provide ready-to-use text that meets the demands of the experimental task). It should be mentioned that in case of no questions to ChatGPT, the researcher asked for motivations for not having used the tool. The students' questions to ChatGPT were coded via Qcamap (Fenzl & Mayring, 2017) by two independent coders, with good inter-coder agreement ($k=0.95$).

Awareness of ChatGPT Use

Awareness of ChatGPT use was assessed through a retrospective think-aloud protocol, which was transcribed and coded following a category system analysis. Participants' responses on their reasons for using or not using ChatGPT were organized into the following categories: (1) awareness about text structure; (2) ChatGPT considered as more trustworthy than the source; (3) confirming prior beliefs; (4) ethic responsibility on text authorship; (5) fear of using an unknown tool; (6) search for new words; (7) not having any question to ask; (8) low expertise in the usage of ChatGPT; (9) searching for new perspectives different from prior beliefs; (10) not sufficient knowledge on the topic; (11) searching for text integration strategies; (12) searching vocabulary; (13) short available time; (14) social human-like interaction with AI; (15)

searching for strategies to improve readability; 16) testing the validity of the information provided in the source-text. Again, Qcemap (Fenzl & Mayring, 2017) was used to code the think-aloud protocol for student responses on strategies and motivations that justified their interaction with ChatGPT. Two independent coders conducted the analysis of the protocol, reaching a high inter-judge agreement ($k=0.91$).

Student-Written Text

The texts written by the students were uploaded to Qcemap for coding and analysis. These texts were coded along three dimensions, which are explained below.

The argumentative text produced by the students was coded taking into account the number of arguments coming from the source-text on glyphosate and the thematic category they referred to; i.e.: (1) general information about glyphosate, (2) the expansion of its use worldwide and in Italy, (3) available data and concerns about the possible danger to people's health from its use, (4) refutation of the data and concerns about its danger to health; (5) data and concerns about its effect on the environment; (6) refutation of the data about its danger to the environment; (7) policies regulating its use in Italy and the USA.

Argument-counterargument integration. After reading the source text on Glyphosate, the students were asked to write an informative article for a popular scientific journal, expressing their opinion on this controversial topic. This argumentative text was evaluated by taking into account the degree to which students were able to integrate arguments and counterarguments about the use of this chemical herbicide. We assessed the level of integration of the writings applying a coding system from 1 to 6 levels (developed by Mateos et al., 2018). The levels of integration are listed in Table 1.

Two independent judges coded the 30% of argumentative texts produced by the students to calculate inter-rater reliability with respect to the level of integration ($k=0.88$).

The argumentative texts produced by the participants were also digitally analyzed with the open-source Tint NLP suite, an extension of Stanford CoreNLP (Manning et al., 2014) written to support Italian, freely available on GitHub (<https://github.com/dhfbk/tint>). Tint is provided also with a text reuse component, which is able to identify and quantify the portions of an input text that overlap with a given docu-

Table 1 Coding system to assess arguments-counterarguments integration

Score	Argumentative strategy
0	Personal opinion not based on source text
1	Does not take a clear position.
2	Argues in support of one perspective only (pro/against Glyphosate)
3	Integration via refutation
4	Minimum integration via weighing or synthesizing. No or partial conclusion.
5	Average integration via weighing or synthesizing. Partially integrated conclusion.
6	Maximum integration via weighing or synthesizing. Conclusion with overall integration.

ment. The text matching a given source is compared with the source text using the FuzzyWuzzy package, a Java fuzzy string-matching implementation. This kind of comparison allows the system to match expressions that are identical, but also those that are slightly different in terms of word forms. If a match is found, a fuzziness score, between 0 and 100, is computed, corresponding to the percentage of overlap between the string from the essay and the source material. In the current study, we have set the fuzziness score at 70 (a standard score for string similarity), and the minimum proportion of matching characters to be considered for the count of matching characters at 0.4.

Adopting the FuzzyWuzzy package, the number of matching characters has been calculated for each participant with respect to the source- text, and with respect to the answers provided by ChatGPT. The analysis allows us to quantitatively measure the number of characters that have been copied either from the source or from ChatGPT in the output text produced by the participants. The data are informative with respect to the writing strategies, and the usage the students make of the output of ChatGPT.

