www.ijte.net

ChatGPT Usage Scale in Education: Validity and Reliability Study

Mustafa Taktak
İstanbul Gelisim University, Turkiye

Görsev Bafralı
İstanbul Gelisim University, Turkiye

To cite this article:

Taktak, M. & Bafrali, G. (2025). ChatGPT usage scale in education: Validity and reliability study. *International Journal of Technology in Education (IJTE), 8*(1), 193-207. https://doi.org/10.46328/ijte.1024

International Journal of Technology in Education
2025, Vol. 8, No. 1, 193-207 https://doi.org/10.46328/ijte.1024

ChatGPT Usage Scale in Education: Validity and Reliability Study

Mustafa Taktak, Görsev Bafralı

Article Info

Article History

Received:
10 September 2024
Accepted:
29 December 2024

Keywords

ChatGPT
Education
Scale development
Reliability
Validity

Abstract

This study aimed to develop a valid and reliable scale to measure individuals' and organizations' attitudes toward the use of ChatGPT, emphasizing the necessity for organizations to adapt to rapidly evolving information and technology environments. The methodology consisted of three stages. In the first stage, a 13-item draft scale was administered to 213 participants, and its construct validity was evaluated through Exploratory Factor Analysis (EFA). In the second stage, based on EFA results, the scale was reduced to 12 items and administered to 175 participants, followed by Confirmatory Factor Analysis (CFA) to validate its structure. In the third stage, test-retest reliability and criterion validity were examined with 67 school administrators and teachers. The analyses revealed that the scale consisted of two dimensions, "opportunities" and "challenges," and CFA confirmed this structure, with fit indices falling within acceptable ranges. The overall reliability coefficient (Cronbach's Alpha) was calculated as 0.71, and further analyses, including independent sample t-tests, item-total and item-residual correlations, inter-factor correlations, and test-retest correlations, supported the scale's validity and reliability. Based on the findings, the scale was recommended to be named the "ChatGPT Usage Scale," offering a valuable tool for assessing attitudes toward ChatGPT use among individuals and organizations.

Introduction

Recent years have seen a rapid development in artificial intelligence (AI), which has led to a variety of applications in various fields, including healthcare (Xu et al., 2021) and education (Zawacki-Richter et al., 2019). With a lot of data, AI systems can be trained to mimic the human brain and carry out repetitive tasks (Bengio et al., 2021). Recent developments in deep machine learning have given rise to platforms that can generate texts that closely resemble human-generated writing due to the impact and efficacy of artificial intelligence (e.g. Rytr, Jasper, ChatGPT and CopyAI). Right now, ChatGPT stands out as the most impressive of the increasing number of content development platforms that incorporate artificial intelligence features.

Its user-friendly interface, which has quickly gained a large number of users, is the primary cause of this (Gleason, 2022). The three things that set ChatGPT apart are its ability to generate text of excellent quality for a variety of uses, its ongoing development, and its free nature. People with varying levels of technical expertise can easily interact and use the platform thanks to its chat format. OpenAI created ChatGPT, a pre-trained language model. The Generative Pre-trained Transformer, or GPT, was developed to efficiently execute speech-based AI tasks, including text generation, dialogue systems, and question answering (OpenAI, 2023).

When completing a particular task, like responding to inquiries or asking questions in a conversational system, ChatGPT provides results that are customized to conversational data. The model is better able to produce responses that resemble a human's because of this process, which enables it to learn particular spoken language nuances and patterns. Apart from text generation, ChatGPT can also comprehend and decipher textual content, which makes it applicable to a broad range of natural language processing applications. Additionally, it is employed in language translation, creative writing, and summarization (OpenAI, 2023). Additionally, ChatGPT is ideally suited to facilitate the use of AI in teaching and learning for educators.

ChatGPT generates responses to user input that are human-like through the use of natural language processing. Its remarkable ability to generate consistent, methodical, and educational answers has drawn attention from all over the world (Zhai, 2022). The University of Minnesota Law School administered four distinct exams to ChatGPT, passing with astonishing success (ChatGPT Passes Exams from Law and Business Schools, 2023). It was determined that the results are sufficient to demonstrate that this artificial intelligence application can earn a university degree, even though the scores are not very good (yet) (Choi et al., 2023). Just two months after its November 30, 2022 launch, ChatGPT reached 100 million active users in January 2023, making it the fastest-growing user app in history (ChatGPT Sets Record for Fastest-Growing User Base—Analyst Note, 2023).

