Writing with, without and against the AI. Insights from a high school writing lab

Scrivere con, senza e contro l'IA. Cosa abbiamo imparato da un laboratorio di scrittura nella scuola media superiore

LUCA BOTTURI[A*], LUCA CIGNETTI[A], SILVIA DEMARTINI[A]

[A] *Department of Education and Learning / Graduate School of Pedagogy, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Locarno, Switzerland, luca.botturi@supsi.ch*, luca.cignetti@supsi.ch, silvia.demartini@supsi.ch*

* Corresponding author

ABSTRACT AI is enabling new practices to potentially extend education, but if such practices enhance or endanger learning is still under discussion. This study presents the findings of an explorative research conducted in a sophomore high school class in Italian-speaking Switzerland, aimed at investigating the impact of ChatGPT on fiction writing education. Students were provided with narrative stimuli and asked to write a short story on their own or with the support of ChatGPT 3.5. The resulting corpus was enriched with stories generated by ChatGPT alone. The analysis followed a dual quantitative and qualitative approach, examining the short stories in combination with feedback from authors and from a reading group composed of students of the same age. The results suggest clear differences among the three types of stories and a shift in focus during the writing process for students who used ChatGPT.

KEYWORDS ChatGPT; Creative Writing; Textual Analysis; High School.

SOMMARIO L'intelligenza artificiale rende possibili nuove pratiche educative, ma resta aperta la discussione su quanto tali pratiche migliorino o frenino l'apprendimento. Questo articolo espone i risultati di una ricerca esplorativa condotta in una classe del secondo anno di liceo nella Svizzera italiana, che ha esplorato l'impatto di ChatGPT sull'insegnamento della scrittura creativa. Dopo aver fornito stimoli creativi, agli studenti è stato chiesto di scrivere un racconto breve, individualmente oppure con il supporto di ChatGPT 3.5. Il corpus

risultante è stato ampliato includendo anche testi generati esclusivamente da ChatGPT. L'analisi ha seguito un approccio misto, quantitativo e qualitativo, esaminando i racconti in relazione ai feedback degli autori e di un gruppo di lettura composto da coetanei. I risultati evidenziano differenze significative tra i racconti dei tre tipi, nonché un cambiamento di prospettiva nel processo di scrittura da parte degli studenti che si sono avvalsi del supporto di ChatGPT.

1. Introduction

AI technologies have reached many domains of human activities. In particular, Generative Artificial Intelligence (GenAI, i.e., algorithms that generate new digital content like texts, images, video or audio files) and of Large Language Models (LLM, i.e., AI systems trained on vast amounts of text data to understand, generate, and manipulate human language) are changing many professional domains, including education. To harness their potential and control risks, re-thinking educational practices at all school levels is required, from homework (Ibrahim et al., 2023) to theses (Lopes, 2024) and from cheating (Lee et al., 2024) to plagiarism detection (Hutson, 2024). At its core, the challenge is reshaping educational contexts to help learners and teachers develop competences for living in an interconnected world populated by AI agents (Markauskaite et al., 2022).

LLMs are basically text-generation machines, and the concerns expressed above have a direct impact on the teaching and learning of writing skills. This paper reports an exploratory case study on the use of ChatGPT 3.5 for fiction writing in a high school class writing lab. Our purpose is twofold: on the one hand, exploring AI-assisted writing processes and experiences; on the other hand, identifying possible ways to exploit LLM-based tools to foster writing skills development.

2. State of the art

The writing process is deeply connected with human cognition and entails (or should entail) multiple phases and different skills, as shown since Flower and Hayes (1981), which go under the umbrella of "writing skills". Writing includes both authorial skills like identifying a context or situation in which the text makes sense, selecting content, organizing content, and choosing a style; and technical skills, like spelling, punctuation and grammar (Smith, 1983; 2013).

Writing is also a dynamic process that evolves in response to the development of the tools and the means through which it will be carried out. The Western contemporary society, which is deeply permeated by writing and by the widespread availability of increasingly sophisticated writing technologies, inevitably shapes both individual and educational writing practices (Williams & Beam, 2019). The recent large-scale accessibility of GenAI calls for a reconsideration of both writing opportunities and of the modalities through which writing is produced, for example in the form of assisted writing (a human author writes with the help of an agent) or delegated writing (a human delegates text production to an agent). Although published a few years ago, Sassoon (2019) provides relevant insights for the issue explored in our research. The study examines AI-driven storytelling in journalism, entertainment, and marketing, questioning its impact on authorship, originality, and reader engagement. It also

addresses ethical concerns, including bias, authenticity, and the role of human creativity in an AI-mediated narrative discourse.

Human and AI interactions in writing, both in education and in the professions, is a new research domain in which multiple strands are being explored. One strand focuses on the quality of texts generated by GenAI-tools, often through comparison with texts by human authors, with controversial results (Sardinha, 2024; Gunser et al., 2022). Kindenberg (2024) conducted a comparative analysis between historical narratives generated by ChatGPT and those written by secondary school students, assessing both historical elements and narrative quality. He concluded that, while ChatGPT's narratives show better stylistic features, they lack the emotional depth present in human-generated narratives.

A second strand of research explores AI-supported writing, or the interactions of human authors with GenAI-tools in writing. Most research in this strand focuses on second language learning, and especially ESL (English as Second Language; Fahti & Rahimi, 2024; Song & Song, 2023). Ghajargar et al. (2022) investigate human-AI co-writing in fiction using GPT-3. Through two autoethnographic studies with varying narrative complexity, the authors examined the system's limitations and emergent capabilities, revealing its impact on creative practices and human-AI collaboration. Evidence so far clearly indicates that AI-assisted writing provides benefits in terms of quality of the textual output (at least for some text types), efficiency and motivation. On the other hand, in the context of writing education, the impact of GenAI-tool is still unclear (Abdullayeva & Musayeva, 2023). Whether the use of AI improves writing quality (i.e., the outcome text) or contributes to the actual development of students' writing skills is still under discussion (Daulay et al., 2024; Ironsi & Solomon Ironsi, 2024).

