

ChatGPT in education: Teachers' and Students' views

Ulrik Söderström

Digital Media Lab, Umeå University
Umeå, Sweden
ulrik.soderstrom@umu.se

Karl Lambertsson

Digital Media Lab, Umeå University
Umeå, Sweden
kala0198@student.umu.se

Elsa Hedström

Digital Media Lab, Umeå University
Umeå, Sweden
elhe0177@student.umu.se

Thomas Mejtoft

Digital Media Lab, Umeå University
Umeå, Sweden
thomas.mejtoft@umu.se

ABSTRACT

Since its launch just over two years ago, the conversational chatbot ChatGPT developed by OpenAI has become integrated into the studies of students within the Swedish educational system. This paper investigates both teachers' perspectives and attitudes toward students using ChatGPT and the students' view of their perceived learning. Insights were gathered through questionnaires (for both students and teachers) and a learning session for students, revealing concerns and enthusiasm regarding ChatGPT's integration. Teachers express a lack of understanding on incorporating ChatGPT into education and perceive a lack of support from school leadership and the Swedish National Agency for Education. Identified scenarios suggest ChatGPT's potential for understanding concepts, but resistance toward its use in examinations. The students have a statistically significant enhancement in confidence and understanding of magnetism concepts after engaging with ChatGPT. This research contributes to ongoing discussions about integrating AI tools in education, emphasizing benefits and addressing ethical concerns and learning outcomes.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design; Interaction design theory, concepts and paradigms.**

KEYWORDS

Education, Sweden, Upper Secondary School, Higher education, Generative Artificial Intelligence, ChatGPT, Perceived Learning

ACM Reference Format:

Ulrik Söderström, Elsa Hedström, Karl Lambertsson, and Thomas Mejtoft. 2024. ChatGPT in education: Teachers' and Students' views. In *European Conference on Cognitive Ergonomics (ECCE 2024)*, October 08–11, 2024, Paris, France. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3673805.3673828>



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

ECCE 2024, October 08–11, 2024, Paris, France
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1824-3/24/10
<https://doi.org/10.1145/3673805.3673828>

1 INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative force in various domains, and its integration into education has become a subject of noteworthy debate [24]. As technologies advance, the impact of AI on learning experiences becomes increasingly pronounced. In recent times many AI resources have been developed. As more open AI platforms emerge, they are capable of delivering text in reply to users' questions, claims, and information requests [25]. Such tools offer students a wide range of self-learning possibilities for preparing homework, text translation, coding, or learning a new language [24], [7]. AI and its resources are useful tools for education and learning, but they are not an alternative to the teacher's role or their ability to support students in a specific learning process [7].

The development of artificial intelligence (AI) and generative AI has accelerated in recent years, and a variety of disciplinary practices, such as healthcare [38] and education, have taken advantage of this technology. ChatGPT, a powerful language model from OpenAI [29], reached 1 million users in just five days. ChatGPT uses natural language processing to generate human-like responses to user text input. The strength of ChatGPT lies in its ability to generate coherent, systematic, and informative responses [39]. ChatGPT also has limitations, including the potential to generate answers with incorrect information [29].

With the use of ChatGPT, unlike traditional teaching methods, students may receive continuous interactive learning support at all times outside of regular classroom hours [24]. The interactive element that ChatGPT provides could enhance the learning experience by catering to individual student queries, offering a more personalized and adaptive approach to education. There has been much discussion on the potential of ChatGPT to transform education and work procedures in schools [5, 14, 34]. For instance, ChatGPT can act as a virtual tutor, answer students' questions and provide personalised learning experiences [27]. On the other hand, ChatGPT does not only influence education positively, it also introduces new challenges and threats for teachers and students. It can be used to complete written assignments and examinations on behalf of students, leading to concerns about AI-assisted cheating and plagiarism [11]. One recommendation to reduce malicious use of ChatGPT is to make students aware of academic integrity policies and help them understand the consequences of academic misconduct; use chatbots ethically and hold the students personally accountable.

Even though a lot of research about using AI in different fields has been conducted, firm conclusions about how AI, and more precisely, how ChatGPT affects students' learning in education cannot be drawn.

1.1 Objective

This paper aims to investigate the impact that ChatGPT has on perceived learning in higher education, as well as to identify Swedish teachers' attitudes towards how students can use ChatGPT in upper secondary education, including situations when utilization is inappropriate. By pinpointing the impact on perceived learning and understanding teachers' attitudes and preparedness, the study aims to provide insights into the effect ChatGPT has in both higher and secondary education settings.

2 THEORY

In this section the theoretical basis will be presented. The focus will be on AI, generative AI, and Learning, with each contributing distinct perspectives to the understanding of perceived learning through the use of generative AI.

2.1 Artificial Intelligence (AI) and Generative AI

AI as a concept can be explained more broadly as a system that imitates cognitive functions, such as learning, speech and problem-solving, usually similar to human-like characteristics [32]. In greater detail, AI empowers systems to solve specific problems or achieve a specific goal through flexible adaption, achieved through comprehension received from processing external data [20].

Generative modelling is an AI technique that generates synthetic artefacts by processing data sets, learning their patterns and distribution, and based on that creating realistic facsimiles [3, 16]. In other words, generative models are methods or techniques that can transform data into a code or subset of codes from which new data can be reconstructed [18]. Consequently, generative AI uses generative modelling and advances in deep learning. It is a subset of AI that generates diverse content by utilising existing content, such as text, graphics, audio, and video [3, 16].

