Contemporary Educational Technology, 2024, **16**(4), ep528
ISSN: 1309-517X (Online)

OPEN ACCESS

Research Article

ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude

Carmen Köhler [1*]
0000-0002-6668-4658

Johannes Hartig [1]
0000-0001-6361-4374

[1] DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, GERMANY
[*] Corresponding author: c.koehler@dipf.de

Citation: Köhler, C., & Hartig, J. (2024). ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude. *Contemporary Educational Technology, 16*(4), ep528. https://doi.org/10.30935/cedtech/15144

ARTICLE INFO

Received: 23 Jul 2024

Accepted: 16 Aug 2024

ABSTRACT

Since ChatGPT-3.5 has been available to the public, the potentials and challenges regarding chatbot usage in education have been widely discussed. However, little evidence exists whether and for which purposes students even apply generative AI tools. The first main purpose of the present study was to develop and test scales that assess students' (1) knowledge about ChatGPT, (2) actual ChatGPT usage and perceived value of use, and (3) attitude towards ChatGPT. Our second aim was to examine the intercorrelations between these scales, and to investigate differences (a) across five academic fields (i.e., human sciences, social sciences, teaching profession, health sciences, and law and economics) and (b) between stages of education (i.e., number of semesters). *N* = 693 students from various German universities participated in our online survey. Quality checks (Cronbach's alpha, MacDonald's omega, and confirmatory factor analyses) show satisfactory results for all scales. The scales all positively relate to each other, except for the knowledge and attitude scales. This means that more knowledge about ChatGPT is connected to a less favorable attitude regarding the generative AI tool. Lastly, MANOVA and subsequent Bonferroni corrected ANOVA tests show that ChatGPT is mostly used by law and economics students, and most frequently by students in the third year of higher education.

Keywords: ChatGPT in higher education, student knowledge, student use, student attitude, scale development, assessment, large language models (LLMs)

INTRODUCTION

Since ChatGPT-3.5 (OpenAI, 2023) has been openly available to the public, the use of natural language processing (NLP) systems for generating texts has been a controversially discussed trend topic. While some try out these tools to explore and illustrate their capability, others find entertainment in evoking inaccuracies and provocative responses (Alkaissi & Mcfarlane, 2023; Mogali, 2023; Rasul et al., 2023; Rudolph et al., 2023; Shen et al., 2023; Xames & Shefa, 2023).

Reactions in the educational field have been mixed. Various discussion papers emerged, highlighting the potentials and opportunities of ChatGPT for student learning, but also voicing concerns about its application in educational settings (Chan & Hu, 2023; Farrokhnia et al., 2023; Kasneci et al., 2023; Kooli, 2023; Lim et al., 2023; Mills et al., 2023; Schön et al., 2023; Sok & Heng, 2023). Government departments and ministries have published handbooks for educational leaders and policy makers, providing guidance on integrating artificial intelligence (AI)-based applications in the classroom to enhance learning (Hessisches Kultusministerium, 2023; Office of Educational Technology, 2023). Several research papers focus on the role of generative AI tools particularly in higher education (Cotton et al., 2023; Fauzi et al., 2023; Firat, 2023; O'Dea & O'Dea, 2023; Rasul et al., 2023; Rudolph et al., 2023a, 2023b; Schön et al., 2023; Sullivan et al., 2023). Schön et al. (2023) proclaim that the use of AI-based applications will change higher education in its entirety, influencing the learning experience, the learning assessment, the teaching concepts, and the degree program regulations.

However, little evidence exists whether and for which purposes students in higher education even apply generative AI tools (Bonsu & Baffour-Koduah, 2023; Strzelecki, 2023). Further, O'Dea and O'Dea (2023) argue that the usefulness of generative AI technologies to improve learning yet has to be demonstrated. In order to examine influences of these tools on educational outcomes in the first place, scales that assess the knowledge about these tools, the extent of generative AI tool usage, and the attitude toward these tools are needed. Therefore, the present study focuses on students' knowledge about ChatGPT, their actual and intended ChatGPT use, and their attitude towards ChatGPT. It investigates how these constructs are related, and whether specific groups of students differ on these constructs.

University Students' Knowledge about Generative AI Tools

Very few psychometrically evaluated instruments exist that assess knowledge about generative AI tools. Wang et al. (2022) developed the "artificial intelligence literacy scale", consisting of twelve items on four dimensions (i.e., *awareness*, *usage*, *evaluation*, and *ethics*), with the dimension *usage* referring to the successful application of the AI tool (e.g., "I can skillfully use AI applications or products to help me with my daily work"). As is evident from the example, the items are based on self-assessments, which might be subjective.

In a similar fashion, Laupichler et al. (2023) developed a scale to assess AI literacy of non-experts with 38 items (e.g., "I can explain how machine learning works at a general level"), which students evaluated on a seven-point Likert scale (from "strongly disagree" to "strongly agree"). As for the scale developed by Wang, the assessment is highly subjective, and individuals with the same level of expertise might select very different categories.

University Students' Use of Generative AI Tools

Measurement instruments assessing student use are also relatively sparse. Grájeda et al. (2023) focused on a private university in Latin America, assessing the five dimensions *effectiveness use of AI tools*, *effectiveness use of ChatGPT*, *student's proficiency using AI tools*, *teacher's proficiency in AI*, and *advanced student skills in AI*. The developed items for the scales were subject (i.e., field of study) and teacher specific, however.