Research in this domain is of course still in progress, confronted with a technological landscape that is also undergoing rapid and in large part unpredictable evolution. New players are entering the market (e.g., the newcomer DeepSeek or Grok), and there is no consensus about the actual effectiveness of LLMs in different languages (Huang et al., 2023). While a few significant contributions are available, further investigations are necessary, especially for languages other than English like Italian.

In this study, we examine a classroom-based creative writing experience involving native Italian-speaking (L1) high school students, with and without the support of ChatGPT 3.5. The focus of our study is both on text quality, as perceived by human readers and through the results of textual analysis, and on the writing experience.

3. Methods

3.1. Research design

The study was conducted with a sophomore class attending a writing lab at a state high school in an urban area in Southern Switzerland. The vehicular language, and the language of writing, was Italian. Before the writing lab, all students already knew and used the free version of ChatGPT, even if with different frequencies and purposes. None of them had a pro subscription.

The 3-session writing lab was jointly facilitated by the Italian teacher of the class and the three authors of this paper. It was presented as a challenge: students would compete with ChatGPT in a mixed-mode fiction writing challenge. Writing fiction is a common topic in high school writing labs. Fiction writing fosters students' ability to shift perspectives and to construct, through imagination, alternative worlds (Lavinio, 1995; Cignetti, 2018; Cignetti et al., 2022). The experimentation workflow is illustrated in Figure 1.
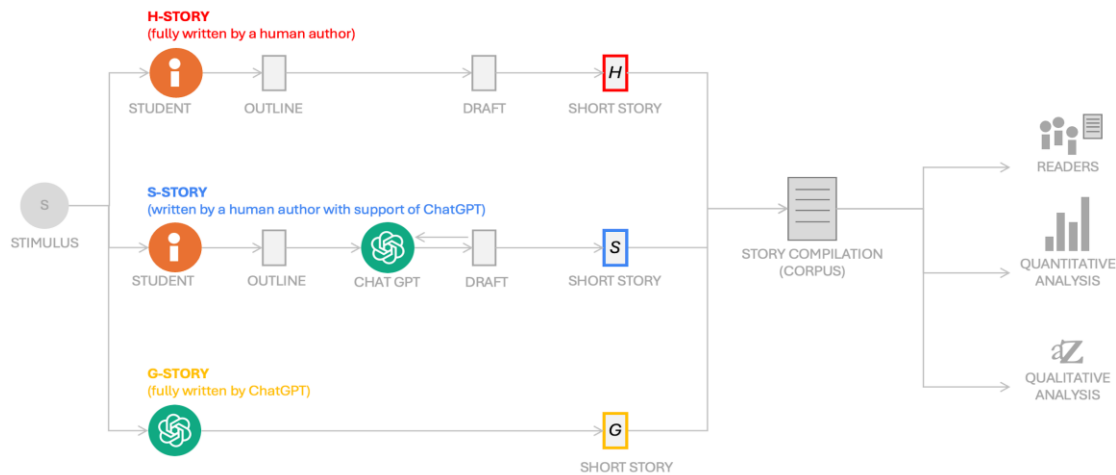
Figure 1. Experimentation workflow

The three sessions were organized as follows:
- *Session 1.* After a short introduction to LLMs, the challenge was presented. Each student received one of three writing stimuli, which consisted of a pair of cards from the *Work Kit Fiction Design* by NearFutureLab, portraying one object and one action (Figure 2), that had to be included in the short story; potential readers would be young people in their age group. Students were asked to develop an outline of the story focusing on the main character, a key challenge, and a resolution. They could then start writing a 5000-character fiction text in either mode: (a) asking ChatGPT to write the short story for them and then refining it, or (b) without the use of ChatGPT. The session was so designed, that at least two students for each stimulus wrote in each mode.
- *Session 2.* The second session was devoted to individual work for completing the story and submitting it. Students were then asked to fill in a Writer Questionnaire about their writing experience during the lab (Table 1). The survey was independent from stimulus and writing mode and addressed focus and satisfaction; surveys could be paired with short stories through an alphanumeric code, so that anonymity by design was guaranteed. No personal data was collected.
- *Session 3.* In this session, students were handed out a compilation of all short stories. The compilation included three additional short stories generated by ChatGPT alone, using the prompt in Figure 3. Some were read aloud to the class for discussion. The researchers illustrated the results of the quantitative and qualitative analyses and discussed them with the students.

The class was composed of 13 students, but only 12 actually attended the lab and one did not submit the short story. The resulting short story corpus was composed of 14 short stories (Table 2).

Figure 2. Sample writing stimulus

ITALIAN

```
Scrivi un racconto di 2 facciate (circa 5000 battute). I tuoi
lettori sono ragazzi e ragazze tra i 15 e i 18 anni,
appassionati lettori.
Nel racconto devono essere presenti:
- l'oggetto STRUMENTO DA CUCINA (può comparire come un
  personaggio o come un altro elemento del racconto)
- l' azione RIUTILIZZARE.
Scegli liberamente l'ambientazione e lo stile narrativo.
```

ENGLISH

```
Write a story of two pages (approximately 5,000 characters).
Your readers are teenagers aged 15 to 18 who are passionate
about reading.
Your story must include:
The object KITCHEN TOOL (it can appear as a character or as
another element of the story).
The action REUSE.
You are free to choose the setting and narrative style.
```

Figure 3. The prompt used with ChatGPT for G-stories

Table 1. Writer questionnaire

Italian	English	Item type
Descrivi come hai lavorato per la scrittura del racconto	Describe how you worked on writing the story	Open answer
Qual è stata la cosa più difficile nella scrittura del racconto, o quella che ti ha chiesto più impegno?	What was the most difficult aspect of writing the story, or what required the most effort from you?	Open answer
Qual è la cosa di cui sei più soddisfatto/a del tuo racconto?	What are you most satisfied with in your story?	Open answer
Come valuti il tuo racconto?	How do you rate your story?	Likert scale [1-6]

Table 2. Short stories dataset (corpus)

N	Label	Description
6	H-stories	Short stories written by students without the use of ChatGPT
5	S-stories	Short stories written by students with the use of ChatGPT
3	G-stories	Short stories written completely by ChatGPT

3.2. Text assessment and analysis

The short story compilation was printed and 21 students of another sophomore class in the same school were assigned the reading and assessment of four short stories each. Assessment was formulated via an online Reader Questionnaire that captured their feedback in terms of perceived quality as readers, namely: overall quality, quality of writing, consistency, engagement, difficulty, imagination, originality, connection with reader's experience (Table 3). Also in this case, no personal data was collected.