2.2 Generative Pre-trained Transformer and ChatGPT

There exist several forms of Generative AI, one common technique is Generative Pre-trained Transformer (GPT) which is a type of language model [6]. It has its roots in natural language processing (NLP), which is an area of AI focused on enabling machines to understand and generate human language [12, 13, 17, 26]. This is done by training the model on a large data set of text, it is therefore able to generate new text that is similar to the text it was trained on. This technique is often used in chatbots, which are computer programs that are designed to hold text- or voice-based conversations with humans [6]. ChatGPT is based on this technique and is trained on a large corpus of text data, including books, articles and websites [1]. As mentioned in the introduction, ChatGPT is developed by the company OpenAI which has made conversions with language models more enjoyable, human-like and realistic. Examples of activities ChatGPT can be used for are asking questions, asking for advice, requesting explanations or just chatting [35].

Even though ChatGPT can be seen as a powerful AI tool there exist limitations and challenges that need to be addressed. Here are some existing problems regarding ChatGPT-3.5:

- Sensitivity to input variations - ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times [29].
- Inappropriate Responses - The model does not always refuse inappropriate requests. Sometimes it responds to harmful instructions or exhibits biased behaviour [29].
- Incorrect information - Answers that initially sound plausible can be factually inaccurate or nonsensical. There are multiple reasons for this, for example, the dataset that the model is trained on may be incorrect [29]. The information can also be outdated since ChatGPT-3.5 is trained on data through September 2021 [2].
- Biases - The model can reproduce or amplify biases present in the training data in the answers that it generates [6].

2.3 Learning

Learning is a human made hypothetical construct: it cannot be directly observed, but only be determined by observable behavior [15]. Psychologists define learning as "a relatively permanent change in behavior due to past experience" or as "the process by which relatively permanent changes occur in behavioral potential as a result of experience" [15].

Another way to define learning is done by Bacon [4] and he explains it as follows:

"The term learning here to mean the gains in the knowledge or skills that a student possesses. Learning reflects a change over time, not a state at a particular moment in time. The term perceived learning refers to a student's self-report of knowledge gain, generally based on some reflection and introspection. The term actual learning distinguishes real learning from perceived learning. Actual learning reflects a change in knowledge identified by a rigorous measurement of learning".

2.4 Vygotsky's zone of proximal development

The zone of proximal development (ZPD) represents the gap between what a learner has already mastered (their current level of development) and what they can achieve with educational support (their potential development). This zone is evident in a child's development during collaborative activities with an adult, but not when the child is working alone. Lev Vygotsky introduced the concept to address two issues in developmental and educational psychology: assessing children's intellectual abilities accurately and evaluating the effectiveness of instructional methods ([36]). He further defines the ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers".

2.5 Utilisation of ChatGPT in teaching and learning

In the realm of education, Generative AI such as ChatGPT can be a useful tool. One example of how it can be used is that it can provide personalised tutoring and feedback to students based on their needs and level of understanding. According to Baidoo-Anu and Owusu Ansah [5], a study carried out by Wang et al. in 2020, had positive results on providing personalised math tutoring to students using a chatbot based on a generative AI model. Results from a study carried out by [37] in 2020 show that a chatbot based on a generative model could provide personalised math tutoring to students. The study claimed improved learning outcomes through explanations tailored to students' misconceptions. Other examples of situations and scenarios where ChatGPT can be effectively employed and used by students are mentioned in a study by Lo [23] in 2023. Examples mentioned in this study were answering questions, summarising information, facilitating collaboration, concept checking, exam preparation, drafting assistance and providing feedback. Furthermore, a study by Rahman and Watanobe [30] in 2023 mentions additional situations where ChatGPT can be useful for students: Understanding and solving complex problems (e.g. by getting explanations and step-by-step solutions to a given problem), developing reading and writing skills by providing suggestions (e.g. syntactic and grammatical), creating practice exercises and quizzes and lastly developing analytical and out-of-the-box thinking.

ChatGPT performs variously in different subject domains. Lo [23] gives a general description of the literature that proves this. The study mentions that ChatGPT-3.5 demonstrated outstanding results in critical and higher-order thinking and economics. However, its performance was not entirely satisfactory in other subject domains, such as law, medical education, mathematics and psychology.

2.6 The Swedish National Agency for Education's advice on ChatGPT and similar tools

The Swedish National Agency for Education (Skolverket) 2024 [33] provides advice to teachers and principals on how to use ChatGPT and other chatbots responsibly and ethically. Additionally, they urge schools to form a detailed written approach concerning the usage of AI. This will bring clearance to both teachers, students and caregivers. The approach of schools should be based on factors such as students' age, the level of knowledge about AI for pedagogical staff and which support teachers and students can get. Skolverket also gives examples, divided into three sections (regarding assignments as a basis for assessment, AI tools in teaching and AI tools in other work), of how the approaches can be formulated.

Furthermore, Skolverket informs about the potential risks of using chatbots like ChatGPT. For example, informing about learning requires effort and students may miss knowledge intake as a consequence of using ChatGPT as a tool. They invite teachers and pupils to reflect upon which type of knowledge benefits and not from letting students use ChatGPT. Language development is given as an example of a negatively exposed field in school. Additionally, factual errors, data security and the fact that chatbots do not proceed from schools' regulatory documents are other risks that teachers should be aware of when using chatbots.

Skolverket gives the advice to not give grading-based assignments if the reliability can not be verified because ChatGPT can create essays.

3 METHOD

To investigate the impact that ChatGPT has on perceived learning in education, two surveys were conducted, one where the target group were students in higher education and one where the participants were teachers in upper secondary schools. For both surveys the region was northern Sweden. Furthermore, a learning session with students was conducted. The chosen methods were designed to provide an understanding of the participants' experiences using ChatGPT and they were influenced by a previous study by Chan and Hu [9] in 2023, investigating students' voices on generative AI. By employing a combination of a questionnaire and a learning session, the aim was to capture the nuances and dynamics associated with the integration of ChatGPT in educational settings.

3.1 Participants

The participants in the two studies are described below.

