4th International Conference on Scientific and Academic Research

July 19-20, 2024 : Konya, Turkey

https://www.icsarconf.com/

Generative Artificial Intelligence (GenAI) Literacy Scale: Validity and Reliability

Sahin Gokcearslan[1], Hatice Yildiz Durak [2] and Mustafa Serkan Gunbatar [3], Nilufer Atman Uslu[4], Aysun Nuket Elci [5]

[1]*Distance Education R&D Center, Gazi University, Turkiye*
[2] *Department of Computer Education and Instructional Technology, Necmettin Erbakan University, Turkiye*
[3] *Department of Computer Education and Instructional Technology, Van Yuzuncuyil University, Turkiye*
[4, 5] *Department of Primary Education, Elementary Education, Dokuz Eylül University, Turkiye*

[*]*(sgokcearslan@gazi.edu.tr) Email of the corresponding author*

Abstract **–** Generative Artificial Intelligence (GenAI) chatbots can make positive contributions to education processes thanks to their advanced natural language processing capabilities. This is a new technology, and it is important to know its proper use. GenAI literature students have more potential to use chatbots effectively. In the literature, there are a limited number of measurement tools to measure the literacy level of AI chatbots, which are in the productive artificial intelligence category. The aim of this study is to adapt an instrument to measure the GenAI literacy of university students. The AI literacy scale developed by (Wang et al., 2023) was adapted into Turkish with the participation of 297 undergraduate and graduate university students. The scale consists of four dimensions and ten items. The Cronbach's alpha internal consistency coefficient for the four dimensions was calculated to be 0.74. The total test scores showed a significant difference between the upper and lower 27% groups. The fit indices were examined to demonstrate the adequacy of the model, and it provides a good foundation to build upon. With this structure, the instrument can be used as a suitable tool to measure the level of GenAI literacy, the effectiveness of learning activities to improve this level, and to conduct research on this topic.

Keywords – Artificial intelligence (AI) chatbots, GenAI literacy, measurement tool, university students

I. INTRODUCTION

We can think of artificial intelligence (AI) as a field of research that focuses on producing intelligent machines and intelligent computer programs (McCarthy, 2007). Today, AI applications developed for many different purposes are used. Chatbots are also a type of AI applications. ChatGPT, a chatbot, works with Generative Artificial Intelligence (GAI) logic, collects data with its own algorithm, performs operations on this training data and provides information as a result of interaction with people (Harrer, 2023). The text-based information it provides is as if it were created by a human (Mohamed, 2023); (Tirado-Olivares et al., 2023) and uses statistical correlations to choose the next word when constructing sentences (Harrer, 2023). ChatGPT, a good representative of chatbots, has impressed its users with its performance since it became available for free use in November 2022 (King & ChatGPT., 2023).

Traditionally, when looking at the working logic of computer technology, computers operate and produce results according to the commands given by people. When we perform this search process through chatbots, we assign the task of selecting and synthesizing the resources to it. However, the ability of chatbots used to produce text-based content to select valid and reliable sources is unclear. Because chatbots cannot distinguish between right and wrong (Sison et al., 2023), they can present information that is biased, tends to maintain existing power dynamics, and has the potential to manipulate people (Weinberg, 2022). A chatbot can return very good, completely wrong, or even irrelevant results to the given prompts. The text it produces

may contain errors. Therefore, text-based content produced by a chatbot should be checked by an expert for accuracy, bias, appropriateness, and logic (Hosseini et al., 2023).

II. Literature Review

Generative AI Literacy

Chatbots are tools that can communicate with people via text or voice by combining technologies such as artificial intelligence and language processing (Pérez et al., 2020); (Yildiz Durak, 2023). In addition to having the potential to improve workplaces and daily life, chatbots bring many opportunities as communication and information tools for digital learning (Wollny et al., 2021); Yildiz Durak & Onan., 2023a); Yildiz Durak & Onan., 2023b). Chatbots play an important role as a tool that emphasizes individualized teaching and supports accessibility (Pérez et al., 2020). A systematic review study reports that research on the use of chatbots in education generally focuses on language, engineering, and computer courses (Hwang & Chang., 2023); (Yildiz Durak & Onan., 2023c). Chatbots might offer significant support in helping students and reinforcing repetitive tasks (Pérez et al., 2020); Yildiz Durak & Onan., 2023c). However, in cases where there are no repetitive processes and require in-depth and instant decisions, chatbots inadequacy can be mentioned. For example, it is reported that there are limitations in ensuring that chatbots are matched with learning strategies in the design of e-activities (Gökçearslan., 2013). Therefore, it is suggested that more comprehensive and student-focused studies are needed to evaluate their effectiveness in education (Hwang & Chang., 2023). At this point, students' artificial intelligence literacy and their awareness of it are important (Gökçearslan et al., 2023; Esiyok et al., 2024).