Table 3. Reader questionnaire

Italian	English	Item type
Il racconto è coerente e non si contraddice	The story is coherent and does not contradict itself	Likert scale [1-5]
Il racconto è coinvolgente	The story is engaging	Likert scale [1-5]
Ci sono parti che non ho capito o non sono chiare	There are parts I did not understand or that are unclear	Likert scale [1-5]
Sono riuscito a immaginarmi luoghi, situazioni e personaggi del racconto	I was able to imagine the places, situations, and characters in the story	Likert scale [1-5]
Avevo già letto un racconto simile a questo	I had already read a story similar to this one	Likert scale [1-5]
Il racconto mette a tema qualcosa che è vicino alla mia esperienza	The story addresses something that relates to my experience	Likert scale [1-5]
L'autore/autrice di questo racconto sa scrivere bene	The author of this story writes well	Likert scale [1-5]
In generale, credo che sia un buon racconto	Overall, I think this is a good story	Likert scale [1-5]
Cosa ti è piaciuto di più di questo racconto?	What did you like most about this story?	Open answer

Secondo te, chi ha scritto questo racconto?	Who do you think wrote this story?	Open answer

The corpus was also analysed from a quantitative perspective. Although the number of texts was limited, we deemed methodologically appropriate to include this type of analysis in order to provide an overview of some significant features. The analysis was conducted using corrige.it[1] and READ-IT[2], and addressed: spelling accuracy and punctuation, syntactic complexity, readability index, and lexical features and quality (Table 4).

Table 4. Linguistic traits considered in quantitative analysis

Considered linguistic traits	Definition
Spelling Accuracy and Punctuation	The correct application of orthographic rules and the use of punctuation marks.
Syntactic Complexity	The structural elaboration of sentences in a text, typically measured through factors such as sentence length, clause embedding, and the use of diverse syntactic constructions.
Readability Index	A measure that evaluates the ease with which a text can be read, based primarily on linguistic factors such as sentence length and word complexity.
Lexical Features and Quality	The characteristics of vocabulary used in a text, including word frequency and diversity.

We selected these linguistic features because they are among those that can be analysed most efficiently using automatic tools. Moreover, such heterogeneous elements offer meaningful insights despite the limited size of the corpus and are relevant from a pedagogical perspective.

To complement the quantitative observations and explore more subtle aspects of textual quality, a qualitative analysis was conducted on selected excerpts from both G-stories and H-stories. The goal was to investigate features of writing that are less easily captured through automatic metrics but are crucial in assessing narrative sophistication, stylistic expressiveness, and authorial voice.

The selection of excerpts followed a purposive sampling strategy, aiming to identify representative passages that exhibited relevant linguistic, rhetorical, or narrative features. The analysis focused on surface-level linguistic patterns (e.g., unusual collocations, syntactic repetition), rhetorical structures (e.g., epiphora, emphasis), and deeper narrative components such as psychological plausibility and emotional framing.

---

[1] https://pro.corrige.it/

[2] https://www.ilc.cnr.it/dylanlab/apps/texttools/

4. Results

4.1. Authors' experiences

The writing experience of authors was analysed through the answers to the Writer Questionnaire (attachment 1). Authors of S-stories were more satisfied with their text than authors of H-stories (4.6 vs. 3.8 on a 5-point scale). Table 5 reports the elements that both groups of authors indicated as satisfying or challenging. Interestingly, authors of S-stories and H-stories emphasize almost completely different aspects, suggesting that, with the introduction of ChatGPT, the focus of the writing process had shifted: S-authors seem mostly concerned about interacting with the AI, and do not mention style consistency, character development and time management, which are frequently mentioned by H-authors.

Table 5. Satisfaction and challenge elements for authors

Authors	Satisfaction	Challenges
H-stories	Idea, plot, conclusion, writing process	Choosing words, describing characters, keeping a consistent style, time management.
S-stories	Idea, text smoothness, interaction with AI	Developing the outline, having the AI doing what I wanted, fine-tuning the text

4.2. Readers feedback

Readers' feedback was analysed through the Reader Questionnaire data. Given the limited number of participants, after confirming the absence of differences dependent on the original stimulus, only descriptive statistics was used.

When asked about writing and overall quality, young readers appreciated all three types of short stories, with a slight preference for S-stories. G-stories seem to have been less appreciated (Figure 4). Specific feedback items investigated text consistency, experienced engagement in the story, reading difficulty, imagination, originality, and connection with the reader's experience. Their scores are also rather balanced across H-, S- and G-stories, again with a slight preference for S-stories, especially when it comes to consistency and engagement (Figure 5). However, two remarkable differences emerge:

- H-stories are more difficult to read (H=2.31 vs. S=1.64, G=1,57)
- G-stories connect less with the reader's experience (G=1,61 vs. H=2.14, S=2.24).
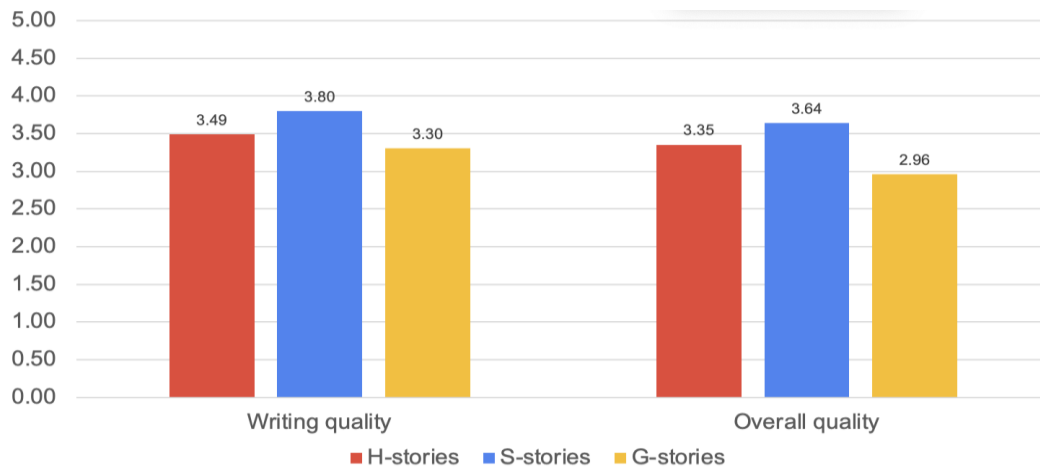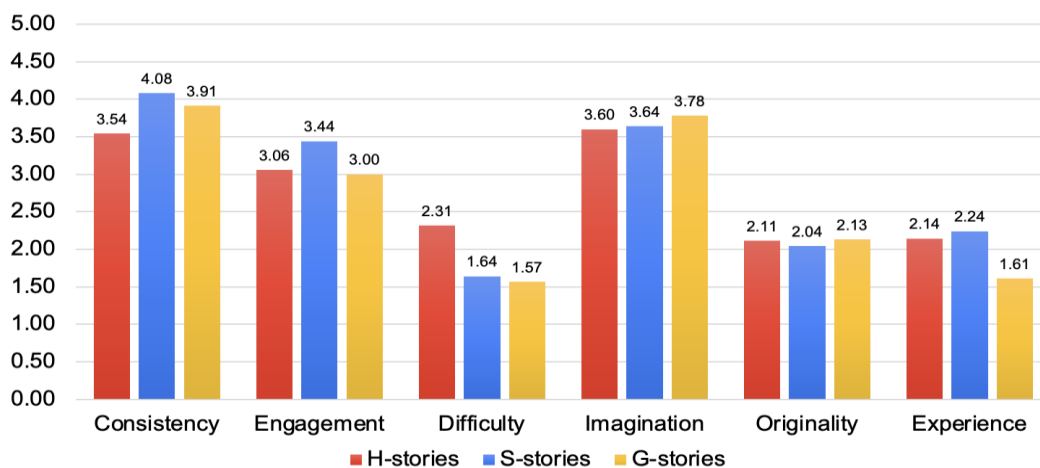
Figure 4. Quality assessment by readers

Figure 5. Detailed feedback by readers

Finally, readers were asked to guess who the author of each text was. Most readers of any type of short story identified the authors as "other boys/girls of our age" (about 60% of all answers). However, this choice is prevalent (about 75%) for G-stories. H-authors were more often identified as "writers", and S-authors were more often identified as "adults with other professions" (i.e., not writers). This might suggest that readers of H-stories perceived a stronger authorial voice.

4.3. Quantitative text analysis

4.3.1. Spelling accuracy and punctuation

As expected, the texts generated by ChatGPT contain no spelling mistakes. In contrast, S-story texts contained two errors, while H-stories, which were written completely by students, presented a total of 13 mistakes, including missing accents and apostrophes, as well as missing commas in parenthetical elements. H-stories also presented additional morphological errors, e.g., plural-singular agreement mistakes like in "breve sequenze" (should be "brevi sequenze") and "costanti e regolare esercizio fisico" (should be "costante e regolare esercizio fisico").

While no quantitative analysis was conducted on punctuation, a general trend could be observed: ChatGPT exhibits a preference for standardized and repetitive solutions (sequences

punctuated by full stops and commas), whereas human writers employ a wider range of strategies. Although their results may not always be entirely satisfactory, they are certainly more varied and effective in conveying the author's feelings and some particular stylistic effects. The following excerpt from an H-story starts with a nominal sentence, and the following paragraph includes a combination of period, comma and colon in order to emphasise the atmosphere of anxiety and tension in the text.

ITALIAN
Imprecazioni e preghiere. Queste erano le componenti distintive delle poche parole pronunciate in mezzo all'aria densa di paura e impazienza che sempre precedeva l'adrenalina innescata dal primo proiettile, dal rumore di eliche assassine: nel ventunesimo secolo i droni commerciali con cui gli appassionati di paesaggi naturali sorvolavano creste innevate in Ucraina trasportavano bombe da mortaio. [S3H]
ENGLISH (translated)
Curses and prayers. These were the distinctive components of the few words spoken amidst the thick air of fear and impatience that always preceded the adrenaline triggered by the first bullet, the sound of murderous propellers: in the twenty-first century, the commercial drones that nature enthusiasts used to fly over snow-covered ridges in Ukraine carried mortar bombs. [S3H]

Finally, in G-texts, the semicolon is not used at all; in contrast, it appears twice in S-stories and 13 times H-stories.

4.3.2. Syntactic Complexity

Syntactic complexity is a highly complex parameter. We present here some preliminary findings obtained through the automatic analysis tools mentioned above. First, average sentence length is higher in H-stories (20,37), slightly lower for S-stories (19,36) and about 25% lower for G-stories (15,37), which suggests that humans tend to use more words. This is likely to have contributed to the readers' impression that G-stories were easier to read.

In addition to this, through the READ-IT tool, we could observe the distribution of parts of speech and some specific structural preferences, including the frequency of coordination and subordination. Our findings indicate that human-written narratives tend to be more dynamic, showing a higher percentage of verbs (17,9%), used in a more varied way. This percentage decreases to 15,6% in the S-texts and further drops to 14,4% in those generated by artificial intelligence, which instead tend to favour nouns and adjectives.