3.1.1 Study for students. The participants targeted for this study were students between the ages of 20 and 30 who were at the time pursuing a university-level education. This demographic group was chosen to focus on a population segment where the impact of tools like ChatGPT may have had significant implications for learning experiences. Specifically, students who utilized ChatGPT as part of their learning processes were included in the target group, allowing for a more specific examination of the tool's influence on their educational experiences. Furthermore, only students who use ChatGPT daily were chosen to participate in the study. In total 15 students participated in the study where 9 (60%) were females and 6 (40%) were males. A total of 12 (80%) participants were between the ages of 24-27, and 3 (20%) were between the ages of 20-23.

3.1.2 Study for teachers. A convenience sampling method was employed to select respondents based on their availability and willingness to partake in the study. 70 participants working at 6 different schools located in northern Sweden were contacted through email. All participants had to fill out a declaration of consent before participating in the survey. The participation was completely voluntary, the responses were anonymous and they could end their contribution whenever without sending in their answers.

Altogether, 38 people participated in the survey whereas 3 respondents were not eligible due to they were not working as teachers in the upper secondary school today. The not eligible respondents were redirected to the end of the survey, resulting in 35 respondents in total, 18 (51%) men and 17 (49%) women. The age distribution among the respondents can be seen in table 1.

3.2 Study for students

To collect quantitative data and gather insights from the selected target group, a structured questionnaire was developed. Google Forms were used as the platform for creating the questionnaire. The questionnaire included questions designed to assess perceived potential learning outcomes related to the use of ChatGPT in educational settings. Close-ended 5-level Likert scale questions were

Table 1: Age distribution for the respondents (n = 35).

Age	Distribution
29 or younger	6%
30–39	23%
40–49	26%
50–59	37%
60 or older	9%

constructed for efficiency, making it faster for participants to complete the questionnaire, as well as to obtain readily quantifiable data for analysis. The Likert scale questions were on a scale of “Don’t agree at all” to “Totally agree”.

The questions in the questionnaire were based on previous research [31], [21], [28], where perceived learning was the main focus. However, since those studies evaluated students’ perceived learning over a longer period (1 month or longer), the questions created for the questionnaire were modified to fit the context of this study.

The study was conducted in four parts:

- (1) Before participating in the study the students were inquired about their ChatGPT usage, if they did not use ChatGPT daily they were not eligible to participate.
- (2) The questionnaire was then sent to the participants. The first part of the questionnaire gathered demographic information about the participants. The second part posed questions about the students’ knowledge of a scientific subject, in this case magnetism. This part of the questionnaire gathered the pre-test data used in the study. The key concepts were magnetic materials, magnetic fields, poles of a magnet and magnetic domains. The expectation was that the students should have only a basic understanding about the subject without having any deeper knowledge about the topic.
- (3) In the third part of the questionnaire, all participants were given 10 minutes to learn about magnetism while using ChatGPT to gather information and explain the key concepts.
- (4) The fourth and final part of the questionnaire gathered information about the students’ perceived learning when using ChatGPT to learn about magnetism. This part of the questionnaire gathered the post-test data used in the study.

A test-study was created in order to gather feedback on the questions and overall design of the questionnaire and study. All participants were recruited through social media. Prior to answering the questionnaire, all participants provided their informed consent.

3.3 Study for teachers

For the teachers, the survey consisted of 13 questions, 3 of which were in free-text format and the rest were either multiple-choice or Likert scale. For the Likert questions could respondents rate totally disagree; disagree; partially agree; totally agree; or do not know. All questions were mandatory to answer except the free-text questions. The questions were divided into three different sections: Initial descriptive questions, How students can use ChatGPT and

Knowledge about and support for the use of ChatGPT. All questions and instructions were written in Swedish and later translated into English for the data analysis.

A pilot test was carried out resulting in valuable insights. This included feedback as changing from five to four options for Likert scale questions, adding a question on how long the respondents have been working as teachers and adding free-text questions letting respondents clarify and further explain their answers.

3.4 Data Analysis

The data gathered for the student study were on a 5-level Likert scale, the data ranged between 1 and 5 for each question. No steps were taken to preprocess or clean the data. Before analyzing the data, two arrays were created filled with all values for each question and each participant, one array for pre-test scores and the other one for post-test scores. The first step taken when analyzing the data was to check for normality. A QQ-plot was plotted as well as a Shapiro-Wilk test for each question in the arrays. The mean and standard deviation for the post- and pre-test for each question was calculated.

Since the data was not normally distributed a normal t-test could not be conducted. Instead a paired Wilcoxon signed-rank test was done with questions 1-6 in both arrays, where question 1 from the pre-test was matched with question 1 from the post-test and so on. This was done to see if there were statistically significant differences between the paired observations.

The data gathered for the teachers was analysed with a descriptive analysis, and a thematic analysis approach was applied to examine the responses to the open-ended questions in the survey. Data from the Likert-scale questions were prepared before calculating the mean and standard deviation. This was done by converting the scale alternatives into 1-4 and removing ‘I do not know’ answers. The decision to undertake a thematic analysis was influenced by the previously mentioned study by Chan and Hu in 2023. The thematic analysis in this study was based on the method described by Braun and Clarke in 2006. The open-ended questions were coded, using open coding, to label phrases and sentences. By identifying patterns among the codes themes were generated. Lastly, the created themes were reviewed, named and defined.

3.5 Limitations

The use of a questionnaire and a learning session for quantitative data carries certain limitations. Potential limitations included sample bias, as the study focused on a specific age group recruited from the same networks. Recruiting participants from the same networks risks creating a homogeneous sample, potentially limiting the variety of perspectives and experiences within the study. Individuals within the same network may share common characteristics or attitudes, affecting the external validity of the study. Additionally, participants recruited from the same networks might be influenced by shared social norms and/or expectations, potentially leading to biased responses. Social desirability bias, where participants respond in a manner perceived favorably by others, could be more pronounced in a closely-knit sample.