Notwithstanding all of its advantages and demonstrated effectiveness, ChatGPT in the educational sector has drawbacks and risks. The use of ChatGPT by students in their exams and assignments, which they are accountable for, raises questions about artificial intelligence cheating and lack of student effort because of its capacity to offer precise answers to user questions. Some schools appear to have banned ChatGPT access on campus in an effort to stop these negatives (Schools Ban ChatGPT Amid Fears of Artificial Intelligence-Assisted Cheating).

In his analysis of eight ChatGPT articles, Mhlanga (2023) discovered that teachers are apprehensive about using ChatGPT in the classroom. They voiced concern that because ChatGPT can produce appropriate texts quickly, students might outsource their work. Mhlanga (2023) thus underlined the significance of using ChatGPT responsibly and morally. Furthermore, Sallam (2023) examined sixty articles pertaining to ChatGPT from the academic, medical, and healthcare domains. He discovered issues with the studies that ranged from plagiarism to wrong responses and citations. Finally, but just as importantly, Filipino students discussed Chat GPT humanities student from Quezon City stated that he occasionally uses Chat GPT for idea generation or concept exploration, but he cited mistakes as a problem (Antivolla, 2023). In contrast to the qualitative research studies mentioned earlier, there have also been published some quantitative research reports (Fabella, 2023; Farhi et al., 2023; Yılmaz et al., 2023). Fabella (2023) made an effort to learn people's opinions about those above positive and negative aspects of Chat GPT. The researcher designed a nine-item, 6-point Likert scale instrument.

Regarding the benefits that Chat GPT is said to offer, the participants are divided on a few points: the platform gives detailed answers to user inquiries; they are divided on whether Chat GPT uses context from past user-to-platform exchanges that are stored in the same prompt tree; they are divided on whether Chat GPT lets users edit their queries until they are happy with the answer; and they are divided on whether Chat GPT is intended to reject inappropriate requests. Concerning the shortcomings of Chat GPT, however, the respondents strongly agreed that

it can be abused, agreed only somewhat that it can be mistake-prone, and agreed only somewhat that it lacks morality in its responses, can be biased, and sometimes gives convincing but false answers.

Furthermore, Yılmaz et al. (2023) sought to develop and validate an instrument to investigate university students' perceptions of Chat GPT and look at possible differences across gender, grade level, major, and prior experience with using the tool. The study, which employed a quantitative research methodology, involved 239 students enrolled in the Science and Mathematics Education Program at a private university in Almaty, Kazakhstan. The results demonstrated that participants' opinions of Chat GPT were largely positive. The utilization of qualitative research methods in these studies highlights the pressing need to create a measurement tool that can pinpoint the benefits and drawbacks of implementing ChatGPT in the educational domain.

Additionally, Farhi et al. (2023) used scales and measurements from previous research, mostly survey-based empirical literature, to examine how students in the United Arab Emirates (UAE) used ChatGPT as well as their thoughts, concerns, and ethical perceptions. Data was gathered from 388 students at two universities in the city of Al Ain using Yamane's formula. Students see ChatGPT as an innovative tool that helps them in many ways, according to research. The data collected showed that ChatGPT Usage continued to have a significant influence on students' opinions. The path analysis also supported the second hypothesis, which proposed that ChatGPT significantly influenced students' concerns. In the end, the findings validated the final hypothesis by indicating the significant influence of ChatGPT usage on students' perceptions of ethics in the United Arab Emirates.

In the light of the aforementioned studies, it is clear that preexisting studies mainly focused on the perceptions, attitudes and perspectives of the participants regarding using Chat Gpt. However, it is for sure that it is very urgent to work toward creating a reliable and valid scale to gauge how the field of education and stakeholders feel about using ChatGPT in order to highlight how critical it is for education to be able to adjust to the rapidly evolving information and technology landscape quickly. For this reason, this study reports the process and findings of developing a valid and reliable scale about ChatGPT usage scale.

