

Effects of ChatGPT-Instruction on Text Comprehension and User Experience

Anke Schmitz, Rahel Baer, Christina Holzwarth, & Michael Ruloff

Theoretical Background

- The rapid advancements in artificial intelligence and LLMs are reshaping educational paradigms worldwide with certain risks and opportunities (Kasneji et al., 2023).
- Students increasingly use LLMs (e.g. ChatGPT) to gather information, summarize texts, solve problems, and complete assignments. At the same time, they may be tempted to shorten learning paths (skill skipping). This forces teachers to reflect on and innovate traditional instruction and assignment formats.
- Focusing on upper secondary schools (grades 11-13), students are required to manage multiple expository texts which pose challenges regarding global coherence and applying knowledge to new contexts. Unfortunately, current research indicates that digital reading, especially longer and complex texts, tends to be more superficial compared to paper-based texts (Clinton, 2019; Delgado et al., 2018).
- Regarding the comprehension of texts with the help of LLMs, students with strong metacognitive skills apply LLMs more strategically and critically (Chen et al., 2024). Also, the use of LLMs positively affect user experience in terms of perceived control and sense of time (Zheng & Fan, 2024). But especially weaker students who struggle with text comprehension and are less experienced with digital tools are disadvantaged and feel less confident (Erzinger, 2023).
- While the majority of studies have been conducted with university students (Chen & Shimada, 2023; Zheng & Fan, 2024), the present study analyzes whether a targeted and reading process-oriented instruction affects secondary school students' usage of LLMs and impacts digital text comprehension and user experience.

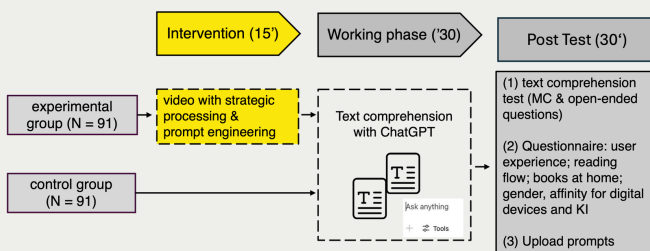
Research Questions

1. How does an instruction on the strategic, reading process-oriented use of ChatGPT affect students' prompting and user experience?
2. How does the instruction influence students' digital text comprehension?
3. Is the effect of the instruction on text comprehension moderated by students' prompting and/or user experience?

Instruments

A 7-minute **video** that explains and models the strategic, reading process-oriented use of ChatGPT for the deep comprehension of multiple complex texts for preventing skill skipping. It also addresses the critical use of LLMs in terms of hallucinations. Established strategies from models such as SQ3R are adapted to the digital context using LLMs including prompt engineering.

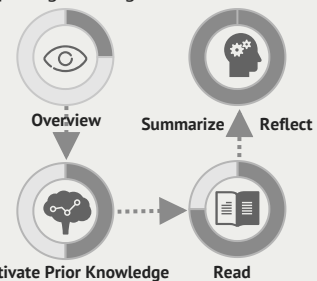
Method & Sample



Sample: N = 182 students in 3 upper secondary schools in Northwestern Switzerland; m = 33.5%, f = 63.7%, d = 2.7%; M = 17.75 (SD = 1.18) years old



+ prompt library



Two expository texts about the topic *satire* (5 pages)

Text comprehension test: 12 Items, MC and open-ended ($\alpha = .65$)

User Experience: 10 Items ($\alpha = .90$), $M = 3.76$, $SD = 1.12$; cf. Strzelecki (2023)

Reader variables: reading motivation (4 Items, $\alpha = .86$), books at home, gender, affinity for digital devices and AI outside school (7 Items, $\alpha = .66$) (PISA, 2022)

Comparison of Students' Prompts

Prompts differed significantly: Students in the experimental group (EG) generated more prompts than those in the control group (CG), $t(1, 110) = 8.65$, $p < .001$, $d = 1.83$. They also formulated more strategic prompts aligned with the reading process (e.g., requesting an overview, identifying main messages, author perspectives, connections and contradictions, or asking the chatbot follow-up questions). In contrast, students in the CG produced prompts of a different nature, focusing on text simplification, explanation of facts, and simple summaries, $t(1, 106) = 4.13$, $p < .001$, $d = 1.02$.

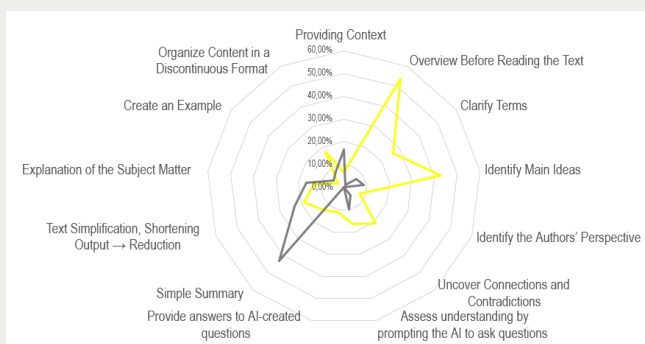


Fig. 1: Distribution of prompts (CG = grey, EG = yellow).

First Results

- MANOVA with prompting, text comprehension and user experience as dependent variables revealed a large effect of the group intervention on prompting ($F(1, 102) = 79.09$, $p < .001$, $\eta^2 = .44$) and on user experience ($F(1, 102) = 12.23$, $p < .001$, $\eta^2 = .11$), but not on text comprehension ($F(1, 102) = 6.76$, $p = .343$).
- A multiple moderated regression including intervention group, mean number of prompts, and interaction of group \times mean prompts and user experience as covariate revealed a significant interaction effect on text comprehension.

	b	SE B	t	p
Constant	10,7308	,8701	12,3324	,0000
Group	,8897	,8865	1,0036	,3181
Prompts (centered)	,3636	,1781	2,0418	,0439
Group x Prompts (centered)	1,0599	,3906	2,7133	,0049
User Experience (centered)	-,7554	,3504	-2,1559	,0336

Note. $R^2 = 0.34$

Tab. 1: Linear model of predictors of text comprehension.

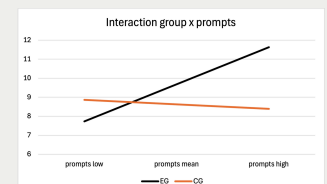


Fig. 2: Simple slopes equations of the regression of intervention on text comprehension at three levels of prompting.

- The interaction effect remains stable when gender and reading motivation were controlled.

Discussion and next steps

- Qualitative analyses of students' prompts regarding their relation to specific text segments (word level, local text level, global text level); coding of individual prompting profiles (sequences of prompts)
- Long-term effects: Replication study with follow-up test and different texts is necessary
- Modelling of adequate prompting seems to be essential to help all students to internalize the method effectively. Urgent need to implement use of LLMs into teacher professionalization.

References Contact



Prof. Dr. Anke Schmitz
University of Applied Sciences
Northwestern Switzerland
School of Education
German Didactics and its' Disciplines
anke.schmitz@fhnw.ch