Strzelecki (2023) conducted a study at a Polish University to examine the predictors of students' intended and actual use of ChatGPT. The intention to use ChatGPT was measured with three items that inquired about the general plan to use ChatGPT "in the future", "in my studies", and "frequently". Actual ChatGPT use was assessed with a single item that asked about usage frequency. Bonsu et al. (2023) investigated the perceived usefulness, the ease of use, and the intentions to use ChatGPT in a diverse student sample at a Ghanaian university. The scales measuring perceptions and intentions of use consisted of seven items each, but actual ChatGPT use was assessed with a single item. Sallam et al. (2023) investigated the usage of ChatGPT in higher education in Jordan. The four usage subscales were *perceived usefulness*, *perceived risks*, *perceived ease of use*, and *behavior*. The behavior subscale consists of three items, inquiring about general ChatGPT use.

The limitations of the scales employed in the described studies are that items were either very general or that the scale consisted of a single item. In several studies, the quality of the scales was not evaluated. Further, many of the studies are based on the *technology acceptance model* (TAM; Davis, 1989) or its extension, the *unified theory of acceptance and use of technology* (UTAUT) model (Venkatesh et al., 2003, 2012). Although TAM has been applied in various contexts involving new technologies (e.g., Liu et al., 2009), the model has limitations. The two scales developed by Davis (1989) assess the perceived usefulness of the technology and the perceived ease of use, but not the actual usage. The UTAUT model also identifies predictors of the intention to use a technology, and although the intention is considered a good predictor for actual technology usage (Davis, 1989), the actual usage is not measured.

University Students' Attitude towards Generative AI Tools

Several scales measuring the general attitude toward AI exist (e.g., Schepman & Rodway, 2020; Sindermann et al., 2021), containing items such as "Artificial intelligence is exciting" (Schepman & Rodway, 2020) or "I fear artificial intelligence" (Sindermann et al., 2021). However, instruments measuring the attitude towards generative AI tools are sparse. Sallam et al. (2023) developed an attitude scale specifically regarding

ChatGPT in higher education, which consists of 13 items assessing the three subscales *perceived risks*, *technology/social influence*, and *anxiety*. They base their scales on the TAM.

The Present Study

The main purpose of the present study was to develop short scales (with a maximum of 10 items) assessing students'

(1) knowledge about ChatGPT,

(2) actual ChatGPT usage and perceived value of use, and

(3) attitude towards ChatGPT.

Note that our interest lies in large language model (LLM) applications in general, not in a specific software. After careful consideration, we deliberately decided to examine ChatGPT in particular, as it is presently the most widely known. Specifically, we developed a knowledge scale, a usage scale, and an attitude scale, which all contain no more than ten items. We further evaluated the dimensionality and the reliability of the scales. A secondary aim was to examine the intercorrelations between the scales, and to investigate group differences

(a) across academic majors and

(b) between stages of education.

The study is relevant insofar that students in higher education are increasingly learning online, using the Internet as their main source of information (Gasser et al., 2012; Maurer et al., 2019). The quality of web-based texts varies considerably, and students are confronted with the task of finding, selecting, and integrating online information in an adequate manner (Molerov et al., 2020; Nagel et al., 2020). In recent years, a growing body of research emerged in this area, investigating students' literacy and critical online reasoning skills (Koltay, 2011; Molerov et al., 2020; Murray & Pérez, n.d.; Nagel et al., 2020; Sparks et al., 2016; Zlatkin-Troitschanskaia et al., 2021). For the 2025 cycle, the Programme for International Student Assessment (PISA) even initiated the assessment *Learning in the Digital World*, which focuses on students' capacity to use technology for acquiring new knowledge and skills (PISA, 2023). The critical evaluation of internet and media content is a highly valued literacy, and one that both politicians and researchers deem relevant to promote (Hessisches Kultusministerium, 2023; National Research Council, 2012). The possibility of using generative AI tools adds to the information landscape, which increases the complexity of internet-based sources. Student knowledge and use of generative AI tools are relevant for investigating students' digital literacy and their ability to critically evaluate online information. It is therefore important to have psychometrically evaluated instruments that assesses students' knowledge about generative AI tools, use of generative AI tools, and attitude toward generative AI tools. In addition to the scale development, the present study investigates the relationship between the dimensions. Further, differences between academic tracks and between stages of education are examined.

METHOD

Construction of the Scales

We constructed the three scales *knowledge* about ChatGPT, ChatGPT *use*, and *attitude* towards generative AI tools in a bottom-up process. For the *knowledge* scale, we searched peer-reviewed articles for common misconceptions about ChatGPT, which, for example, resulted in the item "ChatGPT may provide content that is not based on facts". We further included well-documented errors ChatGPT makes, which a common user might not expect from an AI tool. This, for example, resulted in the item "ChatGPT can solve math problems reliably". Lastly, we used our own experience from teaching courses, which, for example, resulted in the item "ChatGPT performs web searches" (for a full list of items, see **Appendix A**).

Regarding the *use* scale, the potential theoretical foundations are the TAM (Davis, 1989) and its extended UTAUT model (Venkatesh et al., 2003, 2012). Since we aim to assess students' actual – and not their intended – ChatGPT use in our study, we constructed our own items. We developed specific study-related tasks based on the recommendations for LLM usage in higher education. Researchers claim that LLMs can facilitate adaptive learning when students use them as 'personal tutors' to help them understand a learning content better, as a writing tool, for example, for proofreading and editing, or for brainstorming (Chan & Hu, 2023; Cotton et al., 2023; Farrokhnia et al., 2023; Firat, 2023; Kasneci et al., 2023; Kooli, 2023; Lim et al., 2023; Rasul et al., 2023; Rudolph et al., 2023a; Sullivan et al., 2023). The items we developed are therefore direct and task specific (e.g., "I have used ChatGPT to get feedback on texts I wrote" or "I have used ChatGPT to have texts summarized"; for a full list of items, see **Appendix A**).