Results

Descriptive Statistics of the Participants

Analysis of the students' responses to the questions on "perceived prior knowledge" and "prior beliefs" about Glyphosate confirmed that students did not perceive themselves as knowledgeable on the topic ($M=1.4 \pm 0.64$), and, in general terms, displayed unfavorable attitudes towards the use of Glyphosate as a chemical pesticide ($M=1.74 \pm 0.81$).

When asked to describe what ChatGPT is, prior to any instruction about the experimental task, 21 participants reported not to have any prior knowledge about ChatGPT; 5 participants out of 27 were able to provide a simple answer indicating they knew the existence of the tool and that it can be used to gather information, but provided a description comparable to the one of a search engine such as a Google; only 1 participant defined ChatGPT as a tool able to generate text. All participants declared they had never used ChatGPT prior to the experiment.

The scores in the Academic Writing Motivation Questionnaire (AW_Tot) suggest a medium-high level of motivation in writing among our sample ($M=123.81 \pm 16.34$, maximum score possible=148). The total score of the Academic Writing scale significantly positively correlated with the percentage of categories transferred from the original source to participants' written essays ($r=0.52$; $p=0.006$).

Following, we will present results from students' interaction with ChatGPT, retrospective think-alouds and written essays.

Interaction with ChatGPT

When interacting with ChatGPT, 11.83% participants mainly asked for Vocabulary; 10.75% of participants did not request any specific information from ChatGPT; 6.45% asked for suggestions to better complete the task of writing an Argumentative

Text for a non-academic scientific journal; 3.23% of participants asked for suggestions on Writing Strategies; 2.15% asked for suggestions to improve their Argumentative Strategies in writing. No participant requested ChatGPT to produce the full output of the text fulfilling the experimental task.

We also calculated the total number of questions asked to ChatGPT, even if questions about the same category were asked (e.g., vocabulary clarification questions). Overall, participants interacted with ChatGPT 3.07 times ($SD=2.13$).

After completing the source-based writing task, subjects were asked to retrospectively think-aloud about why each prompt was used in ChatGPT. The participants mainly used ChatGPT to: collect further information about the topic of the source (Glyphosate) (36.75%), followed by the intention to search for integration strategies (11.97%); the intention to challenge the information available in the source (10.26%); the search for a new perspective on the conflictual topic presented in the source about the usage of Glyphosate (8.55%); the search for appropriate vocabulary (8.55%), the search for information about the structure of the text requested by the task (6.41%) and the search of information confirming prior beliefs (5.13%). Among the less frequent answers in the think-aloud session, participants also reported they asked ChatGPT for more information considering it more trustworthy than the source provided in the task (2.56%). Finally, some participants did not request any information to ChatGPT, except for a few questions on vocabulary (2.56%). When asked about the reasons for not having extensively interacted with the tool, few participants declared not to have had any question to ask. One participant only declared to have searched for synonyms not to be too repetitive in their essay. Finally, a set of answers in the think-aloud sessions addressed some feelings of uncertainty and self-perceived low competence in completing the experimental task: one participant considered themselves not sufficiently skilled to extensively use ChatGPT; one participant evaluated the given time to conclude the task (1 h) as too short to explore the tool. One participant requested some information to ChatGPT mainly to experience the use of an AI since it was their first time. One participant searched for strategies to improve readability. Finally, one participant declared to be afraid of the new tool, therefore decided not to use it. Remarkably, when asked about why ChatGPT was not used at all (2.6% of the participants), they reported ethical issues in solving the task through the facilitation of AI. Overall, participants reported an average of 4.59 ($SD=2.79$) strategic reasons associated with their interactions with ChatGPT.