Tables 6a and 6b provide an overview of the key aspects of the stories' syntactic profile: the percentage distribution of selected parts of speech (nouns, proper nouns, adjectives, verbs, coordinating and subordinating conjunctions; Table 6a), and the presence of coordinated and subordinated clauses (Table 6b). In addition to the previously mentioned features, the higher percentage of subordinate clauses is also a distinctive characteristic of H- and S-stories. This suggests a "human preference" for a more hypotactic structure in Italian-language writing.

Table 6a. Comparison of frequency of elements of the morpho-syntactic profile

Part of speech	H-stories	S-stories	G-stories
Noun	18,4%	20,9%	20,8%
Proper name	2,7%	1,5%	3,3%
Verb	17,9%	15,6%	14,4%

Adjective	7,1%	8,6%	9,8%
Conjunction	4,9%	5%	4,5%
./. coordinating	*64,1%*	*68,1%*	*77,6%*
./. subordinating	*35,9%*	*31,9%*	*22,4%*
Main clauses	52,9%	57,5%	71,4%
Subordinate clauses	47,1%	42,4%	28,6%

Table 6b. Presence of main and subordinate clauses

Kind of clause	H-stories	S-stories	G-stories
Main clauses	52,9%	57,5%	71,4%
Subordinate clauses	47,1%	42,4%	28,6%

4.3.3. Readability Index

The text analysis tool also allowed computing the readability index of the texts, namely GULPEASE, a standardized measure in linguistic studies specifically calibrated for Italian (Lucisano & Piemontese, 1988). GULPEASE provides indications about the relative "ease" or "difficulty" of a text in relation to the reader's educational level, considering word and sentence length. While these are not the only factors contributing to textual complexity, they offer an initial indication of readability, to combine with further lexical analysis. Corrige.it generates a text score between 1 (difficult) and 100 (easy).

The average scores of our subcorpora are as follows: H-stories 55, S-stories 54, G-stories 57. Overall, we observe no significant differences among the texts, which are generally very difficult for readers with primary or lower secondary education, but easily comprehensible for those with higher levels of education. However, some texts written by humans reach higher peaks of complexity (35). This is not inherently problematic, because in narrative texts complexity can be an interesting and appropriate feature, to be considered within the broader context of the author's stylistic choices.

4.3.4. Lexical Features and Quality

In general, H-stories show a slightly higher lexical richness; this measure refers to the range and variety of vocabulary deployed in a text. However, it is not possible to identify significant differences in this regard, as the measure obtained using Corrige.it ranges from 2,94 (H-stories) to 2,72 (S-stories) and 2,18 (G-stories).

Table 7. List of occurrences in the corpus (G stories were in general shorter)

H-stories	S-stories	G-stories
essere (163)	essere (107)	essere (30)
avere (77)	avere (45)	fonderia (14)

cavallo (34)	fabbrica (21)	Alvaro (13)
fare (28)	trovare (16)	antico (12)
Geremia (26)	Alex (15)	potere (11)
potere (24)	giorno (14)	fare (10)
Margaret (22)	tempo (14)	chiave (9)
mamma (21)	nuovo (13)	segreto (8)
ora (20)	fare (13)	biblioteca (8)
Zeno (18)	casa (13)	vecchio (7)

What is particularly interesting from a didactic perspective, however, is identifying the lexical preferences emerging in the different sub-corpora to examine the vocabulary of younger generations in written compositions. If, overall, more than 90% of the words belong to the Basic Vocabulary of Italian[3], lemmatization allowed us to obtain the lists of occurrences in Table 7.

The list highlights notable differences in the frequency of certain lemmas. For instance, the auxiliary verbs *essere* ("to be") and *avere* ("to have") appear with high frequency in H-stories but are underrepresented in G-stories. Similarly, the verb *fare* ("to do" / "to make") is used more frequently in H-stories and S-stories than in G-stories. One possible explanation is its occurrence in the first two groups within idiomatic expressions (*aveva fatto la sua parte* ["he/she had done their part"]; *poteva fare la differenza* ["he/she could make a difference"]) or in collocations (*vieni a fare colazione* ["come have breakfast"]), a pattern less evident in G-stories. It is also possible to observe the use of the verb *fare* as a generic verb (applicable in various phrasal contexts), a phenomenon that has been identified analysing a large corpus of handwritten texts produced at school in Canton Ticino (Cignetti et al. 2016), but further investigation is needed, along with a more precise balancing of the three sub-corpora. However, to formulate hypotheses about these frequency differences with a methodologically sound approach, the analysis should be extended to a larger number of texts and to different text types for a more comprehensive investigation.

4.4. Qualitative text analysis

4.4.1. Lexical choices

The qualitative analysis of the texts has revealed several noteworthy aspects. A primary linguistic feature concerns the use, in H-stories, of lexically possible choices that are uncommon in contemporary Italian usage. Examples of this phenomenon include the noun-adjective pair *mente vigilante* ("vigilant mind"), the phrase *destinata alla dimenticanza* ("destined for

---

3

https://www.dropbox.com/scl/fi/zg2y99xqik4k11nj19fgi/nuovovovocabolariodibase.pdf?rlkey=s0uf8ggv11kf44ip6a2ldz16n&dl=0

oblivion"; although attested in some literary texts of the past, is perceived in ChatGPT's usage as a marked shift in register), and specific noun pairs such as un *futuro fatto di reinvenzione e creatività* ("a future built on reinvention and creativity") and *un'oasi di storia e curiosità* ("an oasis of history and curiosity").