While a 10-minute learning duration is a practical choice for efficiency, it poses limitations as to the assessment of potential benefits and/or drawbacks of using ChatGPT for learning. Assessing learning outcomes immediately after a short learning session does not capture the sustainability of knowledge retention over time. Evidently, long-term retention is crucial for evaluating the effectiveness of a learning tool, and a longer duration would allow for a more comprehensive examination of this aspect. Other aspects, such as subject complexity and varied learning paces also had to be considered. The complexity of the subject matter (in this case magnetism) needs to be considered. Certain topics may need longer sessions for effective learning, especially when using a tool like ChatGPT. Individuals have different learning paces, and a fixed 10-minute duration may not be enough time for all participants to learn. Some participants might grasp concepts quickly, while others may require more time for in-depth comprehension.

Further, the participants might not use ChatGPT the same way. Some might use ChatGPT more as an interactive conversation whilst others might use ChatGPT primarily for information retrieval. This then may lead to differences in learning outcomes where one participant might learn more than the other.

4 RESULT

This section presents the results of the conducted surveys.

Table 2 presents descriptive statistics for the students' responses to their understanding and confidence in various aspects regarding magnetism. The responses were measured on a 5-level Likert scale, with scores ranging from a low of 1 to a high of 5. The "Mean" column provides the average score for each pre-test question and the "SD" column indicates the variability around the mean for each pre-test question. Furthermore, an internal reliability consistency estimate of reliability was calculated using Chronbach's coefficient alpha. The reliability score for the pre-test scores was 0.78, which shows an acceptable score of reliability.

Table 2: Mean and SD for each pre-test question.

Question (Likert scale: 1-5)	Mean	SD
1. I can produce a subject guide for other students.	1.27	0.59
2. I feel that I have an understanding of the basics within magnetism.	2.33	1.11
3. I know what magnetic materials are.	2.87	1.25
4. I feel confident in explaining the concept of magnetic fields and their impact on materials.	1.80	0.94
5. I have an understanding of the poles (north and south) of a magnet and how they interact.	3.33	1.05
6. I know what magnetic domains are.	1.20	0.41

Table 3 provides a summary of the students' responses to post-test questions assessing their perceived learning outcomes following an instructional learning session on magnetism. The Likert-scale questionnaire consisted of eight statements, each capturing different aspects of the participants' experiences and perceptions. The "Mean" column reports the average scores for each post-test question, offering insights into the central tendency of participants' perceived learning outcomes. Simultaneously, the "SD" (Standard Deviation) column indicates the dispersion or variability around the mean for each statement. An internal reliability consistency estimate of reliability was calculated using Chronbach's coefficient alpha for the post-test questions as well. The reliability score for the post-test scores were 0.75, which also shows an acceptable score of reliability.

Table 3: Mean and SD for each post-test question.

Question (Likert scale: 1-5)	Mean	SD
1. I can produce a subject guide for other students.	2.20	1.01
2. I feel that I have a better understanding of the basics within magnetism after this session.	4.33	0.49
3. The information about magnetic materials enhanced my knowledge in this area.	4.00	0.38
4. I feel more confident in explaining the concept of magnetic fields and their impact on materials.	3.40	0.99
5. My understanding of the poles (north and south) of a magnet and how they interact is clearer to me now.	3.93	0.70
6. The explanation of magnetic domains improved my understanding of the subject.	3.80	1.01
7. The interactive session where I could ask questions positively contributed to my learning experience.	4.33	0.82
8. In general, I feel that I have understood the subject content well.	3.47	0.74

4.1 Descriptive information

Figure 1 depicts the distribution of the teachers' fields of study. It can be seen that Natural science subjects (Chemistry, Physics and Biology), Swedish or Swedish as a second language and Social science subjects (Geography, History, Religion and Social science) were over-represented. Note that the respondents could mark multiple fields of study.

The distribution of the time the teachers have been working as teachers in the upper secondary school is presented in Table 4.

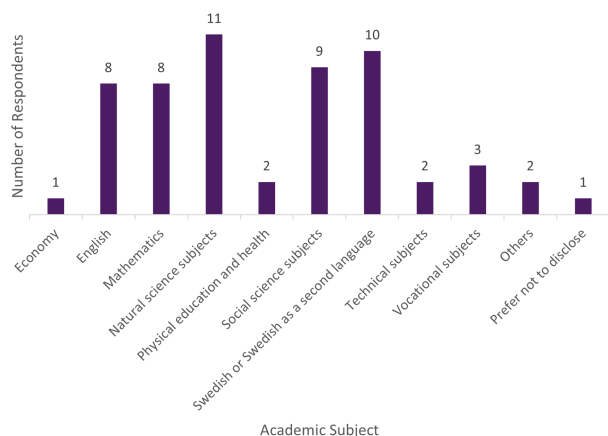


Figure 1: Distribution for responding teachers' fields of study (n = 35).

Table 4: How many years the teachers had worked as teachers (n = 35).

Years	Distribution
0–5	14%
6–10	23%
11–15	11%
16–20	23%
21–25	6%
Longer than 25	11%

4.2 Teachers' knowledge of ChatGPT

The distribution for teachers' usage of ChatGPT is presented in Table 5. The majority had tested ChatGPT a few times, while others had heard about it but never used it themselves. The remaining respondents indicated regular usage, either monthly, weekly or daily. The survey was constructed so that participants who were unfamiliar with ChatGPT were redirected to the end. However, the results show that none of the respondents reported that alternative.

Table 5: The teachers' usage of ChatGPT (n = 35).

Statement	Distribution
Heard about, but never used	11%
Tried a few times	43%
Use it monthly	20%
Use it weekly	23%
Use it daily	3%

One survey section examined teachers' perceived understanding and information regarding incorporating ChatGPT in upper secondary education. The questions asked in this section can be found in Table 6 together with the number of respondents that took a stand for each statement. The mean and standard deviation for the Likert questions are also presented. A higher mean indicates a greater level of agreement with a given statement, signifying an elevated level of knowledge and information within that particular context. The highest value that the mean could take was 4.