Analysis of Written Outputs

The main text about the Glyphosate included seven informational sections. In this section we report the percentages of transfer of each section from the source to the output text of the participants. Participants' texts mainly targeted a description of Glyphosate and its use (27.75%), the reported danger to people's health (22.51%), and the confutation of health danger data (20.94%). The other sections of the source were considered less relevant in the process of writing: a minority of students' essays targeted the worldwide expansion of use of Glyphosate along the years (5.24%); a small percentage of produced essays targeted the risks and reported issues for the environment in the usage of Glyphosate (7.85%) and a slightly higher percent-

age reported the confutation of the argumentation on the risks for the environment (8.33%); finally a small percentage of produced essays targeted the policies adopted in the USA to contain the use of Glyphosate (7.33%).

We also counted the total number of times a source information was mentioned in students' written outputs, thus including categories mentioned several times. Overall, students referred to literal information from the source 7.07 times ($SD=2.91$).

On average, participants produced texts with a low level of integration ($M=3.19 \pm 2.02$). This result is consistent with previous studies suggesting that students struggle with intertextual integration even at the university level (Tarchi & Villalón, 2021; Mateos et al., 2018). Students are generally exposed to argumentative writing during secondary schooling but are very rarely explicitly taught about intertextual integration strategies. The analysis of distribution of scores across integration categories (see Fig. 1) shows that most students either did not integrate at all (levels 1 and 2) or implement high-integration strategies, i.e., weighting or synthesis.

We also investigated the extent to which participants included literal information from the original source (matching percentage with source) and from ChatGPT (matching percentage with ChatGPT) in their essays. The two scores are somewhat similar, as participants "plagiarized" 18.28% ($SD=20.61$) of the original text and 20.06% ($SD=24.52$) of the ChatGPT output.

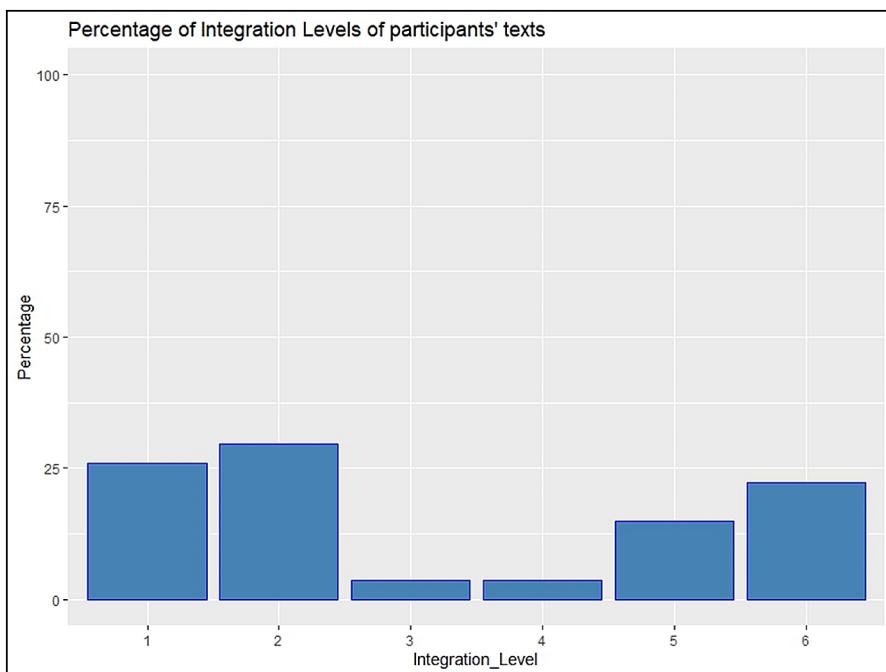


Fig. 1 Percentage of participants for each integration level

Associations Between Use of ChatGPT and Quality of Written Essays

The intertextual integration score was significantly not associated with any variables related with interactions with ChatGPT or think-aloud reflections. The total number of information categories transferred from the source to participants' written essays negatively correlated with the matching percentage with ChatGPT ($r=-0.40$; $p=0.034$) and positively correlated with the matching percentage with source ($r=0.53$; $p=0.005$). The number of questions asked to ChatGPT significantly correlated with the matching percentage with ChatGPT ($r=0.61$; $p<0.001$), and negatively correlated with the number of information categories transferred from the source text to the participants' written essay ($r=-0.40$; $p=0.04$).