ITALIAN
[1] Ogni giorno, con occhi penetranti e MENTE VIGILANTE, sorvegliava la biblioteca e il potere che essa racchiudeva. [S2G1]
[2] La vecchia frusta, una volta DESTINATA ALLA DIMENTICANZA, aveva trovato una nuova vita e una nuova bellezza. [S1G1]
[3] E mentre la notte scendeva sulla tranquilla cittadina di San Giorgio, il loro portacandele illuminava il cammino verso UN FUTURO FATTO DI REINVENZIONE E CREATIVITÀ. [S1G1]
[4] Nella pittoresca cittadina di San Giorgio, la bottega di antiquariato "La Stamberga delle Meraviglie" si stagliava contro il panorama urbano, UN'OASI DI STORIA E CURIOSITÀ. [S1G1]

ENGLISH (translated)
[1] Every day, with piercing eyes and a VIGILANT MIND, he watched over the library and the power it contained. [S2G1]
[2] The old whip, once DESTINED FOR OBLIVION, had found a new life and a renewed beauty. [S1G1]
[3] And as night fell over the quiet town of San Giorgio, their candle holder illuminated the path toward A FUTURE BUILT ON REINVENTION AND CREATIVITY. [S1G1]
[4] In the picturesque town of San Giorgio, the antique shop "The Wonders' Attic" stood out against the urban landscape, AN OASIS OF HISTORY AND CURIOSITY. [S1G1]

4.4.2.   Lexical repetitions

Anaphoric chains in texts generated by ChatGPT frequently rely on lexical repetition, even in contexts where the use of pronouns or synonyms would render the text more natural and effective. This is evident in the repetition of the noun *fonderia* ("foundry") in the following example from a G-story:

ITALIAN
[5] Un giorno, durante una passeggiata solitaria, Alvaro notò un'inconsueta attività intorno alla FONDERIA. Uomini e donne dall'aspetto losco si aggiravano furtivamente, portando con sé strani attrezzi e materiali misteriosi. Deciso a scoprire cosa stesse accadendo, Alvaro si avventurò oltre il cancello arrugginito della FONDERIA, conscio del pericolo che lo attendeva. Una volta dentro, Alvaro rimase sbalordito. La FONDERIA non era più l'antica officina abbandonata che credeva. [S3G1]

ENGLISH (translated)
[5] One day, during a solitary walk, Alvaro noticed an unusual activity around the FOUNDRY. Men and women with a suspicious appearance lurked furtively, carrying strange tools and mysterious materials. Determined to uncover what was happening, Alvaro ventured beyond the rusty gate of the FOUNDRY, fully aware of the danger that awaited him. Once inside, Alvaro was astonished. The FOUNDRY was no longer the abandoned old workshop he had believed it to be. [S3G1]

Lexical repetitions also occur in texts written by humans; however, in such cases, their rhetorical-expressive function is evident. The repetition of *stanco* ("tired") in the following H-story excerpt demonstrates a deliberate and expressively effective use of the rhetorical figure known as *epiphora*, emphasizing the phrasing:

ITALIAN
[6] Geremia era molto STANCO. Ogni lunedì arrivava a casa STANCO. Ogni giorno che andava a scuola tornava STANCO. Ma il lunedì era peggio. A Geremia piaceva la scuola ma il lunedì era faticoso. Soprattutto se sei un bambino di 10 anni e vedi tuo padre solo il fine settimana. [S1H2]

ENGLISH (translated)
[6] Geremia was very TIRED. Every Monday, he came home TIRED. Every day he went to school, he returned TIRED. But Mondays were worse. Geremia liked school, but Mondays were exhausting. Especially if you are a ten-year-old child and see your father only on weekends. [S1H2]

4.4.3.   Character development, narrative development and causality

Another aspect that emerged from the qualitative analysis – undoubtedly deserving of further investigation in a future study – is the absence of affective and emotional factors related to the protagonists in G-stories. By contrast, such factors appear especially in H-stories, where they enrich the psychological portrayal of the character and can even serve as key elements in the narrative development of the story, as demonstrated by examples 7 and 8. Another emerging pattern in the qualitative analysis concerns the use of causality. G-stories often display a linear but shallow causal structure, where events follow each other with limited elaboration on motivations or consequences. This results in narratives that are formally coherent but semantically weak in terms of depth and psychological realism. By contrast, H-stories frequently construct causal relationships between actions and emotional or contextual elements. Consider the example in 7, where the character's decision to leave home is clearly motivated by emotional turmoil, adding plausibility and engagement to the storyline. This richer causal connectivity appears linked to the human author's greater awareness of psychological plausibility and narrative logic.

ITALIAN
[7] Preso DALL'IRA E DAL DOLORE, il ragazzino corse in camera sua e preparò lo zaino pronto a partire verso quella maledetta fabbrica. Ora più che mai era convinto di voler andare e DIMENTICARE DEL SUO CARO CANE, non poteva convivere anche con QUEL DOLORE… ne aveva passate già abbastanza. [S3H1]
[8] Così uscì di casa silenziosamente ARRABBIATO E ADDOLORATO, SENZA NEANCHE LANCIARE UN ULTIMO SGUARDO A SUA MAMMA, che non si era ancora mossa. NON PROVAVA NEMMENO A PREOCCUPARSI DI LUI. [S3H1]

ENGLISH (translated)
[7] Overcome with RAGE AND PAIN, the boy ran to his room and packed his backpack, ready to head toward that cursed factory. Now more than ever, he was determined to go and FORGET ABOUT HIS BELOVED DOG; he could not bear THAT PAIN as well… he had already endured too much. [S3H1]

[8] Thus, he left the house silently, ANGRY AND HEARTBROKEN, WITHOUT EVEN CASTING ONE LAST GLANCE AT HIS MOTHER, who had not yet moved. SHE DIDN'T EVEN SEEM TO WORRY ABOUT HIM. [S3H1]

5. Discussion

The small-scale experimentation presented in this study was localized in a specific high school context, involving a small sample, a specific language (namely, Italian), a specific genre (fiction short story), and a particular chatbot (ChatGPT 3.5). Moreover, the research design only considers a specific way of using GenAI-tool in writing (namely, in generating the first draft of the text). For these reasons, its results cannot be generalized, especially to learning scenarios that include other ways of using AI in writing (e.g., for generating ideas, or improving specific passages, etc.) or different tools. Nonetheless, the results help shedding some light both on the textual quality that can be expected from GenAI-tools and on the impact of such tools in the writing process and in writing education.