AI knowledge encompasses the proficiency to assess AI technologies critically, engage in effective communication and collaboration with AI systems, and utilize AI as a versatile tool across online platforms, households, and professional environments (Long & Magerko, 2020). It is an important starting point for learners to have AI literacy, to understand the underlying mechanisms of these technologies and thus to gain awareness about their capabilities and limitations, and to develop a critical attitude towards the information produced (Aktaş et al., 2024).

The Present Study

Various conceptualizations and frameworks have been proposed in the literature regarding AI literacy. Recently, there have been studies conducted by researchers on how AI literacy can be measured for different education levels. There are studies focusing on the scope of AI literacy or how it can be improved for early childhood education (Su et al., 2023), secondary school (Lee et al., 2021) or K-16 (Wang et al., 2023) Although it is important to prepare students in AI literature from an early age, GenAI has very quickly entered the business world and has begun to be seen as an innovation that has the potential to transform and change expectations from the workforce. In this respect, it is necessary to first measure the GenAI literacy levels of university students so that their current situation can be determined. The concept of AI literacy is applicable to intelligent agents, automata, problem-solving through data structures, sorting, searching, and machine learning, with the exception of the latter in terms of measuring AI literacy in students within the computer science education system. The instrument that refers to digital literacy to measure the AI literacy of users in the general population is based on the emotions and behaviours that users exhibit in relation to knowledge and skills, along with values, attitudes and ethics. The term "AI literacy" is defined as "awareness and understanding of AI technology in practical applications; the ability to apply and use AI technology to perform tasks competently; and the ability to analyse, select, and critically evaluate data and information provided by AI, while promoting awareness of one's own personal responsibilities and respect for mutual rights and obligations." Four constructs have been proposed for AI literacy: awareness, use, evaluation, and ethics. These dimensions cover the general skills of the general population related to AI. While the measurement tool is adapted to productive AI, which is a type of artificial intelligence, there is no adaptation that would disrupt this scope. In this respect, the adapted instrument was used for measuring general productive AI literacy in the general population

In this way, educators can have the necessary starting data to prepare them to adapt to their professional lives. Based on this point, the current study aims to adapt AI literacy survey developed by (Wang et al., 2023) to Turkish culture and obtain the necessary findings regarding its validity and reliability.

MATERIALS AND METHOD

A. *Study Group*

In this study, 66.7% of the 297 participants were female and 33.3% were male. Their mean age was 21.75 years. 85.9% were associate and undergraduate students, 14.1% were graduate students. This demographic distribution provides a diverse representation of young adults in higher education, primarily focusing on those in early stages of their academic careers.

B. *Data Collection Instrument*

The measurement tool developed by (Wang et al., 2023) to measure AI literacy was adapted to generative AI. A 12- item instrument was obtained to measure AI literacy. The scale items have a 5-point Likert scale. A high score on the instrument indicates a high level of AI literacy. 12 items provide a general measure of generative AI literacy. These general statements were subjected to confirmatory factor analysis with GenAI adaptation.

C. *Data Collection*

Data collection was conducted through an online Google Form, which allowed for an efficient and widespread distribution among the target audience. "Voluntary status of the participants" was accepted as the main criterion for data collection.

D. *Data Analyses*

The collected data were then analyzed using Lisrel 8.0 and IBM SPSS 24. These statistical software tools enabled detailed examination of the data, including confirmatory factor analysis and other relevant statistical tests. The use of Lisrel 8.0 enabled an understanding of whether the AI literacy instrument adapted by confirmatory factor analysis was both reliable and valid for measuring AI literacy in productive AI contexts. IBM SPSS 24 was used for additional validity and reliability analyses.

RESULTS

Descriptive Findings

Table 1 presents the mean, standard deviation, skewness, and kurtosis statistics and standard error values of the items in this scale.