Overall, these correlations suggest that when interacting with ChatGPT, our participants mainly used its output in their texts in a literal way, without transforming the text to integrate it with the other information. When using ChatGPT our participants made less references to the original source.

Annotated Examples from Students' Interaction with ChatGPT, Think-Alouds and Essays

In appendix A we included examples of the three outputs of this study (interaction with ChatGPT, think-aloud and essays) from two subjects.

The first subject wrote an essay in which they first described how glyphosate began to be used as a pesticide (“*due mainly to the increase in population worldwide, there has been an increasing demand for food, so new agricultural techniques have been developed and devised. One of these was the creation in the 1970s of a particular herbicide called glyphosate*”), then explained how its use in agriculture is reasonably safe (“*the results showed that it was low in danger because it was present in small amounts in the samples analyzed*”), and finally it concluded by supporting its use (“*The use of glyphosate as an herbicide is therefore reasonably safe*”). The essay is overall one-sided as it supports one perspective only (glyphosate as a useful pesticide) with just a minor mention of the opposing perspective (“*It may be right, as in Italy, that farmers should not use it in busy urban areas because it has been classified as a “probable carcinogen,”*”), which is immediately after refuted (“*but its dangerousness, being compared to that of red meat or household chimney fires, I think we need not be overly concerned about it*”).

While writing the essay the subject asked only two questions to ChatGPT (a synonym and a date), which underlie a rather superficial strategic use of this tool, an interpretation confirmed by the retrospective think-aloud (“*I needed precisely a synonym, since I did not want to make a repetition within the sentence of the same verb and so, in short, I had this need*”).

In their essay, the second subject defined what glyphosate is (“*Glyphosate is a chemical herbicide that is used in agriculture to eliminate unwanted plants*”), discussed its positive effects (“*The importance of glyphosate lies in its efficacy*”), presented results about its toxicity (“*A second study was done by collecting data from rat experiments, through which they deduced that this substance was capable of causing cancer*”) and presented their own conclusion (“*the results of this research make us*

understand that the use of glyphosate is not carcinogenic to human beings and is not toxic to environmental health, so its use appears to be safe"). The essay is characterized by a higher level of intertextual integration than the first example, as both pro-glyphosate and against-glyphosate arguments are presented and given consideration.

When interacting with ChatGPT this subject took a more strategic approach, asking ChatGPT how to structure their essay ("How should an article for a popular magazine be set up?"), and to provide some more information about glyphosate ("What is the importance of glyphosate?"). The think-aloud confirmed the goal-oriented use of ChatGPT ("I thought about How could I precisely set up an article for a popular magazine. I really wanted to have an outline").

Discussion

Artificial intelligence has been part of our lives for several years, but only recently it has been brought to people's attention. The year 2023 has witnessed the explosive growth of generative AI tools among the general public. ChatGPT is one vivid example of this effect. ChatGPT gained one million users in its first week after launch and has a currently estimated 100 million active users (<https://www.tooltester.com/en/blog/chatgpt-statistics/>). ChatGPT represents both a threat and a resource of writing education, although the first aspect is more stressed than the second one. This depends also on the fact that there is a lack of well-developed guidelines and ethical codes around generative AI in the education sector (Dwivedi et al., 2023). The issue is complicated by the fact that students are struggling with source-based writing tasks even at the higher education level (Tarchi & Villalón, 2021). To this end, in this study we adopted an educational perspective to investigate the strategic use of ChatGPT in a source-based writing task in higher education students. We were interested in whether students use ChatGPT strategically, whether doing so leads to better writing and what use is made of the text generated by ChatGPT. We framed the study as AI-collaborative writing.