The data collected from the Writer Questionnaire clearly indicated that for the authors of S-stories, the focus of writing was the interaction with ChatGPT, and not writing itself. It can be argued that such a shift in focus is more conducive to the development of AI Literacy skills (how do I interact effectively with a chatbot?) than of actual writing skills (e.g., developing an own writing style). However, the balance between the two could be fine-tuned by providing appropriate tasks and scaffolding learners at different stages.

Both Readers Questionnaire results and the text analyses indicated that stories generated by ChatGPT were error-free and easier to read, e.g., they had shorter sentences, less lexical variety, used less verbs and more nouns, and included more repetitions. While these textual features can be adapted with proper prompting, they suggest that AI-generated texts can be used as models of writing only to a limited extent, and only in relation to specific features. ChatGPT writing style tends to follow standard plain writing rules.

However, the qualitative text analysis revealed deeper insights into the nature of human vs. AI-generated writing. First, lexical choices in H-stories revealed a greater stylistic variety and a deliberate use of marked or literary expressions, contributing to a richer textual fabric. These expressions were virtually absent in G-stories, where vocabulary tended to be more standardized. This signals a lower lexical experimentation in AI-generated texts, potentially limiting students' exposure to more expressive or nuanced language.

Secondly, lexical repetition was employed differently across the two text types. While ChatGPT-generated stories frequently repeated nouns mechanically, human writers used repetition rhetorically and expressively, for example through epiphora (as in the repeated use of *stanco*). This finding emphasizes how repetition, when human-driven, contributes to rhythm and emotional emphasis, whereas in G-stories, it often results from limitations in co-reference management.

A third important distinction regards character development, where H-stories incorporated affective and emotional states that shaped the protagonists' decisions and narrative progression. Emotions such as anger, sorrow, and frustration are not only described but causally linked to plot events, as seen in examples. G-stories, in contrast, tended to lack emotional depth and internal perspective.

Such observations suggest that writing without AI provides more opportunities to develop authorial skills, exactly because the overall focus remains on writing as such and is not shifted to interacting with a machine or partner. We also interpret these results as a call to continue the conversation about the artistic value of AI-generated artefacts (Kraaijeveld, 2024).

6. Conclusions

The key results of this small-scale exploratory study were also presented and discussed with the participants and their teacher. The framework of a "challenge with AI" proved to be an effective motivational element for engaging in writing, and the whole process provided several opportunities to discuss not only writing, but also how LLM work and our individual and societal relationship with AI, in the framework of AI Literacy education (Ranieri et al., 2023). These are also venues that deserve further exploration.

This study explored student's use of AI in writing, ignoring the obviously detrimental situation in which students fully hand over writing to the AI. Our results support the idea that AI can be beneficial for learning to write, as AI-generated texts indeed possess relevant features, e.g., they are easier to read and use shorter sentences. However, our participants' experience suggests that AI can be beneficial to writing education only if framed within an appropriate teaching and learning approach. New AI-supported instructional strategies should be developed, to prevent the focus shifting from writing to use an AI tool that students reported. Such strategies might identify more contextualized and task-oriented modes of integration of Gen-AI tools into specific phases in the writing process. Following are some examples:

- Conversating with a chatbot to generate useful ideas (for fiction) or arguments (for argumentative texts) before developing the outline of a text.
- Enhancing lexical choices by expanding the generic vocabulary frequently used by students encouraging them to find more precise alternatives using AI support, critically reflecting on contextual appropriateness and on pragmatic effects.
- Interacting with a chatbot for *revising* drafts (as opposed to *generating* drafts), using a prompt that asks the system not to simply "improve the text", rather to "highlight passages that can be improved and suggest possible improvements", until a satisfactory result is achieved. Then, the various versions produced (by different people) can be compared and discussed.

These examples limit the use of AI to specific writing phases (e.g., *inventio*, or *elocutio*), so that students get feedback that they should elaborate. In this way, the output of the GenAI tool is never the final target text, leaving the main agency to the student. Implementing such interactions requires a high level of control on the writing process, which might be difficult to achieve for younger students. Such interactions might also be embedded in a game-like situation like "continue each other's story" (Yang et al., 2022).

More generally, our results suggest that integrating AI in teaching and learning processes requires a careful analysis of learning processes – in our case, the writing process. Using AI to enhance education is not just a matter of generating higher-quality products (e.g., texts) or making the process more efficient (e.g., spend less time writing longer texts). Learning requires time and effort, and AI can be easily used as a shortcutting that does not necessarily imply better learning (Bohacek, 2023). To generate meaningful integration of AI in teaching and learning situations, both teachers' and students' basic AI Literacy is required in order to avoid a "magic approach", in which Gen-AI tools are considered like omniscient oracles or humanized.

From a research point of view, it would certainly be valuable to replicate the study in at least three directions: (a) expanding the sample size and varying its composition, e.g., controlling for writing skills or habits; (b) including different text types, for enriching and strengthening the analyses with also with e.g., descriptive, normative or argumentative texts; and (c) using different Gen-AI tools. We also expect that the comparison of similar studies across different languages would generate insights in how LLMs generate language and interact with human writers.

7. References

Abdullayeva, M., & Musayeva, Z. M. (2023, April). The Impact of Chat Gpt on Student's Writing Skills: An Exploration of Ai-Assisted Writing Tools. In *International Conference of Education, Research and Innovation* (Vol. 1, No. 4, pp. 61-66). https://academicsresearch.ru/index.php/ICERI/article/view/1613

Bohacek, P. (2023). Engaging Students to Avoid Unproductive Shortcuts. *EdCircuit.* https://edcircuit.com/engaging-students-to-avoid-unproductive-shortcuts/

Cignetti, L. (2018). Sulla didattica del testo narrativo: dalla lettura alla rielaborazione. *Serie Treccani lingua italiana*. http://www.treccani.it/magazine/lingua_italiana/articoli/scritto_e_parlato/Cignetti2.html.

Cignetti, L., Demartini, S., & Fornara, S. (2016). Come TIscrivo? La scrittura a scuola tra teoria e didattica. Aracne.