The development of the items for the *attitude* scale was also a bottom-up process that was largely based on print- and online-media articles discussing ChatGPT, and on our own experience in constructing items that assess attitude. For example, several articles pointed out high energy costs (Kasneci et al., 2023; Rasul et al., 2023), which resulted in the item "ChatGPT consumes too much electricity". "Chatbots such as ChatGPT are among the most important inventions of the 21st century" is an item we developed to enquire about the perceived relevance of generative AI tools.

After the construction, the items were evaluated by a group of subject matter experts. We revised the items accordingly.

Procedure

This study was approved by the ethics committee of the DIPF | Leibniz Institute for Research and Information in Education in Frankfurt. We sent the participation-invitation with the link to the survey to several colleagues from different fields and universities, who distributed it to students via e-mail, teaching platforms, and WhatsApp groups. The sample was therefore a convenience sample. We promoted the study by giving out amazon vouchers á 25 Euros to twenty randomly drawn participants. Entering the e-mail address to partake in the prize draw was optional. All participants gave their informed consent to the study.

Participants

Since we aimed at comparing students between different fields of studies and between stages of education, we used G*Power (version 3.1; Faul et al., 2009) to determine the necessary sample size prior to the data acquisition. Based on our recruiting strategy, we expected a larger body of students in at least four academic majors. For finding a medium effect of $f = .25$ between four groups when calculating a one-way multivariate analyses of variance (MANOVA) with four dependent variables when fixing the type-I-error at $\alpha = .05$ and the power at $1 - \beta = .9$, the required sample size was *n* = 31 students per group. With respect to the six stages of education (1. & 2. semester, 3. & 4. semester, 5. & 6. semester, 7. & 8. semester, 9. & 10. semester, and 11. & >11. semester), the required sample size was *n* = 19 students per stage of education. We kept the online survey accessible until those requirements were met.

The participants of the study were *N* = 730 students at various German universities, who filled out the questionnaire in November and December of 2023. We excluded 5% of students with the fastest response times, since the majority of those students showed suspicious response patterns (i.e., consistently selecting the same category) on at least one of the three scales. The remaining sample of *N* = 693 students were at various stages in their curriculum (see **Table 1**), ranging from the first undergraduate semester, which typically lasts for six semesters in Germany, to students in an advanced phase of their graduate degree, which typically lasts for four semesters. The students fell into one of ten academic fields (see **Table 1**). Mainly represented in the sample were students studying human sciences, social sciences, teaching profession, health sciences, and law and economics.

Measures

Additional to the three main scales knowledge, use, and attitude, we included three global items in the questionnaire. The first one assessed how well students feel informed about ChatGPT ("I feel well informed about ChatGPT") on a 4-point Likert scale from *absolutely disagree* to *absolutely agree*. The second global item asked about the general frequency of ChatGPT use ("How often do you use ChatGPT in context with your studies per month?") on a 4-point Likert scale with the response options *never*, *1-2 a month*, *weekly*, and *(almost) daily*. These two items were assessed prior to the three main scales. The third global item was included at the end of the questionnaire, inquiring whether universities should inform more about generative AI tools ("More

Table 1. Number of students listed by academic field and semester

Academic field	1./2. semester	3./4. semester	5./6. semester	7./8. semester	9./10. semester	11./>11. semester
Human Sciences	28	21	22	21	11	28
Sport Sciences	0	0	1	0	0	0
Law and Economics	8	11	6	10	10	6
Mathematics, Natural Sciences	3	3	1	0	1	0
Health Sciences	0	44	4	1	2	3
Agricultural, Forestry and Nutritional Sciences, Veterinary Medicine	0	0	0	0	0	0
Engineering Sciences	2	3	1	1	3	3
Art Sciences	3	3	1	1	1	0
Social Sciences	48	40	28	35	37	37
Teaching Profession	40	33	32	31	38	27
Total	132	158	96	100	103	104

information should be provided about chatbots (such as ChatGPT) in university courses") on a 4-point Likert scale from *absolutely disagree* to *absolutely agree*.

The *knowledge* scale consisted of six items with a dichotomous response format (*correct* vs. *incorrect*) and an *I don't know* option. We included a note on the top of the page, informing the students that the questions pertain to the free-of-charge version 3.5 of ChatGPT. In the *use* scale, items were formulated in the present perfect (i.e., "I have used ChatGPT to …"). The ten items had four response options:

a) *No, I don't intend to either*,

b) *no, but I can imagine it*,

c) *yes, but was not helpful*,

d) *yes, was helpful*.

This response scale therefore reflects two dimensions:

(1) dichotomous information about the usage (i.e., yes vs. no), and

(2) the perceived usefulness of ChatGPT for the specific task (i.e., people who do not intend to use it or do not consider it helpful vs. people who intend to use it or are already using it and finding it helpful).

The *attitude* scale consisted of eight items with a 4-point Likert scale from *absolutely disagree* to *absolutely agree*.

Statistical Analyses

In an initial step, we scored the responses on the *knowledge* scale such that responses in the *I don't know* category were considered incorrect. We also rescored items such that a correct response was scored as 1 and an incorrect response as 0. For the *use* scale, we scored the responses on two dimensions: *actual usage* and *use value* (see **Table 2**). The actual usage dimension therefore informs about whether a particular ChatGPT interaction takes place whereas the value dimension informs about the perceived usefulness of a particular ChatGPT interaction.