Methodology

Research Design

A measuring tool was created for this study to determine the possible benefits and drawbacks of implementing ChatGPT in the field of education. This measurement tool made use of a survey model. To comprehend, characterize, and explain the characteristics of people in a particular field as well as historical or contemporary circumstances, researchers employ the survey model (Büyüköztürk et al., 2012; Ural & Kılıç, 2013; Kaptan, 1995; Karasar, 2012).

Participants

The study group of the research consists of undergraduate students receiving formal education at public and private universities in Istanbul. In scale development studies, it is recommended to use a sample size of 5 to 10

times the number of items (MacCalum, Widaman, Zhang, & Hong, 1999 as cited in Erkuş, 2014). In this context, 10 times the number of items in the draft scale (13*10=130) was taken into account when determining the sample size. In this context, 213 participants were reached for Exploratory Factor Analysis (EFA), 175 participants for Confirmatory Factor Analysis (CFA) and 67 participants for test-retest application. In the analysis made on the demographic profile of the participants in the EFA study, 60.6% (129 people) of the participants were women; In terms of age distribution, it was seen that 15.9% of the participants were 20 years old and below, 75.6% were 21-25 years old and 8.5% were 26 years old and above. Regarding technology use, it was found that 74.2% of the participants stated that it was good, while 2.1% stated that it was bad. Additionally, it was determined that 69.7% of the participants used ChatGPT mostly on the computer.

Scale Development Steps

A scale development study was conducted to evaluate the effective use of ChatGPT in education. In this study, it was first determined how education stakeholders define and perceive ChatGPT. To identify these perceptions, the relevant literature was scanned, and the findings were transferred to the item pool. Then, the content and scope of the items in this item pool were evaluated by consulting expert opinions. Then, the items in the item pool were examined using statistical techniques such as descriptive analysis, EFA (Exploratory Factor Analysis), inter-factor correlation analysis with the scale total score, and CFA (Confirmatory Factor Analysis).

As a result of an in-depth review of the literature, a draft scale with 21 items was created. To evaluate this draft scale in terms of its form, content, understandability and grammar, opinions were taken from three people who are experts in the fields of measurement-evaluation and educational management. 8 items deemed inappropriate by the experts and 1 item as a result of the analysis were removed from the scale. All items in the draft scale prepared in a five-point Likert type were scored positively. The scale consists of 12 items and 2 sub-dimensions. Participants received scores ranging from 12 to 60 on the scale. The high total score indicates that the opportunities and challenges offered by ChatGPT are at a high level.

Analysis of Data

EFA and CFA methods were used to statistically verify the validity of the scale. The feasibility of EFA was evaluated using the Kaiser-Meyer-Olkin (KMO) test followed by the Bartlett test. To determine the construct validity of the scale, principal component analysis and Varimax orthogonal rotation methods were applied in EFA. In order to determine the relationship between the scale and the items, correlation coefficients between the total score of the scale and the total scores of the factors were calculated.

In CFA, chi-square value and fit indices were examined to test the validity of the model. In order for the model to be valid in CFA, fit indices must be at an acceptable level. Cronbach's Alpha value was calculated to test the reliability of the developed scale. In addition, the necessary correlation coefficients were calculated to prove the relationship between the whole scale and the items. Dependent groups t-test analyses were performed to determine the discrimination of the items. Data analysis was carried out using statistical packages (SPSS and AMOS).

Findings

In this part of the study, the findings obtained from the validity and reliability studies of the developed scale are included.

Findings Regarding Validity Studies

In the research, Exploratory Factor Analysis (EFA) was used to determine the factor structure of the scale. However, before proceeding with EFA, a normality test was used to determine whether the data was normally distributed. The Kolmogrov-Smirnov test result ($p>.05$) was not significant. This indicates that the data shows a normal distribution. Additionally, since the skewness (.697) and kurtosis (1.077) values were within ±1.5, it was accepted that the scores obtained from the scale showed a normal distribution (Tabachnick and Fidell, 2007; George and Mallery, 2016).

In line with the purpose of the research, Kaiser-Meyer-Olkin (KMO) and Bartlett's Test were used to determine whether the data structure obtained from the scale draft was valid for Exploratory Factor Analysis (EFA). According to the analysis results, Kaiser-Meyer-Olkin value = .813 and Bartlett's Test result = 1314.313 ($p <.001$). These findings indicate that the scale is suitable for examining the factor structure and that the measurement has a separable structure among the factors.