Table 6: Teachers' perception about whether they have sufficient information and knowledge to integrate ChatGPT in upper secondary schools.

Statement	n	Mean	SD
Inform students about the advantages and disadvantages of using ChatGPT in school	35	2.86	0.97
Inform students about how they can use ChatGPT in their studies	35	2.71	0.96
Use ChatGPT within the subject I teach	35	2.71	0.96
Adjust the education due to ChatGPT	35	2.66	1.03
Adjust examination due to ChatGPT	34	2.65	1.12

One teacher wrote in an open-ended question that teachers and principals have a deficient understanding of ChatGPT: *"Most teachers/principals are far too unfamiliar with ChatGPT to even form an opinion on how it should/should not be used"*.

4.3 Support

Table 7 shows how respondents perceived that they received support and knowledge on the integration of ChatGPT into their professional roles. Colleagues emerged with the highest mean score, followed by the Swedish National Agency for Education, and school management with the lowest mean. It is worth mentioning that five teachers expressed uncertainty about whether they received support or knowledge from the Swedish National Agency for Education. Here a higher mean indicates an enhanced perception of support and knowledge.

Table 7: Teachers' perceived received support and knowledge regarding the integration of ChatGPT in upper secondary education.

Statement	n	Mean	SD
Colleagues	34	2.46	0.98
Swedish national agency for education	30	2.00	0.95
School management	35	1.74	0.90

The questionnaire included a concluding question where the teachers could provide additional comments if necessary. One of

the teachers explained the situation regarding support from their point of view: “Regarding support for AI, at the moment, I respond with “disagree” at the school level, as we lack genuine routines and guidelines, as well as in-service training on AI, even though we have mentioned it in a couple of meetings. However, the school formed a special teacher group last week to work on these issues and create guidelines and support that more colleagues can use, and I am also part of it. So it is at least in progress, but at this pace, it will take several years before AI is considered comprehensively, both regarding cheating and potential”. Another teacher conveyed a sense of not receiving sufficient assistance or support at a level that meets their needs: “My experience is that teachers are left entirely alone in this, or offered short ‘training’ sessions on the subject that feel extremely outdated for an experienced computer user”.

4.4 Teachers voice on how ChatGPT should be used by students

The teachers answered Likert scale questions regarding in which situations or activities ChatGPT could be used by students in an educational situation. The results showed that they were most positive about using ChatGPT during the learning process (mean = 3.51), for understanding phenomena and concepts (mean = 3.41) and for being creative (mean = 3.39). On the contrary, teachers showed the least positive attitude towards ChatGPT usage for improving the content in an assignment (mean = 2.66), for solving a problem (mean = 2.81) and for identifying mistakes (mean = 2.84). See these results together with other situations in Table 8.

Table 8: Teachers’ attitudes towards students’ usage of ChatGPT in different school situations.

Situation	n	Mean	SD
During the learning process	35	3.51	0.61
For understanding phenomena and concepts	34	3.41	0.78
For being creative	31	3.39	0.80
For structuring oral presentations and texts	33	3.18	0.92
For guidance on the approach to completing a school assignment	32	3.16	0.85
For clarifying school tasks and assignments	32	3.06	0.84
For identifying mistakes	31	2.84	1.07
For solving a problem	26	2.81	1.06
For improving the content in an assignment	32	2.66	1.04

In addition to the situations in Table 8 the teachers could suggest other situations. This was done through two open-ended questions: “You can add additional situations in school where it is **not** appropriate for students to use ChatGPT” filled in by 25 respondents and “You

can add additional situations in school where it is appropriate for students to use ChatGPT” answered by 14 respondents.

4.4.1 Situations where ChatGPT may not be suitable for use. The answers showed two common additional situations when teachers do not think ChatGPT is appropriate to use in upper secondary school. These two themes of situations can be seen in table 9. There it can also be seen how many percent of the respondents mentioned the theme in one way or another. The two situations will be described further in the following paragraphs.

Table 9: Situations when ChatGPT should not be used by students according to the teachers (n = 35).

Situation	Mentioned by
Not during or related to any type of examination	54%
Not when a task primarily aims to practice a student’s abilities	9%

Firstly, the most frequently mentioned situation theme in the answers was that ChatGPT should not be used during or related to any type of examination. One teacher formulated as follows: “All situations where the student completes assignments that are intended to serve as the basis for grading”. Another teacher said “In all cases where the student’s knowledge is assessed. It is not possible to evaluate a student who only demonstrates what chatbots can do”. One examination form that is mentioned explicitly is assignments, which one teacher means should not be a form of examination no matter the existence of ChatGPT or not: “Assignments. It’s just a matter of stopping handing them out. Which, to be honest, should have been done a long time ago since it’s not possible to determine who has done what when the process is not monitored”. Additionally, preparing an oral presentation should not be done by letting ChatGPT do the job according to a teacher: “At examinations, assignments and when creating oral presentations”.

Secondly, teachers emphasised that students should refrain from using ChatGPT when a task primarily aims to practice a student’s abilities. This could be backed up by a quotation from one of the teachers: “It’s also important that the student doesn’t solely turn to GPT for assistance but tries on their own first. The brain needs to be challenged and exercised, and it’s beneficial if this is done both individually and collaboratively with other people”. Additionally, another teacher wrote “Using GPT to write texts or respond to analysis questions is not ideal because it does not train the skills that the student needs to develop”.

4.4.2 Situations where ChatGPT may be suitable for use. Regarding situations where teachers thought ChatGPT was appropriate to use, four themes could be identified. These can be seen in Table 10 together with the corresponding percentage of respondents that mentioned something related to the theme.