Table 2. Scoring of usage scale items to obtain the two dimensions actual usage and use value

Response options	No, I don't intend to either	No, but I can imagine it	Yes, but was not helpful	Yes, was helpful
Actual usage	0	0	1	1
Use value	0	1	0	1

All analyses were conducted in R (R Core Team, 2023). The syntax for our empirical analyses in R and the data are available on OSF (https://osf.io/dqby5/). To check the internal consistency among the items within each of the dimensions, we estimated Cronbach's alpha (Cronbach, 1951) and McDonald's omega (McDonald, 2013) using the psych package (version 2.3.9; Revelle, 2023) and the GPArotation package (version 2023.11-1; Bernaards & Jennrich, 2005), respectively. To test the unidimensionality assumption, we conducted confirmatory factor analyses (CFA) for each of the four dimensions *knowledge*, *actual usage*, *use value*, and *attitude* with the lavaan package (version 0.6-16; Rosseel, 2012). Model fit was assessed with the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). According to Hu and Bentler (1998), CFI and TLI values above .9, RMSEA values below .05, and SRMR values below .08 indicate a satisfactory to good model fit.

After the evaluation of the scales and the decision of whether to drop any of the items, we estimated the latent correlations between the four scales, running a four-factor CFA model. To investigate differences

(a) across academic majors and

(b) between stages of education, we conducted two one-way MANOVA, one for each factor.

Note that the data-base slightly differed between the two: For the MANOVA across academic majors, we only included students from the five main academic fields health sciences, human sciences, law and economics, social sciences, and the teaching profession; between stages of education, all students were included. For each MANOVA, the four dependent variables were the mean scale scores on the four dimensions. Note that the number of groups regarding the academic major comparison slightly differs from the one we used in our power analysis. However, for a comparison of five groups, a smaller number of students per group is necessary (20 students per group), which means our sample suffices to detect a medium effect. After testing whether the assumptions for conducting a MANOVA hold, we calculated the MANOVAs, and added post-hoc Bonferroni adjusted ANOVAs separately for each scale.

RESULTS

Descriptive Analyses

Regarding the global items, the results showed that most students rather agree to the statement that they feel well informed about ChatGPT (*M* = 2.66, *SD* = 0.76). Most students use ChatGPT only once or twice a month (*M* = 2.17, *SD* = 0.91). The majority of the students want to learn more about ChatGPT at their university (*M* = 3.37, *SD* = 0.74).

With respect to knowledge about ChatGPT, most students know that ChatGPT sometimes provides content that is not based on facts. The item which most students answer incorrectly was the one asking about ChatGPT reliably solving mathematical tasks, which many students believed to be true (see **Table B1**, **Appendix B**). Students mostly use ChatGPT to clarify subject content, and to get an overview of a new topic (see **Table B2**, **Appendix B**). They rarely use it for motivation or advice about their management. Regarding the items on the attitude scale, students are hardly concerned about ChatGPT needing too much energy or not being transparent with regard to the underlying text basis. Most agree to the item stating that Chatbots are among the most important inventions of the 21st century (see **Table B3**, **Appendix B**).

Scale Evaluation

Internal consistency was acceptable for all four dimensions (see **Table 3**). For the knowledge scale with a Cronbach's alpha below .6, we checked if removing any of the items would improve reliability, which was not the case. Note that this scale is the shortest one with only six items, and Cronbach's alpha increases with test length (Cronbach, 1951).

Table 3. Reliabilities and model fit indicators of CFA for the four scales knowledge about ChatGPT, actual usage of ChatGPT, the value of ChatGPT use, and attitude towards ChatGPT

	α	ω	CFI	TLI	RMSEA	SRMR
Knowledge	0.57	0.84	.95	.92	.06	.07
Actual usage	0.78	0.91	.95	.94	.08	.11
Use value	0.78	0.92	.93	.91	.10	.12
Attitude	0.70	0.82	.96	.94	.08	.06

The CFA for the knowledge scale revealed good model fit indicators (see **Table 3**). For the actual usage and the use value scales, the values were moderate (see **Table 3**). Given that the two dimensions were constructed by scoring the students' item responses on two dimensions, the results are still remarkable. The attitude dimension showed acceptable values (see **Table 3**).

Taking all quality checks into account, the scales work well in measuring *knowledge*, *actual usage*, *use value*, and *attitude*. We decided against removing any of the items for the consecutive analyses, since none of them decreased the reliability except for the item "Chatbots such as ChatGPT are amongst the most important inventions of the 21st century" on the attitude scale, which is probably due to the fact that people respond to it irrespective of other concerns regarding the chatbot. The correlation to the remaining rather critical statements is therefore lower. In terms of its content, we argue that it measures the attitude dimension and therefore kept it in our analyses. Future users of this scale might consider removing it.

Relationships Between Dimensions

The latent correlations between the four dimensions are depicted in **Table 4**. The highest correlation is between actual usage and use value, which makes sense and is also the result of both dimensions stemming from the same item response. However, the two dimensions measure different constructs, which is also evident in the deviating correlations to the other dimensions. For example, ChatGPT knowledge has a medium correlation with actual ChatGPT use but a low correlation with the value of ChatGPT use. The attitude toward ChatGPT has a medium positive relationship with actual usage and use value, but a weak negative relationship to knowledge. Obviously, people who are better informed about ChatGPT have a more critical attitude toward it.