In the research, principal component analysis was used to determine the structural integrity of the scale. In the first analysis, two factors with eigenvalues greater than 1 were identified for the scale consisting of 13 items. These factors explain 60.82% of the total variance. During the factor analysis process, it was taken into consideration that the load value of each item should be .30 or above, that the items should belong to a single factor, and that there should be at least a .10 difference between the load values of the items included in more than one factor (Çokluk et al., 2012 ; Tavşancıl, 2002).

Items with loading values (below .30) in the draft scale were identified, and item 13, which also loaded on different factors but had a difference of .10 or less between the loading values, was removed from the scale. Later, after removing this item, the analysis was carried out again, thus obtaining the final version of the scale. As a result of the procedures, it was determined that the scale consisted of two factors. The factor loadings of the scale and the proportion of variance they explain are shown in Table 1.

Table 1. Factor Eigenvalues and Total Variance Values

Factor	Eigenvalue	Variance	Cumulative
1.	4.771	39.756	39.756
2.	2.528	21.066	60.822

When Table 2 is examined, it is seen that the first factor is stronger than the other factor in terms of factor loadings. Similarly, of the 60.82% variance explained in total, 39.75% is explained by the first factor and 21.06 by the

second factor. Additionally, the Scree Plot given in Figure 1 was examined to provide evidence about whether the items would be distributed into factors.

Figure 1. Scree Plot of Pilot Scale

When Table 2 is examined, it is seen that the 12-item structure has two factors. Varimax orthogonal rotation analyses were performed to determine the distribution of items to factors. According to the literature, Varimax orthogonal rotation is preferred if the factor loading of an item is .32 or above, there is at least 10% overlap between factors, and there are no compelling reasons (Tabachnick and Fidell, 2007). The items under the factors and their load values are presented in Table 2.

Table 2. Factor Item Loadings of the Scale

	Factor 1	Factor 2
Item6	0,849	
Item2	0,838	
Item3	0,838	
Item1	0,819	
Item5	0,773	
Item4	0,769	
Item7	0,677	
Item9		0,809
Item10		0,805
Item12		0,678
Item11		0,671
Item8		0,656
Total Variance	37.62	23.19

As seen in Table 2, 12 items remain in the draft scale. While the item load values of the first factor of the scale range from .67 to .84, the second factor varies between .65 and .80. By examining the items under the factors, the first factor of 7 items covering the positive tendencies of the use of ChatGPT in education was named "Opportunities". The second factor, which consists of 5 items and covers the negative aspects of using ChatGPT in education, is called "Difficulties". After the exploratory factor analysis, correlation analysis was performed to determine the relationship between the factors for the construct validity of the scale and the total score. The results of this analysis are given in Table 3.

Table 3. Correlations Coefficients Between Item-Total Score

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Total Score	.623	.597	.589	.559	.523	.630	.649	.355	.440	.436	.219	.400

**p < .01.

When Table 3 is examined, it is observed that item-total score correlations vary between r=.219 and r=.649 and that these values reflect positive and significant relationships. Additionally, the first level confirmatory factor analysis based on the results of Exploratory Factor Analysis is presented in Figure 2.