Table 1 Mean, standard deviation, kurtosis and skewness

Items	Mean	Standard Deviation	Skewness		Kurtosis	
			Statistic	Std. Error	Statistic	Std. Error
L1	5.38	1.208	-0.202	0.141	-0.981	0.282
L3	4.39	1.321	0.233	0.141	-0.461	0.282
L4	4.77	1.333	-0.063	0.141	-0.588	0.282
L6	5.46	1.205	-0.325	0.141	-0.844	0.282
L7	5.25	1.226	-0.235	0.141	-0.844	0.282
L8	5.32	1.163	-0.185	0.141	-0.778	0.282
L9	5.05	1.155	0.040	0.141	-0.730	0.282
L10	4.96	1.538	-0.392	0.141	-0.603	0.282
L11	2.40	1.528	1.064	0.141	0.384	0.282
L12	4.96	1.505	-0.176	0.141	-1.037	0.282

Table 1 indicates that the mean scores of the items range from 5.47 to 2.40, with standard deviations varying from 1.733 to 1.155. The skewness and kurtosis values calculated for each item fall within the range of +1.5 and -1.5. Upon analysis of the normal distribution curve of the items, it was concluded that the scores were normally distributed.

Confirmatory Factor Analysis

In order to adapt the scale to Turkish, confirmatory factor analysis was applied. The fit indices of the confirmatory factor analysis were as follows: $\chi 2(29)=60.85$, $\chi 2/df=2.063$, RMSEA= 0.061, GFI= 0.96, NFI= 0.96, NNFI=0.97, CFI=0.98, IFI=0.98. The fit indices were examined to demonstrate the adequacy of the model, and it was found that they demonstrated acceptable and excellent fit values. Items 2 and 5, which had estimated factor loadings of 0.2 and 0.3, were not removed from the scale. The presented analyses are the values obtained after item removal.

Confirmatory factor analysis was employed to assess the factorial validity of the model comprising four factors and 12 items. Dimensions are: FF1: Awareness, FF2: Usage, FF3: Evaluation, FF4: Ethics. The resulting model is presented in Figure 1.

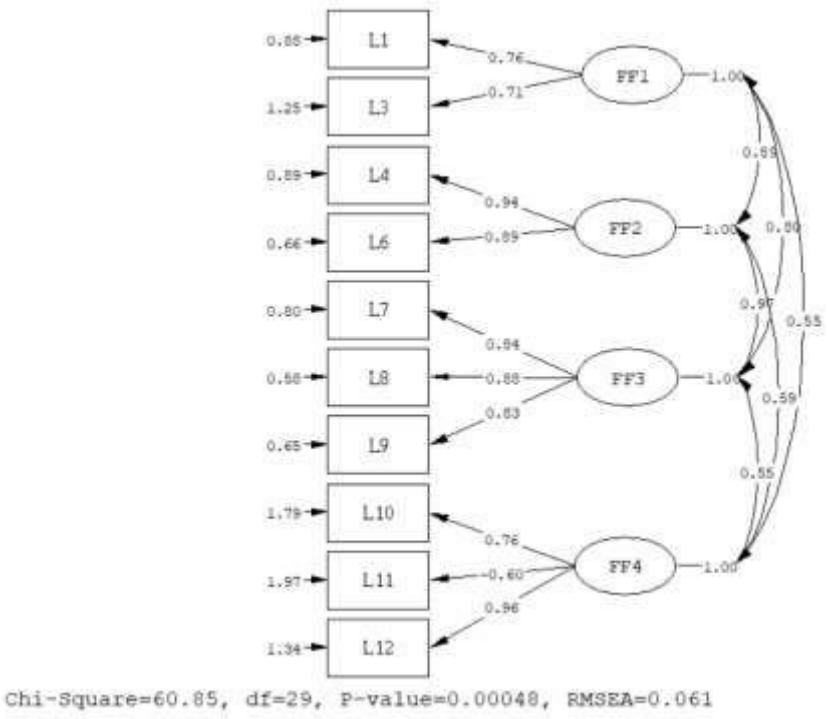

Figure 1: CFA Model

Factor loadings are between 0.94 and 0.60. According to the t-test findings, all connections are statistically significant. It can be said that factorial validity is achieved.

Internal Consistency Analysis

The Cronbach α internal consistency coefficient of this scale was calculated as 0.739. The Cronbach α internal consistency coefficients for the sub-factors of the scale were 0.60, 0.682, 0.761, 0.60, respectively.

Item-Total Scores Correlation and Item Discrimination Index

Item-total score correlations were calculated on the relationship between the score for each item and the total score obtained from the scale. It was found that the item-total score correlation values ranged between 0.300 and 0.656 and had a significant relationship.

In addition, it was examined whether there was a statistically significant difference between the scores of the group in the upper 27% group (N=80) and the group in the lower 27% group according to the total score obtained from the scale. According to the total test scores, a significant difference was observed between the upper and lower 27% groups (t=26.301, p=0.000).