Table 4. Latent correlation coefficients (and standard errors) between the four dimensions knowledge about ChatGPT, actual usage of ChatGPT, the value of ChatGPT use, and attitude towards ChatGPT (confidence intervals are displayed in the upper triangle)

	1	2	3	4
1. Knowledge	1	[.45, .63]	[.08, .30]	[-.30, -.08]
2. Actual usage	.52 (.046)	1	[.59, .72]	[.23, .41]
3. Use value	.19 (.054)	.66 (.032)	1	[.41, .57]
4. Attitude	-.19 (.056)	.32 (.046)	.49 (.041)	1

Group Comparisons

Between different study fields

The mean scale scores for the five main academic fields health sciences, human sciences, law and economics, social sciences, and the teaching profession are depicted in **Figure 1**. Differences between the five fields are smallest on the knowledge scale and largest regarding actual usage.

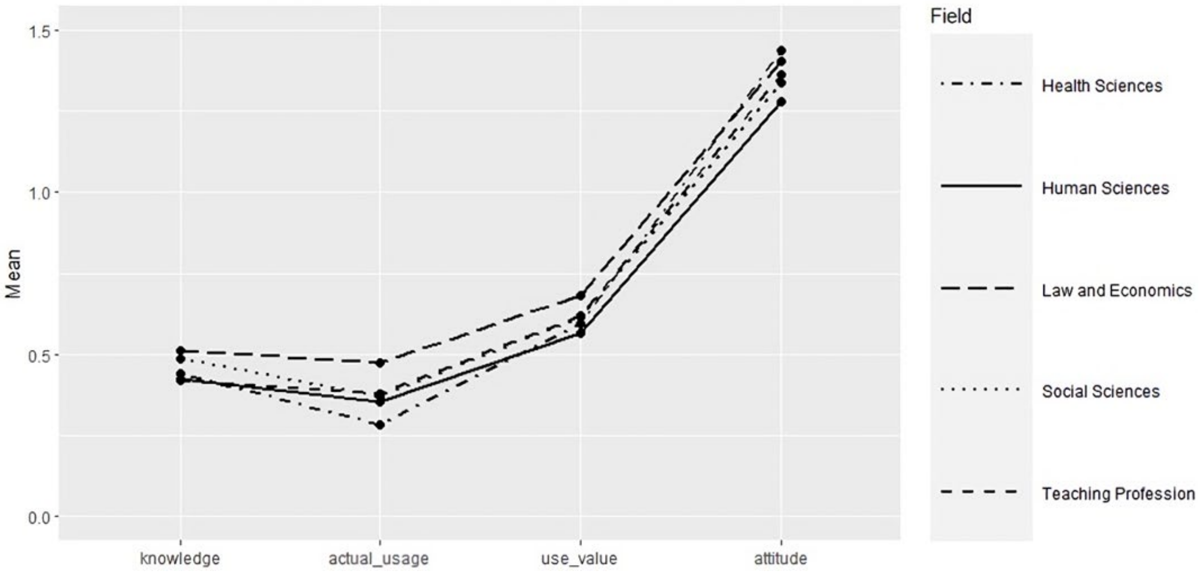

Figure 1. Mean scale scores for different study fields. Note that the attitude scale has a different metric (scale from 0 to 3) than the others (dichotomous 0/1). (Source: Authors)

The assumptions for conducting a MANOVA were met: The Q-Q plot to assess the multivariate normality of the variables showed no outliers, and Box's M test for homogeneity of the variance-covariance matrices was not significant. Using the Mahalanobis distance test, we detected one outlier, which was removed from the MANOVA. The one-way MANOVA to determine whether there is a difference between the five study fields on ChatGPT variables (*knowledge*, *actual usage*, *use value*, and *attitude*) was significant ($F(4, 656) = 2.45$, $p = .001$, partial eta squared $= 0.01$). Subsequent Bonferroni-adjusted ANOVA tests per dependent variable showed significant differences between the five study fields only regarding actual usage ($F(4, 656) = 3.89$, $p = .004$). Students in health sciences use ChatGPT the least while students studying law or economics use it the most. The effect size was low ($\eta^2 = .02$), however, meaning that that the study field did not have much explanatory value.

Between years of education

Descriptively, the knowledge about ChatGPT is greater for students who are at least in their second year of higher education as compared to those who are in their first year (see **Figure 2**). Actual usage is greatest for students who are in their third year of higher education. The value of usage seems rather unaffected by years of education. A positive attitude is highest in the first year of education.