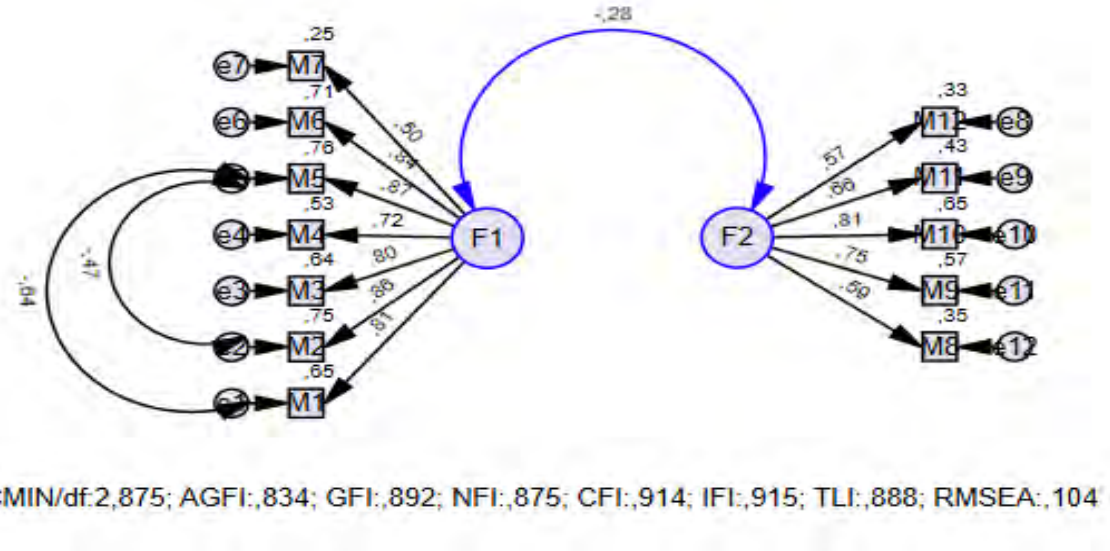

Figure 2. Confirmatory Factor Analysis

Fit indices were used to judge the accuracy of the structure tested by confirmatory factor analysis. Perfect fit with χ^2/df ratio=2.875; Acceptable fit with GFI=.892 value; Acceptable fit with AGFI=.834 value; Excellent fit with CFI=.914 value; It was determined that the fit was excellent with IFI=.068 value and acceptable fit level with TLI=.888 values (Tabachnick and Fidell, 2013; Hair et al., 2014; Schumacker and Lomax, 2010; Kline, 2011).

Findings Related to the Reliability of the Scale

In order to determine the reliability level of the scale, Cronbach Alpha, item-total and item residual correlation, independent groups t-test between the upper 27% and lower 27% groups, test-retest correlation and dependent

groups t-test analyzes were performed. Factor and overall Cronbach-Alpha reliability coefficients of the scale are given in Table 4.

Table 4. Cronbach Alpha Reliability Coefficients

Factor	Cronbach's Alpha
Opportunities	.742
Challenges	.796
Total	.714

In Table 4, the reliability coefficients of the sub-dimensions of the scale are presented. Accordingly, the reliability coefficient for the Opportunities subscale is .74 and for the Challenges subscale is .68. Additionally, the overall reliability coefficient of the scale was calculated as .71. After calculating the Cronbach α reliability coefficient, item-total and item-remainder analyzes were performed to determine whether the items were discriminatory. Table 5 lists the item-total and item-remainder correlation coefficients.

Table 5. Item-Total and Item-Remainder Correlation Analysis Results

Factor	Item	MT	MK
Opportunities	M1	.682	.675
	M2	.687	.679
	M3	.692	.681
	M4	.693	.684
	M5	.701	.691
	M6	.687	.673
	M7	.679	.669
Challenges	M8	.729	.721
	M9	.715	.704
	M10	.721	.708
	M11	.745	.741
	M12	.723	.717

When Table 5 is examined, the item-total correlation coefficients of the scale vary between r=.679 and r=.745, and the remaining item correlation coefficients vary between r=.669 and r=.741. In order to determine the distinctive features of each of the two factors that make up the scale, the total scores received from the scale by the 213 participants in the study group were ranked from lowest to highest. Independent group t-test analysis was performed to determine whether the arithmetic means between the lower and upper 27% groups in the study group were different. The analysis results are presented in Table 6.

As a result of the analysis of Table 6, it was determined that there were statistically significant differences between the lower and upper groups of 27% of the items of the scale (at the $p < .001$ level).

Table 6. Independent t-Test Results for the Difference between the Lower and Upper 27% Groups