The first situation theme when teachers believed that students could use ChatGPT was for getting explanations of phenomena and

Table 10: Situations when ChatGPT can be used by students according to the teachers (n = 35).

Situation	Mentioned by
For getting explanations of phenomena and concepts	14%
For summarising	11%
As a sounding board	9%
When studying for an exam	6%

concepts (mentioned by 5 respondents, 14%). One of the respondents compared using ChatGPT to using Google: *“To get a quick answer, it’s like googling”*. Respondents meant that ChatGPT is useful to explain things in other words so the student can understand it better. This is one example quotation of this: *“To simplify complex concepts or passages of text”*.

Secondly, summarising was another theme that could be identified among the answers (mentioned by 4 respondents, 11%). Related to this theme teachers explained how they believed that ChatGPT could be used to summarise key takeaways from texts. One respondent wrote *“To summarise the most important parts in various studies”*, and another wrote *“To summarise large amounts of text”*.

Several respondents mentioned that students could use ChatGPT as a sounding board in their studies (mentioned by 3 respondents, 9%). One respondent explained that ChatGPT could be used as a sounding board in a creative process *“As a sounding board to progress in a creative process, for example”*. Another respondent emphasised that ChatGPT contributes to making learning more like a dialogue rather than just pure reading: *“A good way to work differently, as it becomes more of a conversation instead of just pure reading”*.

Lastly, teachers mentioned that ChatGPT can be a useful tool when studying for an exam (mentioned by 2 respondents, 6%). Related to this theme wrote participants *“Research for analogue exams”*, and *“Requesting study questions for a specific area”*.

4.5 ChatGPT generates incorrect answers

The answers also revealed that teachers experienced that ChatGPT sometimes generates false, misleading and incorrect answers. This was mentioned by 7 teachers, which corresponds to 20%. Some teachers gave examples of tasks and different fields of study when ChatGPT generates incorrect information, one of these examples was regarding language and grammar *“It is generally difficult to know how ChatGPT can be suitable, as our experience when testing in the subject area has revealed that ChatGPT provides incorrect solutions on a language/grammar level”*. Another teacher mentioned calculations as a field where ChatGPT’s performance can be questioned *“Calculations perform poorly, so that should be avoided”*. The last situation mentioned here was mentioned by another respondent *“When people (students), for example, are supposed to learn how to argue (based on strong arguments, with a strong burden of proof, relevant factual arguments) and also be able to perform argumentation analysis, I believe that AI falls short and is almost useless to use. Additionally, a Likert question investigated whether teachers*

believe ChatGPT can give students incorrect or false information. The mean for this question was 3.47 and the standard deviation was 0.56. The high mean indicates that teachers believe that ChatGPT generates incorrect and false information.

5 DISCUSSION

The results of these studies suggest a positive influence of ChatGPT on students’ perceived learning outcomes, aligning with established research [10, 24] that underscores the positive impact of AI on student learning in higher education.

Analysis of the paired Wilcoxon Signed-Rank test indicates a statistically significant difference between pre-test and post-test scores for all questions except for question 5, which focused on magnetic poles (south and north). The reliability scores, both pre-test (0.78) and post-test (0.75), demonstrate acceptable internal consistency, supporting the reliability of the study’s measures. Notably, question 5 received a p-value of 0.098, close to the chosen 0.05 significance level. If a significance level of 0.1 was chosen for this study, all questions would be statistically significant, underscoring the continued relevance of question 5 and warranting further investigation. Despite efforts to select topics unfamiliar to students, question 5 suggests some pre-existing knowledge among participants, explaining the lack of a significant difference in pre-test and post-test scores.

In the learning session, participants took advantage of ChatGPT’s ability to comprehend and generate human-like text through NLP, enabling ChatGPT to respond to inquiries about topics in a linguistically natural and coherent manner [22]. ChatGPT’s flexibility was demonstrated as participants adjusted their queries to delve deeper into specific aspects of magnetism, highlighting its adaptability to individual needs [19].

Upon completion of the learning session, the students provided self-reports on their perceived learning. It is important to differentiate actual learning, defined as a relatively permanent change in behavior due to experience, from perceived learning, which refers to a student’s self-report of knowledge gain through self-reflection [4, 15]. While perceived learning is interesting to investigate, it is subjective. Thus, to gain a comprehensive understanding of ChatGPT’s educational utility, both actual and perceived learning should be investigated.

The results furthermore reveal a nuanced perspective on teachers’ attitudes towards the integration of ChatGPT in upper secondary education. Both concerns and enthusiasm can be identified in the respondents’ answers. The positive inclination towards incorporating ChatGPT into activities supporting learning, understanding complex topics and concepts, and fostering creativity aligns with previous research findings [30]. Teachers express neutral to reserved views on using ChatGPT for identifying mistakes, solving problems, and improving the content of an assignment, suggesting reservations about relying on ChatGPT for tasks involving critical thinking, problem-solving, and content creation. Additionally, some teachers mentioned concerns about ChatGPT generating incorrect or false information, highlighting the challenge of verifying answers provided by ChatGPT.

The survey results also indicate that teachers may lack sufficient information and knowledge regarding the use of ChatGPT in upper

secondary education. Around 50% of the teachers expressed concerns about using ChatGPT during various forms of examinations. Moreover, teachers reported a lack of support or knowledge on how to address ChatGPT in their work, particularly from school management and educational authorities. This underscores the need for comprehensive support and guidance for educators in implementing AI technologies like ChatGPT in educational settings.

Both teachers and students see the potential in using generative AI within educations, but in different ways. Teachers worry about how it can be used in a good way, where students simply use it and figure out good ways of implementing it. The teachers participating in this study had ideas about where ChatGPT can be useful and where it should not be used, but the students are actually using it, and by doing so, they find where it is possible to use. As most teachers pointed out, there are cases where ChatGPT can be used, but it would not be beneficiary for the students to actually do so; learning is a process and the path to a learning outcome is sometimes more important than the actual lesson.