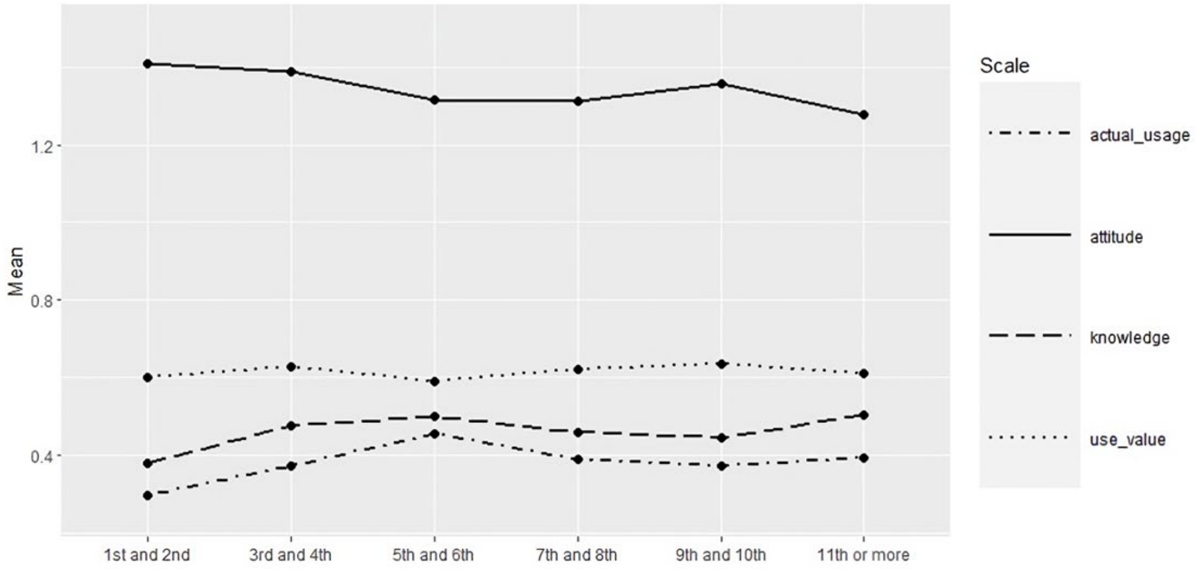

Figure 2. Mean scale scores for different semesters. Note that the attitude scale has a different metric (scale from 0 to 3) than the others (dichotomous 0/1). (Source: Authors)

As before, we conducted tests of whether the assumptions for conducting a MANOVA all held, which was the case. Again, we excluded one outlier – the same one – from the MANOVA. Results reveal a significant difference between years of education on ChatGPT variables ($F(5, 686) = 2.73$, $p < .001$, partial eta squared $= 0.02$). Post-hoc Bonferroni-adjusted ANOVA tests per dependent variable showed significant differences between the years of education regarding knowledge ($F(5, 686) = 4.03$, $p = .001$, $\eta^2 = .03$) and actual usage ($F(5, 686) = 4.41$, $p < .001$, $\eta^2 = .03$). Students at the beginning of their studies know the least about ChatGPT and are also the ones who use it the least. Students towards the end of their bachelor's or master's degree know the most and use it the most, especially those in their 5th or 6th semester (i.e., typically toward the end of a bachelor's degree).

DISCUSSION

The release of ChatGPT has raised controversies and concerns in the education sector. Practitioners and educational researchers are only just beginning to investigate and understand its implications for students and educational institutions. The purpose of the current study was to provide measurement instruments with which students’ knowledge about ChatGPT, their actual usage and perceived value of ChatGPT use, and their attitude towards ChatGPT can be assessed. These could be beneficial for future research, for example, to answer the question of whether heavy ChatGPT use hinders the development of certain skills such as problem solving or critical thinking. We further investigated the relationship between the four dimensions, and whether differences exist between academic tracks or between stages of education.

Our results showed satisfactory qualities of the developed scales. This means that they can be used in future studies that aim to assess any of the four dimensions knowledge, actual usage, value of use, or attitude. We further found that students who use ChatGPT more frequently also perceive it as more valuable, they know more about it, and their attitude toward it is more positive. At the same time, students with more background knowledge about the tool are more critical towards it. Another finding of the study was that general usage was not very high, with most students using it only once or twice a month. Law and economics students show the highest use, whereas students in health sciences use it the least. There were no significant differences between study fields on the other three dimensions. Regarding the stages of education, actual usage and knowledge were lowest for the first and second-semester students.

Our findings are in line with previous studies investigating students’ attitudes toward ChatGPT, indicating that these are generally positive (Abdaljaleel et al., 2024; Grájeda et al., 2023; Sallam, 2023). However, differences in student ChatGPT knowledge, use, and attitude are probable between different countries and other sub-populations, especially where exposure to generative AI tools is not as prevalent. Abdaljaleel et al. (2024) for example showed that the country of residence and the grade point average (GPA) were significant predictors of ChatGPT usage. Grájeda et al. (2023) even showed differences between students from distinct academic disciplines, which also existed in some of the study fields we investigated.

More future research is needed to understand how the use of generative AI affects the educational process (Firat, 2023). A current line of research focuses on integrating generative AI tools into the feedback process, for example, for essay writing or to give personalized formative feedback during problem-solving tasks (Banihashem et al., 2024; Küchemann et al., 2024). Despite the voiced fears that the use of generative AI tools might hinder the development of problem-solving skills (Chan & Hu, 2023; Farrokhnia et al., 2023; Kasneci et al., 2023; Sok & Heng, 2023; Sullivan et al., 2023), Kasneci et al. (2023) also see the possibility of LLMs enhancing students’ critical thinking skills in the classroom. Overall, generative AI tools have the potential to improve educational outcomes (Noroozi et al., 2024). However, a beneficial use of such tools must be trained and need a responsible integration into educational practices (Küchemann et al., 2024; Noroozi et al., 2024).

A limitation of our study is that it is not representative of all students in Germany. Several study fields are not represented at all. Also, the fact that we drew a convenience sample may distort the average values on the scales – overall and for the group comparisons. Since some of the colleagues we asked to distribute the link to the online survey only taught one or two courses at the time, students in a particular semester dominated the entire study field. This was particularly prevalent in health sciences, where the majority of the participants were in their third or fourth semester. This means that the study year and the study field are not completely independent of each other in our sample, which is relevant to keep in mind when interpreting results on group differences. Another limitation of the study is that we used manifest aggregated values, which contain measurement error, for the group comparisons. The differences between the groups could theoretically also be estimated in a structural equation model with the four dimensions as latent factors and the group variables as predictors. However, we decided against this approach, since our sample size per study field was rather small for such a complex model with potentially free item loadings for all groups.

Overall, the scales we developed lay the groundwork for further empirical investigations in investigating students’ literacy and critical online reasoning skills, as they allow the assessment of four relevant and person specific dimensions regarding generative AI use. They are applicable to students at various educational stages and can be adapted to other generative AI tools. Clearly, the item content needs to be updated continuously as changes in technology and the performance of the LLMs occur.