CONCLUSION AND RECOMMENDATIONS

The results of confirmatory factor analysis for the GenAI literacy scale (Appendix 1), adapted into Turkish by integrating with the GenAI structure, indicated that the fit indices were acceptable and had an excellent fit. The scale comprises four dimensions and ten items. The Cronbach α internal consistency coefficient for the four dimensions was calculated as 0.739. The Cronbach α internal consistency coefficients for the sub-factors of the scale were 0.60, 0.682, 0.761, and 0.60, respectively. These values of 0.60 and above provide evidence of reliability, as stated by Dijkstra and Henseler (2015). The low

number of items in the dimensions may explain the Cronbach alpha level. The item total score correlation values were found to range between 0.300 and 0.656, with a significant relationship. According to the test total scores, a significant difference was observed between the upper and lower 27% groups (t=26.301, p=0.000). This structure enables the instrument to be used as a suitable tool for measuring the level of GenAI literacy, evaluating the effectiveness of learning activities to improve this level, and conducting research on this topic.

This study has some limitations. A larger number of participants can be reached. Hair et al. (Hair et al., 2010) argued that the number of items should be at least 5 times. Nunnally (Nunnallyet, 1994) argues that at least 10 times the number of participants should be reached. Although more than 10 times the number of participants was reached in this study, the number of participants can be expanded. The instrument is limited to the adaptation of general AI literacy to generative AI literacy. Recently, GenAI tools have been rapidly integrated into social systems, which may lead to the perception of artificial intelligence being associated only with large language models. On the other hand, artificial intelligence works with many disciplines and has many application areas, such as modeling and simulation. In addition, rapid developments in GenAI tools, expansion of the capacity, and the increase in the diversity of the problems they solve may lead to the need to update the adapted instrument, perhaps more quickly than another measurement tool. Another limitation is that countries still need more institutional and government work on ethical concerns in productive artificial intelligence tools. Although this is not a limitation of the research, it should be kept in mind that awareness of the violation of ethical use of productive artificial intelligence tools is in its early stages, not only at the individual but also at the system level.

Another point is related to the fact that the data collection process is self-reported. However, the current study shows that the measurement tool has achieved sufficient statistical findings regarding the construct validity and reliability. It should be considered that errors may arise due to social desirability. Therefore, in the future, it is recommended that task- and performance-based tools be developed that consider users' actual usage.

REFERENCES

Akkaş, Ö. M., Tosun, C., & Gökçearslan, Ş. (2024). Artificial Intelligence (AI) and Cheating: The Concept of Generative Artificial Intelligence (GenAI). In *Transforming Education With Generative AI: Prompt Engineering and Synthetic Content Creation* (pp. 182-199). IGI Global.

Esiyok, E., Gokcearslan, S., & Kucukergin, K. G. (2024). Acceptance of educational use of AI chatbots in the context of self-directed learning with technology and ICT self-efficacy of undergraduate students. *International Journal of Human–Computer Interaction*, 1-10.

Gökçearslan, Ş. (2013). Developing a Scale for the Sense of Community in Online Learning: A Validity and Reliability Study. *Turkish Librarianship*, 27(1), 154-165.

Gökçearslan, Ş., Tosun, C., & Erdemir, Z. G. (2024). Benefits, challenges, and methods of artificial intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education*, 7(1), 19-39. https://doi.org/10.46328/ijte.600

Hair, J. F., Black, W. C., Tatham, R. L., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*, 90, 104512. https://doi.org/10.1016/j.ebiom.2023.104512

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research*, 1-9. https://doi.org/10.1080/08989621.2023.2168535

Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099-4112.

King, M. R., & chatGPT. (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*, 16(1), 1–2. https://doi.org/10.1007/s12195-022-00754-8

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021, March). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM technical symposium on computer science education* (pp. 191-197).

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).

McCarthy, J. (2007). What is artificial intelligence? [PDF]. Retrieved May 1, 2024, from http://cse.unl.edu/~choueiry/S09-476-876/Documents/whatisai.pdf

Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL Faculty Members. *Education and Information Technologies*, 1-23. https://doi.org/10.1007/s10639-023-11917-

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549-1565.

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). ChatGPT: More than a weapon of mass deception, ethical challenges and responses from the human-Centered artificial intelligence (HCAI) perspective. arXiv preprint arXiv:2304.11215.

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4, 100124.

Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P., & Cózar-Gutiérrez, R. (2023). From human to machine: Investigating the effectiveness of the conversational AI ChatGPT in historical thinking. *Education Sciences*, 13(8), 803. https://doi.org/10.3390/educsci13080803

Wang, B., Rau, P. L. P., & Yuan, T. (2023). 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

Weinberg, L. (2022). Rethinking fairness: An interdisciplinary survey of critiques of hegemonic ML fairness approaches. *Journal of Artificial Intelligence Research*, 74, 75-109. https://doi.org/10.1613/jair.1.13196

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? - A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924.

Yildiz Durak, H. (2023). Conversational agent-based guidance: Examining the effect of chatbot usage frequency and satisfaction on visual design self-efficacy, engagement, satisfaction, and learner autonomy. *Education and Information Technologies*, 28(1), 471-478.

Yildiz Durak, H., & Onan, A. (2023). Adaptation of the behavioral intention scale for using and learning chatbots in education into Turkish. *Journal of Ahmet Keleşoğlu Faculty of Education*, 5(3), 1162-1172. https://doi.org/10.38151/akef.2023.104

Yildiz Durak, H., & Onan, A. (2023a, July). Adaptation of Chatbot Confirmation and Usage Continuance Scale into Turkish. In *1st International Conference on Modern and Advanced Research*. https://doi.org/10.59287/icmar.1259

Yildiz Durak, H., & Onan, A. (2023b, July). An examination of studies on the use of chatbot technology in the field of education. In *International Conference on Applied Engineering and Natural Sciences*, 1(1), 121-124. https://doi.org/10.59287/icaens.978

Yildiz Durak, H., & Onan, A. (2023c, July). Turkish adaptation of the chatbot system, information and service quality scale. In *International Conference on Applied Engineering and Natural Sciences*, 1(1), 114-117. https://doi.org/10.59287/icaens.976

Appendix 1

Generative Artificial Intelligence Literacy Scale for University Students

Üretken Yapay Zeka Okuryazarlığı Ölçeği Maddeleri Turkish	GenAI Literacy Scale Items English
1. Akıllı (yapay zekayı kullanan) ve akıllı olmayan araçlar arasındaki farkları ayırt edebilirim.	1. I can distinguish between intelligent vehicles (using AI) and non-intelligent vehicles.
2. Üretken yapay zekanın bana nasıl yardımcı olabileceğini bilmiyorum (ölçekten çıkarılmış madde).	2. I do not know how generative AI can help me (removed item).
3. Üretken yapay zeka uygulama ve ürünlerinde kullanılan teknolojiyi tespit edebilirim.	3. I can identify the technology used in generative AI applications and products.
4. Günlük işlerime yardımcı olması için üretken yapay zeka uygulamalarını veya ürünlerini ustalıkla kullanabilirim.	4. I can skillfully use productive AI applications or products to help me in my daily work.
5. Yeni bir üretken yapay zeka uygulamasını veya ürününü kullanmayı öğrenmek genellikle zordur. (ölçekten çıkarılmış madde).	5. Learning to use a new productive AI application or product is often difficult (removed item).
6. İş verimliliğini artırmak için üretken yapay zeka uygulamalarını veya ürünlerini kullanabilirim.	6. I can use productive AI apps or products to improve work efficiency.
7. Bir süre kullandıktan sonra bir üretken yapay zeka uygulamasının veya ürününün yeteneklerini ve sınırlamalarını değerlendirebilirim.	7. I can evaluate the capabilities and limitations of a productive AI application or product after using it for a while.
8. Üretken yapay zeka tarafından sağlanan çeşitli çözümler arasından uygun bir çözüm seçebilirim.	8. I can choose an appropriate solution from a variety of solutions provided by generative AI.
9. Farklı görevler için çeşitli üretken yapay zeka uygulamaları arasından en uygun ürünü seçebilirim.	9. I can choose the most suitable product from a variety of generative AI applications for different tasks.
10. Üretken yapay zeka uygulamalarını kullanırken her zaman etik ilkelere uyarım.	10. I always follow ethical principles when using generative AI applications.
11. Üretken yapay zeka uygulamalarını veya ürünlerini kullanırken gizlilik ve bilgi güvenliği konularına hiç dikkat etmem.	11. I never pay attention to privacy and information security issues when using generative AI applications or products.
12. Üretken yapay zeka teknolojisinin kötüye kullanımına karşı her zaman dikkatliyim.	12. I am always alert to the misuse of generative AI technology.

*FF1-Awareness:1,3; FF2-Usage: 4,6; FF3- Evaluation: 7,8, 9; FF4- Ethics: 10,11,12