Dimension	Item	Group	N	Av	SS	t	Sd	p
Opportunities	M1	Lower 27%	58	3.84	0.679	-13.859	213	.000
		Upper 27%	58	4.84	0.392	-13.859	213	
	M2	Lower 27%	58	3.6	0.849	-13.409	213	.000
		Upper 27%	58	4.78	0.436	-13.409	213	
	M3	Lower 27%	58	3.85	0.735	-13.287	213	.000
		Upper 27%	58	4.88	0.417	-13.287	213	
	M4	Lower 27%	58	3.72	0.727	-14.549	213	.000
		Upper 27%	58	4.83	0.399	-14.549	213	
	M5	Lower 27%	58	3.57	0.722	-12.764	213	.000
		Upper 27%	58	4.65	0.576	-12.764	213	
	M6	Lower 27%	58	3.81	0.584	-14.787	213	.000
		Upper 27%	58	4.81	0.425	-14.787	213	
	M7	Lower 27%	58	3.56	0.768	-12.175	213	.000
		Upper 27%	58	4.59	0.51	-12.175	213	
Challenges	M8	Lower 27%	58	3.73	0.7	-10.603	213	.000
		Upper 27%	58	4.58	0.528	-10.603	213	
	M9	Lower 27%	58	3.79	0.705	-12.689	213	.000
		Upper 27%	58	4.81	0.499	-12.671	213	
	M10	Lower 27%	58	3.67	0.68	-11.383	213	.000
		Upper 27%	58	4.64	0.622	-11.383	213	
	M11	Lower 27%	58	3.47	0.688	-13.671	213	.000
		Upper 27%	58	4.59	0.573	-13.671	213	
	M12	Lower 27%	58	3.14	0.777	-14.047	213	.000
		Upper 27%	58	4.44	0.634	-14.047	213	

In addition, it was observed that these items had a high level of reliability and had distinctive features at the expected level. Another criterion used to evaluate the reliability of the scale is the test-retest method. This method shows that reliability and stability are high if the correlation scores obtained by re-administering the scale to the same participants at a certain time interval are strong (DeVellis, 2017: 51-52; Özdamar, 2016: 85). In this study, the test-retest method was applied to the participants three weeks apart.

In Table 7, the Correlation Coefficients between the scores obtained as a result of the test-retest application are given. When Table 7 was examined, it was determined that the dependent t-test result in the test-retest application of the scale was significant at the $p < .05$ level. Since the p-value obtained in the dependent t-test is greater than 0.5, it can be interpreted that there is no significant difference between the two values (Patton, 2017: 272). These findings show that the data obtained is stable and has sufficient reliability for the field of social sciences.

Table 7. Test-Retest Application Result Correlation Coefficients

Dimension	Item	N	r	p
Opportunities	M1-M1	67	.507	.000
	M2-M2	67	.487	.000
	M3-M3	67	.601	.000
	M4-M4	67	.587	.000
	M5-M5	67	.576	.000
	M6-M6	67	.617	.003
	M7-M7	67	.532	.000
Challenges	M8-M8	67	.384	.000
	M9-M9	67	.548	.000
	M10-M10	67	.560	.000
	M11-M11	67	.409	.000
	M12-M12	67	.513	.001

Discussion and Conclusion

In the field, there have been many other quantitative and qualitative research reports regarding with the aim of investigating perceptions of people about ChatGPT usage (Fabella, 2023; Farhi et al., 2023; Yılmaz et al., 2023). Specifically speaking, Fabella (2023) tried to find out what people thought about the advantages and disadvantages of Chat GPT. A 6-point Likert scale with nine items was created by the researcher. The items were also about Chat GPT's drawbacks such as; it can be abused, it can be prone to mistakes, and it lacks morality in its responses, can be biased, and occasionally provides convincing but inaccurate answers. Despite having a similar objective, the scale developed in this research mostly aims to investigate the probable challenges and opportunities that ChatGPT usage offers with the questions such as; ChatGPT usage offers individual learning opportunity, ChatGPT can reduce student-teacher interaction (see Appendix A for the whole scale in Turkish and Appendix B for the whole scale in English).

Additionally, it is evident that previous research primarily concentrated on participant attitudes, perceptions, and viewpoints regarding ChatGPT use (Fabella, 2023; Farhi et al., 2023; Yılmaz et al., 2023). To emphasize how important it is for education institutions to be able to swiftly adapt to the quickly changing information and technology landscape, it is imperative that efforts be made to develop a valid and trustworthy scale to measure how stakeholders feel about using ChatGPT.