Cignetti, L., Demartini, S., Fornara, S., & Viale, M. (2022). *Didattica dell'italiano come lingua prima*. Il Mulino.

Daulay, S. H., Berutu, H., Dalimunte, M., Nasution, M. M., Apriani, E., Muthmainnah, M., & Shaumiwaty, S. (2024). What AI-Based Writing Assistant Actually Improved: Writing Quality or Writing Skills?. In *Impacts of Generative AI on Creativity in Higher Education* (pp. 423-442). IGI Global. https://www.igi-global.com/chapter/what-ai-based-writing-assistant-actually-improved/355440

Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: a qualitative study. *Computer Assisted Language Learning*, 1-40. https://www.tandfonline.com/doi/full/10.1080/09588221.2024.2374772

Flower. L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*(4), Dec., 365-387. https://doi.org/10.2307/356600

Ghajargar, M., Bardzell, J., & Lagerkvist, L. (2022, November). A redhead walks into a bar: experiences of writing fiction with artificial intelligence. In *Proceedings of the 25th international academic MindTrek conference* (pp. 230-241). https://dl.acm.org/doi/abs/10.1145/3569219.3569418

Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., Çakir, D., & Gerjets, P. (2022). The pure poet: How good is the subjective credibility and stylistic quality of literary short texts written with an artificial intelligence tool as compared to texts written by human authors?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Huang, H., Tang, T., Zhang, D., Zhao, W. X., Song, T., Xia, Y., & Wei, F. (2023). Not all languages are created equal in llms: Improving multilingual capability by cross-lingual-thought prompting. *arXiv preprint arXiv:2305.07004*. https://doi.org/10.48550/arXiv.2408.02237

Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. *Journal of Intelligent Communication*, *4*(1), 20-31.

Ibrahim, H., Asim, R., Zaffar, F., Rahwan, T., & Zaki, Y. (2023). Rethinking homework in the age of artificial intelligence. *IEEE Intelligent Systems*, *38*(2), 24-27. https://doi.org/10.1109/MIS.2023.3255599

Ironsi, C. S., & Solomon Ironsi, S. (2024). Experimental evidence for the efficacy of generative AI in improving students' writing skills. *Quality Assurance in Education*. https://www.emerald.com/insight/content/doi/10.1108/qae-04-2024-0065/full/html

Kindenberg, B. (2024). ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis. *Education Sciences, 14*(5), 530 https://doi.org/10.3390/educsci14050530

Kraaijeveld, S. R. (2024). AI-generated art and fiction: signifying everything, meaning nothing?. *AI & Society*, 1-3. https://link.springer.com/article/10.1007/s00146-023-01829-4

Lavinio, C. (1995). Scrivere testi narrativi. In M.T. Calzetti, & L. Panzeri Donaggio (eds), *Educare alla scrittura* (137-156). La Nuova Italia.

Lee, V. R., Pope, D., Miles, S., & Zárate, R. C. (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. *Computers and Education: Artificial Intelligence*, *7*, 100253.

Lopes, C., Forgas, R. C., & Cerdà-Navarr, A. (2024). A PhD dissertation in education written by artificial intelligence?. *Revista Brasileira de Educação*, *29*, e290065.

Lucisano, P., & Piemontese, M. E. (1988). GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana. *Scuola e Città, 3*, 31, marzo 1988.

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Jo Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D. & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?. *Computers and Education: Artificial Intelligence*, *3*, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Ranieri, M., Cuomo, S., & Biagini, G. (2023). Scuola e intelligenza artificiale. Percorsi di alfabetizzazione critica. Carocci.

Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), 100083.

Sassoon, J. (2019). Storytelling e intelligenza artificiale: quando le storie le raccontano i robot. FrancoAngeli.

Smith, F. (2013). *Writing and the writer*. Routledge.

Smith, F. (1983). Reading like a writer. *Language Arts*, *60*(5), 558-567.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1260843/full

Williams, C., & Beam, S. (2019). Technology and writing: Review of research. *Computers & Education*, *128*, 227-242.

Yang, D., Zhou, Y., Zhang, Z., Li, T. J. J., & LC, R. (2022, March). AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the ACM*

IUI Workshops (Vol. 10, pp. 1-11). CEUR-WS Team. https://hai-gen.github.io/2022/papers/paper-HAIGEN-YangDaijin.pdf

8. Attachments

8.1. Attachment 2. Short story assessment

Original Italian version
- Il racconto è coerente e non si contraddice [Likert scale 1-5]
- Il racconto è coinvolgente [Likert scale 1-5]
- Ci sono parti che non ho capito o non sono chiare [Likert scale 1-5]
- Sono riuscito a immaginarmi luoghi, situazioni e personaggi del racconto [Likert scale 1-5]
- Avevo già letto un racconto simile a questo [Likert scale 1-5]
- Il racconto mette a tema qualcosa che è vicino alla mia esperienza [Likert scale 1-5]
- L'autore/autrice di questo racconto sa scrivere bene [Likert scale 1-5]
- In generale, credo che sia un buon racconto [Likert scale 1-5]
- Cosa ti è piaciuto di più di questo racconto? [open answer]
- Secondo te, chi ha scritto questo racconto? Ad esempio, un ragazzo, una ragazza, uno scrittore di professione, una giornalista, uno scienziato, ecc. [open answer]

Translated English version
- The story is coherent and does not contradict itself. [Likert scale 1-5]
- The story is engaging. [Likert scale 1-5]
- There are parts I did not understand or that are unclear. [Likert scale 1-5]
- I was able to imagine the places, situations, and characters in the story. [Likert scale 1-5]
- I had already read a story similar to this one. [Likert scale 1-5]
- The story addresses something that relates to my experience. [Likert scale 1-5]
- The author of this story writes well. [Likert scale 1-5]
- Overall, I think this is a good story. [Likert scale 1-5]
- What did you like most about this story? [open answer]
- Who do you think wrote this story? For example, a boy, a girl, a professional writer, a journalist, a scientist, etc. [open answer]