Overall, the analysis highlights that the group of participants taking part in the study, being both non-experts in the use of ChatGPT and non-expert on the topic of the impact of glyphosate on the environment and the public health, made limited use of all the potential benefits of the use of ChatGPT while writing argumentative texts. The collected samples of written essays that, overall, mainly described the function of glyphosate and its implications for public health, also considering the counter-positions in favor of its use, and only a minor part of the sample also addressed the concerns for pollution and the environment. The participants copied (if not plagiarized) around 20% of their text that they produced from either the source or from ChatGPT. They mainly used ChatGPT to solve problems related to the search of vocabulary, not exploiting the full potential of the tool. Among the reasons for using ChatGPT the main motivation showed to be related to the lack of knowledge on the topic of the text, and in the second instance, to the search for an integration strategy of conflicting information provided in the source. No participant used ChatGPT to request the full production of a finished output, resolving the experimental task, or to help them advance in the writing task. Of the all sample, two participants did not use

the tool, and one-third of the sample made very limited use of the tool; when asked about the reasons for not fully exploiting the AI tool, the answers target both a low level of expertise in using digital tools, and as well ethic concerns in allowing the AI to solve the task instead of the participant.

Overall, our results provide some emerging indications on how ChatGPT is used by non-expert writers. Despite some students reporting resistance in using ChatGPT because of ethical concerns, our impression is that students struggle understanding how ChatGPT can support them in their task because they are not strategic writers. Students seem to lack awareness on how to deal with controversial issues and integrate perspectives across sources, and write, in most cases, one-sided essays with limited references to the texts.

The results of the present study are influenced by the research setting. Indeed, task instructions and research design can influence learner's task model, which is their mental representation of the task, including the set of goals and plans that drives the learner's decisions and actions in reading. Firstly, this was a source-based writing task: participants were assigned a source to read before writing an essay. Reading a source provides some knowledge to participants, which may have inhibited their need for more information and, consequently, their usage of ChatGPT. Moreover, the task was timed, which may have induced readers to adopt a performance learning goal (trying to complete the task as fast as possible) rather than a mastery learning goal (trying to complete the task as well as possible).

Conclusion

Despite being a small-scale preliminary study, we analyzed in depth how non-expert writers interact with ChatGPT in a source-based writing task. Our data suggest a few reflections that can drive the design of large-scale quantitative studies. Firstly, academic writing motivation was associated with task performance, suggesting that intrinsic motivation is a key-element even when collaborating with AI in a writing task. Secondly, participants seemed hesitant to use ChatGPT and, when doing so, they mostly used it as an Internet source to simply transfer content from there to their written output. Moreover, by doing so, they referred less to the assigned source.

The research on the use of generative AI in education is still in its emerging phase. While several guidelines have been created and shared in educational platforms, evidence-based research on ChatGPT as a writing partner is missing. Current AI tools generating text are capable of multiple tasks that can scaffold students' reading and writing competences. Concerning reading, such AI tools can, for example, generate summaries or outlines of complex texts, to facilitate their comprehension, or provide information about a topic, to suggest the reader other aspects of the issue they are investigating that they may have neglected. Concerning writing, such AI tools can, for example, provide suggestions about the structure of their written output, provide feedback on what they have written, proofread in case they are writing in L2. While AI could be valuable in education, its inaccuracy requires users to use it with critical thinking.

Despite being exploratory, the study presents educational implications. It shows that the potential of ChatGPT in supporting writing competences is not fully understood by students. Part of the problem may depend on some ethical aspects (i.e., using ChatGPT feels like cheating), but most of the problem may depend on students' competence in argumentative thinking and writing. Low levels of metacognitive awareness or intertextual integration skills do not allow students to create effective prompts and interact with ChatGPT in an evidence-based way. Future research should further inquiry evidence-based implementations of ChatGPT in writing tasks assigned in educational settings. Firstly, the present research design could be replicated by adding a comparison group of expert writers. Secondly, we suggest evaluating the efficacy of an intervention in which students are provided with strategic prompts that can be used in ChatGPT. Future versions of the study could include training sessions and/or synthesis writing and effective ChatGPT usage to better assess ChatGPT's impact on writing skills.

We suggest giving priority to studies investigating how students can be taught to use ChatGPT as a partner when writing, by collaborating with it rather than non-critically using its content in their written reflections. One path forward if we really want students to improve may be to construct a task where the product is not a text but efficient prompts. That is, target the ability we want to enhance (i.e., writing prompts) and assess both, whether the ability itself improves and whether such improvement is associated with the writing of better texts.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40593-024-00413-1>.