Given that it focuses specifically on feelings regarding ChatGPT's usage across a range of fields from the perspective of university students with diverse majors, this research study is very important. Although previous research has looked at ChatGPT's various facets, such as its language model and ethical issues, attitudes, and perceptions, this study fills in a critical gap by focusing on how users feel about the advantages and disadvantages of using the platform by developing a valid and reliable scale for this specific purpose. In accordance with the

scale development process recommended by Wang (2003), Hinkin (1998), and Churchill (1979) (including steps such as item generation, questionnaire administration, initial item reduction, CFA, convergent/discriminant validity, and replication), this study conceived a conceptual definition of the ChatGPT Usage construct, generated an operative plan of an initial item pool, and empirically verified the generic scale development. Item purification and factor extraction were conducted in the EFA stage, and the factor structure of the scale was confirmed in the CFA stage. The analysis results indicate that the scale consists of two factors: opportunities and challenges in using ChatGPT. Opportunities refers to the extent to which people view ChatGPT as an AI product or technology as beneficial or worthwhile for their tasks. Challenges, on the other hand, refers to the extent to which ChatGPT causes drawbacks in their assignments or tasks.

The methodology for the study was three-staged. In the first phase, 213 participants completed a 13-item draft scale, and the construct validity of the scale was assessed using exploratory factor analysis (EFA). In the second phase, the 12-item scale was administered to 175 participants based on the EFA results, and Confirmatory Factor Analysis (CFA) was used to confirm the construct validity of the scale. The test-retest method and criterion validity were examined in 67 school administrators and teachers during the third stage of the study. The two dimensions of the scale's structure—referred to as "opportunities" and "challenges"—were found to exist. CFA gave its approval to the suggested two-dimensional structure, and it was noted that the fit indices were within reasonable bounds. The Cronbach Alpha calculation yielded an overall reliability coefficient of 0.71 for the scale. To further assess the scale's reliability, the independent groups t-test, item residual and total correlation analysis, inter-factor correlation analysis, and test-retest correlation analysis were carried out. The scale's validity and reliability were confirmed by these analyses, which also recommended renaming it the "ChatGPT Usage Scale".

The content validity and reliability of the 12-item scale are adequate. Additionally, the convergent validity, discriminant validity, and criterion-related validity of the scale were confirmed. As a result, the results imply that this scale has established psychometric qualities and offers numerous significant practical implications for AI environments. This scale can be used to measure people's feelings of ChatGPT usage as a result of its satisfactory validity and reliability. In spite of this, comparing a person's level of ChatGPT usage to a norm - that is, the total distribution of that person's application degrees as determined by others - is a more accurate way to evaluate them. The scale is appropriate for developing preliminary relevant criteria because of the variety of samples used in the study.

This study has various applications in the real world. The scale can be used by practitioners to compare an individual's ChatGPT Usage across various dimensions, in addition to conducting an overall assessment. The 12-item test offers a general framework for comparative analysis and is made to be compliant with ChatGPT. This instrument can be adopted or modified as needed to fit the requirements of a particular study or setting.

References

Antivola MHL. (2023) *What Filipino students are saying about CHATGPT* [Internet]. Available from: https://www.bworldonline.com/technology/2023/04/19/517952/what-filipino-students-are-sayingabout-

ChatGPT/

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*(7), 58-65.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2012). *Bilimsel araştırma yöntemleri* (12. baskı). Ankara: Pegem A.

Choi, Y., Jung, C. M., Kang, G., Kim, P. H., Chae, L., & de Viveiros, P. H. (2023). 155 Evaluation of in silico tools and chat-GPT to determine the impact of missense variants of immune-related genes associated with immunotherapy outcomes for solid tumors.

Çokluk, Ö., Şekerçioğlu, G. ve Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve Lisrel uygulamaları*. Ankara: Pegem Akademi.

DeVellis, R. F. (2017). *Ölçek Geliştirme: Kuram ve Uygulamalar*. Ankara: Nobel.

Erkuş, A. (2014). *Psikolojide ölçme ve ölçek geliştirme I: Temel kavramlar ve işlemler*. Ankara: Pegem A.

Fabella, F. E. T. (2023). Attitudes toward the positive and negative features of chat gpt by selected first year college students. *International Research Journal of Modernization in Engineering Technology and Science.*

Farhi, F., Jeljeli, R., Aburezeq, I., Dweikat, F. F., Al-shami, S. A., & Slamene, R. (2023). Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Computers and Education: Artificial Intelligence,* 100180.

Gleason, N. (2022). ChatGPT and the rise of AI writers: How should higher education respond? THE Campus Learn, Share, Connect. Retrieved April, 2024 from https://www.timeshighereducation.com/campus/ChatGPT-and-rise-ai-writers-how-should-higher-education-respond

Hair, Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis*. London: Pearson Education Limited.