CONCLUSION

The usage of generative AI in educational settings is a pertinent topic, which brings opportunities and challenges. This study provided first valuable insights into students' current knowledge, actual use, perceived usefulness, and attitude towards ChatGPT as well as tools for assessing these dimensions. Generative AI literacy is closely linked to digital literacy and critical online reasoning skills and needs to be taken into perspective in terms of interrelations and a differentiation. All three are highly valued skills that need to be trained and promoted. It is therefore relevant to pursue this line of research, and to integrate its results into recommendations for generative AI implementation in educational settings.

Author contributions: CK: design and conceptualization of study, literature review, writing the first draft of the manuscript, data acquisition, data analysis; **JH:** design and conceptualization of study, revision of the manuscript. Both authors approved the final version of the article.

Acknowledgments: The authors would like to thank the reviewers who provided constructive feedback for this article.

Funding: This article was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the research group "Critical online reasoning in higher education (CORE)" (Project Number: 462702138).

Ethics declaration: This study was approved by the Ethics Committee of the DIPF | Leibniz Institute for Research and Information in Education in Frankfurt on 15 November 2023 (approval number: DIPF_EK_2023_36). All participants gave their informed consent to the study and to sharing the obtained data.

Declaration of interest: The authors declare no competing interest.

Data availability: The data and the syntax for the empirical analyses in R are available on OSF: https://osf.io/dqby5/

REFERENCES

Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A. R., Wayyes, A. M., Farhan, S. S., Khatib, S. E., Rahal, M., Sahban, A., Abdelaziz, D. H., Mansour, N. O., AlZayer, R., Khalil, R., Fekih-Romdhane, F., Hallit, R., ..., & Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports, 14*(1983), 1–14. https://doi.org/10.1038/s41598-024-52549-8

Alkaissi, H., & Mcfarlane, S. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus, 15*. https://doi.org/10.7759/cureus.35179

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1), Article 23. https://doi.org/10.1186/s41239-024-00455-4

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*(5), 676–696. https://doi.org/10.1177/0013164404272507

Bonsu, E. M., & Baffour-Koduah, D. (2023). From the consumers' side: Determining students' perception and intention to use ChatGPT in Ghanaian higher education. *Journal of Education, Society & Multiculturalism, 4*(1), 1–29. https://doi.org/10.2478/jesm-2023-0001

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00411-8

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education, 5*(4), 14886–14891. https://doi.org/10.31004/joe.v5i4.2563

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57–63. https://doi.org/10.37074/jalt.2023.6.1.22

Gasser, U., Cortesi, S., Malik, M. M., & Lee, A. (2012). Youth and digital media: From credibility to information quality. *SSRN*. https://doi.org/10.2139/ssrn.2005272

Grájeda, A., Burgos, J., Córdova, P., & Sanjinés, A. (2023). Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. *Cogent Education, 11*(1). https://doi.org/10.1080/2331186X.2023.2287917

Hessisches Kultusministerium. (2023). Handreichung "Künstliche Intelligenz (KI) in Schule und Unterricht" [Handout "artificial intelligence (AI) in schools and teaching"]. *Digitale Schule Hessen*. https://digitale-schule.hessen.de/unterricht-und-paedagogik/handreichung-kuenstliche-intelligenz-ki-in-schule-und-unterricht

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., …, & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*. https://doi.org/10.1016/j.lindif.2023.102274

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society, 33*(2), 211–221. https://doi.org/10.1177/0163443710393382

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), Article 5614. https://doi.org/10.3390/su15075614

Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *The Physics Teacher, 62*(5), 400–402. https://doi.org/10.1119/5.0212374

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "scale for the assessment of non-experts' AI literacy" – An exploratory factor analysis. *Computers in Human Behavior Reports, 12*, Article 100338. https://doi.org/10.1016/j.chbr.2023.100338

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education, 21*(2). https://doi.org/10.1016/j.ijme.2023.100790

Liu, S.-H., Liao, H.-L., & Pratt, J.-A. (2008). Impact of media richness and flow on e-learning technology acceptance. *Computers and Education, 52*(3), 599–607. https://doi.org/10.1016/j.compedu.2008.11.002

Maurer, M., Schemer, C., Zlatkin-Troitschanskaia, O., & Jitomirski, J. (2019). Positive and negative media effects on university students' learning: Preliminary findings and a research program. In O. Zlatkin-Troitschanskaia (Ed.), *Frontiers and advances in positive learning in the age of information (PLATO)* (pp. 109–119). Springer. https://doi.org/10.1007/978-3-030-26578-6_8

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16–30. https://doi.org/10.37074/jalt.2023.6.1.34

Mogali, S. R. (2023). Initial impressions of ChatGPT for anatomy education. *Anatomical Sciences Education, 17*(2), 444–447. https://doi.org/10.1002/ase.2261

Molerov, D., Zlatkin-Troitschanskaia, O., Nagel, M.-T., Brückner, S., Schmidt, S., & Shavelson, R. J. (2020). Assessing university students' critical online reasoning ability: A conceptual and assessment framework with preliminary evidence. *Frontiers in Education, 5*. https://doi.org/10.3389/feduc.2020.577843

Murray, M. C., & Pérez, J. (n.d.). Unraveling the digital literacy paradox: How higher education fails at the fourth literacy. *Issues in Informing Science and Information Technology, 11*, 85–100. https://doi.org/10.28945/1982

Nagel, M.-T., Schäfer, S., Zlatkin-Troitschanskaia, O., Schemer, C., Maurer, M., Molerov, D., Schmidt, S., & Brückner, S. (2020). How do university students' web search behavior, website characteristics, and the interaction of both influence students' critical online reasoning? *Frontiers in Education, 5*. https://doi.org/10.3389/feduc.2020.565062