Author Contributions **CT:** Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **AZ:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft. **LCL:** Conceptualization, Methodology, Formal analysis, Resources, Writing - Original Draft. **EWB:** Conceptualization, Writing - Original Draft.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. The study was co-funded by the Erasmus+ programme of the European Union (2022-1-IT02-KA220-HED-000086063; Project ORWELL “Online Reflective Writing in higher Education and Lifelong Learning”).

Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Data Availability Data are available upon request.

Declarations

Ethics Approval Ethics approval for this study was obtained by the University of Florence.

Non-financial and financial interests The authors have no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamopoulou, E., & Moussiades, L. (2020). An overview of Chatbot technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (Eds.), *Artificial Intelligence Applications and Innovations*. AIAI 2020. IFIP Advances in Information and Communication Technology, 584. Springer. https://doi.org/10.1007/978-3-030-49186-4_31.
- Baker, T., & Smith, L. S. (2019). *Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta.
- Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36, 584–607. <https://doi.org/10.1080/09588221.2021.1936071>.
- Baykasoglu, A., Ozbel, B. K., Dudakli, N., Subulan, K., & Senol, M. E. (2018). Process mining based approach to performance evaluation in computer-aided examinations. *Computer Applications in Engineering Education*, 26, 1841–1861. <https://doi.org/10.1002/cae.21971>.
- Beach, R., Newell, G., & VanDerHeide, J. (2016). A sociocultural perspective on writing development: Toward an agenda for classroom research on students' uses of social practices. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 88–101). Guilford Press.
- Beck, S. W., & Levine, S. R. (2023). Backtalk: ChatGPT: A powerful technology tool for writing instruction. *Phi Delta Kappan*, 105, 66–67. <https://doi.org/10.1177/00317217231197487>.
- Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology*, 50, 1035–1048. <https://doi.org/10.1111/bjet.12771>.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>.
- Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research*, 13, 9–19.
- Dever, D. A., Azevedo, R., Cloude, E. B., & Wiedbusch, M. (2020). The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes. *International Journal of Artificial Intelligence in Education*, 30, 581–615. <https://doi.org/10.1007/s40593-020-00215-1>.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albasrahi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). Opinion Paper: So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-Augmented teaching and learning practice. *Journal of University Teaching & Learning Practice*, 20. <https://doi.org/10.53761/1.20.5.02>.
- European, & Parliament (2019). Directorate-General for Parliamentary Research Services. In T. Meyer, & C. Marsden (Eds.), *Regulating disinformation with artificial intelligence— effects of disinformation initiatives on freedom of expression and media pluralism*. European Parliament. <https://doi.org/10.2861/003689>.
- Fenzl, T., & Mayring, P. (2017). QC Amap: Eine interaktive webapplikation für qualitative inhaltsanalyse. *Zeitschrift für Soziologie Der Erziehung Und Sozialisation ZSE*, 37, 333–340.
- Floridi, L. (2023). AI as Agency without Intelligence: On ChatGPT, large Language models, and other Generative models. *Philosophy & Technology*, 36, 15. <https://doi.org/10.1007/s13347-023-00621-y>.
- Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4, 9–24. <https://doi.org/10.1002/tesj.65>.
- Graham, S., Harris, K. R., MacArthur, C., & Santangelo, T. (2017). Self-regulation and writing. *Handbook of Self-Regulation of Learning and Performance*, 138–152. <https://doi.org/10.4324/9781315697048-9>.

- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15, ep464. <https://doi.org/10.30935/cedtech/13605>.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>.
- Klein, P. D., & Boscolo, P. (2016). Trends in research on writing as a learning activity. *Journal of Writing Research*, 7, 311–350. <https://doi.org/10.17239/jowr-2016.07.03.01>.
- Kumar, A. H. (2023). Analysis of ChatGPT tool to assess the potential of its utility for academic writing in biomedical domain. *Biology Engineering Medicine and Science Reports*, 9, 24–30. <https://doi.org/10.5530/bems.9.1.5>.
- Lametti, D. (2022). AI could be great for college essays. *slate.com*. <https://slate.com/technology/2022/12/chatgpt-college-essay-plagiarism.html>.
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40, 26–29. <https://doi.org/10.1108/LHTN-01-2023-0009>.
- Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). Chatgpt and a new academic reality: Artificial intelligence-written research papers and the ethics of the large Language models in scholarly publishing. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24750>.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). *The Stanford CoreNLP natural language processing toolkit*. <https://aclanthology.org/P14-5010>.
- Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction*, 36, 119–138. <https://doi.org/10.1080/07370008.2018.1425300>.
- Nussbaum, M. E., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, 76, 59–92. <https://doi.org/10.3200/JEXE.76.1.59-92>.
- Oskoz, A., & Elola, I. (2014). Promoting foreign language collaborative writing through the use of web 2.0 tools and tasks. In M. González-Lloret, & L. Ortega (Eds.), *Task-based language teaching* (vol. 6) (pp. 115–148). John Benjamins.
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caai.2021.100020>.
- Payne, A. R. (2012). *Development of the academic writing motivation questionnaire*. Doctoral dissertation, University of Georgia.
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11, 887. <https://doi.org/10.3390/healthcare11060887>.
- Spivey, N. (1997). Reading, writing and the making of meaning. The constructivist metaphor. *Academic Press*. <https://doi.org/10.2307/358470>
- Stacey, S. (2022). Cheating on your college essay with ChatGPT. *Business Insider*. <https://www.businessinsider.com/professors-say-chatgpt-wont-kill-college-essays-make-education-fairer-2022-12>.
- Stock, L. (2023). *ChatGPT is changing education, AI experts say—but how?* DW.com-science-global issues. <https://www.dw.com/en/chatgpt-is-changing-education-ai-experts-say-but-how/a-64454752>.
- Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52, 119–158. <https://doi.org/10.1111/1467-9922.00179>.
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters.
- Storch, N. (2019). Collaborative writing. *Language Teaching*, 52, 40–59. <https://doi.org/10.1017/S0261444818000320>.
- Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research*, 17, 31–48. <https://doi.org/10.1177/1362168812457530>.
- Strobl, C. (2014). Affordances of web 2.0 technologies for collaborative advanced writing in a foreign language. *CALICO Journal*, 31, 1–18. <https://www.jstor.org/stable/10.2307/calicojournal.31.1.1>.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.aww.2023.100752>.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition*. Cambridge University Press.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning*. Oxford University Press, 97–114.

- Taecharungroj, V. (2023). What can ChatGPT do? Analyzing early reactions to the innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing*, 7, 35. <https://doi.org/10.3390/bdcc7010035>.
- Tarchi, C., & Villalón, R. (2021). The influence of thinking dispositions on integration and recall of multiple texts. *British Journal of Educational Psychology*, 91, 1498–1516. <https://doi.org/10.1111/bjep.12432>
- Vandermeulen, N., Van Steendam, E., & Rijlaarsdam, G. (2023). Introduction to the special issue on synthesis tasks: Where reading and writing meet. *Reading and Writing*, 36, 747–768. <https://doi.org/10.1007/s11145-022-10394-z>.
- Verdú, E., Regueras, L. M., Gal, E., de Castro, J. P., Verdú, M. J., & Kohen-Vacs, D. (2017). Integration of an intelligent tutoring system in a course of computer network design. *Educational Technology Research and Development*, 65, 653–677. <https://doi.org/10.1007/s11423-016-9503-0>.
- Wenzlaff, K., & Spaeth, S. (2022). Smarter than humans? Validating how OpenAI's ChatGPT model explains crowdfunding, alternative finance and community finance. *SSRN Scholarly Paper*. <https://doi.org/10.2139/ssrn.4302443>.
- Weston-Sementelli, J. L., Allen, L. K., & McNamara, D. S. (2018). Comprehension and writing strategy training improves performance on content-specific source-based writing tasks. *International Journal of Artificial Intelligence in Education*, 28, 106–137. <https://doi.org/10.1007/s40593-016-0127-7>.
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: an exploratory investigation. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-11742-4>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>.
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Scholarly Paper*. <https://doi.org/10.2139/ssrn.4312418>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.