5.1 Limitations and future work

For optimal results, it is crucial that the study's subject matter is unfamiliar to students, and the questions do not pertain to well-known topics. However, it is essential to acknowledge certain limitations to the study, particularly regarding the sample size. Expanding the participant pool to a larger, more diverse population may enhance the ability to generalize the results. Since participating students were solely recruited from social media, the demographic group was limited, and thus, results cannot be generalized.

Future research avenues could explore learning retention, closely tied to perceived learning. Extended tests lasting more than 10 minutes could provide additional insights into how ChatGPT influences overall learning among higher education students.

Regarding the study on teachers' attitudes towards the integration of ChatGPT in upper secondary education, it is evident that 35 respondents may not be sufficient to represent the targeted study population accurately. The respondents do not fully represent the population of all teachers in Swedish upper secondary schools, given their selection through convenience sampling. Specifically, teachers working at schools in northern Sweden were contacted and recruited through the mail, potentially introducing biases in the results.

Furthermore, the scenarios presented in Table 8 were not derived from previous research but were made up and refined during pilot testing. This approach raises the possibility that certain noteworthy and relevant situations may have been overlooked.

Suggestions for further work include conducting a similar study on a larger group of respondents, better representing the targeted population and using random sampling during recruitment. Additionally, exploring different fields of study and comparing them, or examining teachers with varying levels of knowledge about AI and ChatGPT, could provide valuable insights. Lastly, investigating how school management in Sweden supports the integration of AI technologies like ChatGPT and suggesting improvements could contribute to enhancing educational practices.

None of the students or teachers mentioned anything about Vygotsky's Zone of Proximal Development (ZPD). The questionnaire

didn't mention this since we wanted to investigate their perceptions of using ChatGPT without having to relate it to theory. The ZPD was also a concept that most students were unfamiliar about and we deemed that these questions wouldn't yield useful answers. A future study should definitely include questions relating to the ZPD since this is the core questions with usage of ChatGPT: does the usage simply save time for the user or does it affect the users actual learning in a bad way?

6 CONCLUSION

One of the studies in this paper explored the impact of ChatGPT on perceived learning in higher education, focusing on students aged 20 to 30. Through a combination of a questionnaire and a learning session, insights into students' perceived learning experiences with ChatGPT as a learning tool were gathered.

In conclusion, the results suggest a positive influence of ChatGPT on students' perceived learning. The statistical significance difference between pre-test and post-test scores indicates that students subjectively felt they increased their learning during the learning session. However, drawing firm conclusions solely based on perceived learning is challenging. Future research should focus on both actual learning and perceived learning to gain a better understanding of how ChatGPT affects students in higher education.

Additionally, despite being launched for around two years, the Swedish National Agency for Education, school leadership, and teachers have not fully adapted to the existence of ChatGPT. Lack of knowledge, information, routines, and support may contribute to the challenges teachers face in identifying suitable scenarios for ChatGPT usage. However, the results provide examples of when teachers believe ChatGPT can be useful for students in their studies and when it is less appropriate to use, such as during or related to any form of examination.

Given the limited sample size of responding teachers, the results should be viewed as indicative. Nevertheless, they can serve as a foundation for future studies exploring the role of ChatGPT within the context of Swedish upper secondary schools, informing educators, policymakers, and researchers alike.

Together, the studies indicate that generative AI will be positive for the learning outcome of the students, but there are several steps to be taken to ensure that this powerful method is used properly and that teachers can trust that it is being used fairly.

REFERENCES

- [1] Malak Abdullah, Alia Madain, and Yaser Jararweh. 2022. ChatGPT: Fundamentals, Applications and Social Impacts. In *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, Milan, Italy, 8 pages. <http://doi.org/10.1109/SNAMS58071.2022.10062688>
- [2] Jad Abi-Rafieh, Hong Hao Xu, Roy Kazan, Ruth Tevlin, and Heather Furnas. 2023. Large Language Models and Artificial Intelligence: A Primer for Plastic Surgeons on the Demonstrated and Potential Applications, Promises, and Limitations of ChatGPT. *Aesthetic Surgery Journal* 44, 3 (2023), 329–343. <http://doi.org/10.1093/asj/sjad260>
- [3] Mohanad Abukmeil, Stefano Ferrari, Angelo Genovese, Vincenzo Piuri, and Fabio Scotti. 2021. A Survey of Unsupervised Generative Models for Exploratory Data Analysis and Representation Learning. *Comput. Surveys* 54, 5 (2021), 1–40. <https://doi.org/10.1145/3450963>
- [4] Donald R Bacon. 2016. Reporting actual and perceived student learning in education research. *Journal of Marketing Education* 38, 1 (2016), 3–6.
- [5] David Baidoo-Anu and Leticia Owusu Ansah. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of