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. https://doi.org/10.17226/13398

Neumann, M., Rauschenberger, M., & Schön, E.-M. (2023). "We need to talk about ChatGPT": The future of AI and higher education. In *Proceedings of the IEEE/ACM 5th International Workshop on Software Engineering Education for the Next Generation* (pp. 29–32). IEEE. https://doi.org/10.1109/SEENG59157.2023.00010

Noroozi, O., Soleimani, S., Farrokhnia, M., & Banihashem, S. K. (2024). Generative AI in education: Pedagogical, theoretical, and methodological perspectives. *International Journal of Technology in Education, 7*(3), 373–385. https://doi.org/10.46328/ijte.845

O'Dea, X., & O'Dea, M. (2023). Is artificial intelligence really the next big thing in learning and teaching in higher education? A conceptual paper. *Journal of University Teaching and Learning Practice, 20*(5). https://doi.org/10.53761/1.20.5.05

OpenAI. (2023). *ChatGPT*. https://chat.openai.com

PISA. (2023). *Learning in the digital world–PISA*. https://www.oecd.org/pisa/innovation/learning-digital-world/

R Core Team. (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Rasul, T., Nair, S., Kalendra, D., Robin, M., Santini, F., Ladeira, W., Sun, M., Day, I., Rather, A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied learning & Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.29

Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research. *Northwestern University, Evanston, Illinois*. https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rudolph, J., Tan, S., & Aspland, T. (2023). Fully automated luxury communism or Turing trap? Graduate employability in the generative AI age. *Journal of Applied Learning and Teaching, 6*(1), 7–15. https://doi.org/10.37074/jalt.2023.6.1.35

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364–389. https://doi.org/10.37074/jalt.2023.6.1.23

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6). https://doi.org/10.3390/healthcare11060887

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence Scale. *Computers in Human Behavior Reports, 1*, Article 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schön, E.-M., Neumann, M., Hofmann-Stölting, C., Baeza-Yates, R., & Rauschenberger, M. (2023). How are AI assistants changing higher education? *Frontiers in Computer Science, 5*. https://doi.org/10.3389/fcomp.2023.1208550

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology, 307*(2). https://doi.org/10.1148/radiol.230163

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI – Künstliche Intelligenz, 35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *SSRN*. https://doi.org/10.2139/ssrn.4378735

Sparks, J. R., Katz, I. R., & Beile, P. M. (2016). Assessing digital information literacy in higher education: A review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Research Report Series, 2*, 1–33. https://doi.org/10.1002/ets2.12118

Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2209881

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31–40. https://doi.org/10.37074/jalt.2023.6.1.17

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning*. https://www2.ed.gov/documents/ai-report/ai-report.pdf

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning & Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.20

Zlatkin-Troitschanskaia, O., Hartig, J., Goldhammer, F., & Krstev, J. (2021). Students' online information use and learning progress in higher education – A critical literature review. *Studies in Higher Education, 46*(10), 1996–2021. https://doi.org/10.1080/03075079.2021.1953336

APPENDIX A

Table A1. Full list of items

Scale	Items
Global	I feel well informed about ChatGPT. How often do you use ChatGPT in the context of your studies per month? There should be more information about chatbots (such as ChatGPT) in university teaching.
Knowledge	ChatGPT can reliably solve mathematical problems[a]. ChatGPT provides the same answer to the same question[a]. ChatGPT is designed to imitate human speech behavior as closely as possible. ChatGPT performs web searches[a]. ChatGPT may provide content that is not based on facts. ChatGPT retrieves daily updated information[a].
Use	I have used ChatGPT to get feedback on the texts I created. I have used ChatGPT to have content explained to me that was not immediately clear to me in class/lecture. I have used ChatGPT to create an outline for a writing project. I have used ChatGPT to have emails pre-written for me. I have used ChatGPT to have texts summarized. I have used ChatGPT to get an overview of a topic that was new to me. I have used ChatGPT to get content ideas for a paper. I have used ChatGPT to create text modules. I have used ChatGPT to motivate myself for a task (e.g. using ChatGPT to talk about my stress and anxiety). I have used ChatGPT to get advice (e.g. on time management).
Attitude	ChatGPT is reliable. ChatGPT harbors risks when using personal data[a]. ChatGPT reproduces stereotypes and prejudice[a]. The use of ChatGPT leads to problems with copyrights[a]. ChatGPT consumes too much power[a]. Chatbots such as ChatGPT are amongst the most important inventions of the 21st century. ChatGPT is not transparent with regard to the underlying text basis[a]. ChatGPT contributes to the spreading of misinformation[a].

[a]Items were reverse coded

APPENDIX B

Table B1. Descriptive data for knowledge scale

Item	M	SD
1[a]	0.16	0.37
2[a]	0.54	0.50
3	0.61	0.49
4[a]	0.27	0.45
5	0.75	0.44
6[a]	0.42	0.49

[a]Items were reverse coded

Table B2. Descriptive data for use scale

Item	M	SD
1	2.39	1.16
2	2.96	1.19
3	2.50	1.19
4	2.19	1.22
5	2.69	1.18
6	2.96	1.18
7	2.71	1.23
8	2.25	1.22
9	1.41	0.84
10	1.56	0.91

Table B3. Descriptive data for attitude scale

Item	M	SD
1	1.61	0.65
2[a]	0.96	0.78
3[a]	1.42	0.83
4[a]	0.96	0.81
5[a]	1.96	0.84
6	1.87	0.87
7[a]	0.87	0.77
8[a]	1.16	0.77

Items were re-coded onto a scale from 0 to 3.

[a]Items were reverse coded