Kaptan, S. (1995). *Bilimsel araştırma ve istatistik teknikleri*. Ankara: Rehber.

Karasar, N. (2014). *Bilimsel araştırma yöntemi.* Ankara: Nobel.

Kline, R. B. (2016). *Yapısal Eşitlik Modellemesi ve Uygulaması*, (Ed. S. Şen). Ankara: Nobel.

Mhlanga, D. (2023). *Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning*. In FinTech and Artificial Intelligence for Sustainable Development: The Role of Smart Technologies in Achieving Development Goals (pp. 387-409). Cham: Springer Nature Switzerland.

Özdamar, K. (2016). *Eğitim, Sağlık ve Davranış Bilimlerinde Ölçek ve Test Geliştirme Yapısal Eşitlik Modellemesi.* Eskişehir: Nisan Kitabevi Yayınları.

Patton, MQ. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory And Practice*. (4th ed.). USA: Sage Publications.

Sallam, M. (2023, March). *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns*. In Healthcare 11,(6), p. 887. MDPI.

Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling* (3rd ed.). New York, NY, US: Routledge/Taylor & Francis Group.

Tabachnick, B. G., ve Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn and Bacon. Boston. USA.

Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel.

Ural, A. ve Kılıç, İ. (2013). *Bilimsel araştırma süreci ve SPSS ile veri analizi*. Ankara: Detay

Xu, L., Sanders, L., Li, K., & Chow, J. C. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer, 7*(4), e27850.

Yilmaz, H., Maxutov, S., Baitekov, A., & Balta, N. (2023). Student attitudes towards chat GPT: A technology acceptance Model survey. *International Educational Review, 1*(1), 57-83.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1-27.

Zhai, X. (2023). *ChatGPT and AI: The game changer for education*. Available at SSRN.

Author Information

Mustafa Taktak
https://orcid.org/0000-0003-2784-1574
İstanbul Gelişim University
Cihangir, Avcılar
Türkiye
Contact e-mail: *mtaktak@gelisim.edu.tr*

Görsev Bafralı
https://orcid.org/0000-0001-6726-3452
İstanbul Gelişim University
Cihangir, Avcılar
Türkiye

Appendix A. ChatGPT Usage Scale in Education (Turkish Version)

	Eğitimde ChatGPT Kullanım Ölçeği	Hiç Katılmıyorum	Katılmıyorum	Kısmen Katılıyorum	Katılıyorum	Kesinlikle Katılıyorum
1.	ChatGPT, bireysel öğrenme imkânı sağlar.					
2.	ChatGPT, öğrenme sürecinde hızlı geri bildirim verir.					
3.	ChatGPT, akademik yazım sürecinde destek sağlar.					
4.	ChatGPT, ilham verici fırsatları sunar.					
5.	ChatGPT, zaman ve maliyet tasarrufu sağlar.					
6.	ChatGPT, planlama süreçlerini kolaylaştırır.					
7.	ChatGPT, dil çevirisinde iyi bir yardımcıdır.					
8.	ChatGPT, öğrenci-öğretmen etkileşimini azaltabilir.					
9.	ChatGPT etik ihlalleri meşrulaştıran bir uygulamadır.					
10.	ChatGPT, teknolojiye bağımlılığı artırarak üretkenliği azaltır					
11.	ChatGPT, aşırı ve gereksiz içerik sunmaktadır.					
12.	ChatGPT'nin denetlenememesi yasal sorunlara yol açabilir.					

Appendix B. ChatGPT Usage Scale in Education (English Version)

ChatGPT Usage Scale in Education	Totally Disagree	Disagree	Partially Agree	Agree	Strongly Agree
1. ChatGPT provides individual learning opportunities.					
2. ChatGPT provides rapid feedback during the learning process.					
3. ChatGPT provides support during the academic writing process.					
4. ChatGPT offers inspiring opportunities.					
5. ChatGPT saves time and money.					
6. ChatGPT facilitates planning processes.					
7. ChatGPT is a good assistant in language translation.					
8. ChatGPT can reduce student-teacher interaction.					
9. ChatGPT is an application that legitimizes ethical violations.					
10. ChatGPT reduces productivity by increasing dependency on technology.					
11. ChatGPT offers excessive and unnecessary content.					
12. Failure to monitor ChatGPT can lead to legal problems.					