- ChatGPT in Promoting Teaching and Learning. *SSRN eLibrary* (2023), 11 pages. <http://doi.org/10.2139/ssrn.4337484>
- [6] Som Biswas. 2023. ChatGPT and the Future of Medical Writing. *Radiology* 307, 2 (2023), e223312. <https://doi.org/10.1148/radiol.223312>
 - [7] Omar Boukker. 2023. From chatting to self-educating: Can AI tools boost student learning outcomes? *Expert Systems with Applications* 238, Article 121820 (2023), 13 pages. <https://doi.org/10.1016/j.eswa.2023.121820>
 - [8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <http://doi.org/10.1191/1478088706qp0630a>
 - [9] Cecilia Ka Yuk Chan and Wenjie Hu. 2023. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 1 (2023). <http://doi.org/10.1186/s41239-023-00411-8>
 - [10] Lijia Chen, Pingping Chen, and Zhijian Lin. 2020. Artificial intelligence in education: A review. *IEEE Access* 8 (2020), 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
 - [11] Peter A. Cotton Debby R. E. Cotton and J. Reuben Shipway. 2024. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* 61, 2 (2024), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
 - [12] Gunther Eysenbach. 2023. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Medical Education* 9 (2023), e46885. <http://doi.org/10.2196/46885>
 - [13] Nino Fijačko, Lucija Gosak, Gregor Štiglic, Christopher T. Picard, and Matthew John Douma. 2023. Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation* 185, Article 109732 (2023). <https://doi.org/10.1016/j.resuscitation.2023.109732>
 - [14] Sukhpal Singh Gill, Minxian Xu, Panos Patros, Huaming Wu, et al. 2024. Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems* 4 (2024), 19–23. <http://dx.doi.org/10.1016/j.iotcps.2023.06.002>
 - [15] Richard D. Gross. 2010. *Psychology: The science of mind and behaviour*. Hodder Arnold.
 - [16] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. 2023. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2023), 3313–3332. <https://doi.org/10.1109/TKDE.2021.3130191>
 - [17] Abid Haleem, Mohd Javaid, and Ravi Pratap Singh. 2022. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 2, 4, Article 100089 (2022). <http://doi.org/10.1016/j.tbench.2023.100089>
 - [18] Geoffrey Hinton and Terrence J Sejnowski. 1999. *Unsupervised learning: foundations of neural computation*. MIT press.
 - [19] Dinesh Kalla, Nathan Smith, Fnu Samaah, and Sivaraju Kuraku. 2023. Study and Analysis of Chat GPT and its Impact on Different Fields of Study. *International Journal of Innovative Science and Research Technology* 8, 3 (2023), 828–829.
 - [20] Andreas Kaplan and Michael Haenlein. 2019. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business horizons* 62, 1 (2019), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>
 - [21] Leila Karimi and Tunku Badariah Tunku Ahmad. 2013. Perceived learning and satisfaction in a blended teacher education program: An experience of Malaysian teacher trainees. *Contemporary Educational Technology* 4, 3 (2013), 197–211. <https://doi.org/10.30935/cedtech/6103>
 - [22] Liddy, Elizabeth D. 2001. Natural language processing. In *Encyclopedia of Library and Information Science* (2 ed.). Marcel Decker, Inc., New York, NY.
 - [23] Chung Kwan Lo. 2023. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences* 13, 4, Article 410 (2023). <http://doi.org/10.3390/educsci13040410>
 - [24] Rose Luckin, Wayne Holmes, Mark Griffiths, and Laurie B. Forcier. 2016. *Intelligence unleashed: An argument for AI in education*. Pearson Education, London, UK.
 - [25] Brady D Lund and Ting Wang. 2023. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News* 40, 3 (2023), 26–29. <https://doi.org/10.2139/ssrn.4333415>
 - [26] Andrianos Michail, Stefanos Konstantinou, and Simon Clematide. 2023. UZH_CLYp at SemEval-2023 Task 9: Head-First Fine-Tuning and ChatGPT Data Generation for Cross-Lingual Learning in Tweet Intimacy Prediction. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, Ritesh Kumar, and Elisa Sartori (Eds.). Association for Computational Linguistics, Toronto, Canada, 1021–1029. <https://doi.org/10.18653/v1/2023.semeval-1.140>
 - [27] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2, Article 100041 (2021). <https://doi.org/10.1016/j.caeai.2021.100041>
 - [28] Thi Thuy An Ngo. 2023. The Perception by University Students of the Use of ChatGPT in Education. *International Journal of Emerging Technologies in Learning (Online)* 18, 17 (2023), 4–19. <https://doi.org/10.3991/ijet.v18i17.39019>
 - [29] OpenAI. 2022. *Introducing ChatGPT*. Retrieved March 27, 2024 from <https://openai.com/blog/chatgpt>
 - [30] Md. Mostafizer Rahman and Yutaka Watanobe. 2023. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences* 13, 9, Article 5783 (2023). <http://doi.org/10.3390/app13095783>
 - [31] Alfred P Rovai, Mervyn J Wighting, Jason D Baker, and Linda D Grooms. 2009. Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *The Internet and higher education* 12, 1 (2009), 7–13. <https://doi.org/10.1016/j.iheduc.2008.10.002>
 - [32] Stuart J Russell and Peter Norvig. 2010. *Artificial intelligence a modern approach*. Pearson Education, Inc.
 - [33] Skolverket. 2024. *Råd om AI, Chat GPT och liknande verktyg*. Retrieved March 27, 2024 from <https://www.skolverket.se/skolutveckling/inspiration-och-stod-i-arbetet/stod-i-arbetet/rad-om-ai-chat-gpt-och-liknande-verktyg>
 - [34] Jiahong Su and Weipeng Yang. 2023. Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education* 6, 3 (2023), 355–366. <http://doi.org/10.1177/20965311231168423>
 - [35] Matheus Vinicius Abadia Ventura and Antonio Carlos Pereira de Menezes Filho. 2023. ChatGPT: limitations, challenges and potential applications. *Brazilian Journal of Science* 3, 1 (2023), 65–68. <https://doi.org/10.14295/bjs.v3i1.427>
 - [36] L. S. Vygotsky. 1978. *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA.
 - [37] W Wang, Y Chen, and N Heffernan. 2020. A generative model-based tutoring system for math word problems. (2020).
 - [38] Lu Xu, Leslie Sanders, Kay Li, James CL Chow, et al. 2021. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR cancer* 7, 4, Article e27850 (2021). <https://doi.org/10.2196/27850>
 - [39] Xiaoming Zhai. 2022. ChatGPT user experience: Implications for education. *SSRN eLibrary* (2022). <https://doi.org/10.2139/ssrn.4312418>