Proceedings of the

16th International Conference on

Technology in Mathematics Teaching

ICTMT 16

ICTMT in the Era of Societal and Digital Transformations

Department of Secondary Education, School of Philosophy,
National & Kapodistrian University of Athens
Athens, Greece
June 7–10, 2023
Editors
Dimitris Diamantidis
Myrto Karavakou
Marianthi Grizioti
Chronis Kynigos

Table of Contents

SOLVING A PROBLEM WITH AI: CONSIDERATIONS ON ASSESSING LEARNING MATHEMATICS

Camilla Spagnolo[1], Giorgio Bolondi[1], Alessandro Gambini[2]

[1]*Free University of Bozen/Bolzano; Camilla.Spagnolo@unibz.it ,Giorgio.Bolondi@unibz.it*

[2]*La Sapienza, University of Rome; alessandro.gambini@uniroma1.it*

The application of artificial intelligence (AI) in STEM education and particularly in mathematics, as an emerging field, is confronted with a challenge of integrating diverse AI techniques and complex educational elements to meet instructional and learning needs. It is well known that assessment is a key point in educational ethics. In addition, central elements of assessment are fairness and validity. On one hand, unfair use of artificial intelligence (AI) by students could invalidate the assessment. On the other hand, teachers' behaviour when faced with AI helps us to construct their identity. It is urgent to investigate how students' use of AI could invalidate the assessment. We discuss a possible use of ChatGPT for investigating whether teachers recognize the use of AI.

Keywords: artificial intelligence, ChatGPT, assessment, mathematics education, computer science.

INTRODUCTION AND THEORETICAL BACKGROUND

Since January 2023, ChatGPT (accessible at the URL https://chat.openai.com/chat) has entered massively into the communication of major newspapers (Aydın & Karaarslan, 2022; Qadir, 2023), even though the chatbot in question (in the first beta version) was released by OpenAI manufacturing company on November 30, 2022. After remarks by Microsoft CEO Nadella in Davos, it is known that ChatGPT will be integrated into many Microsoft products, primarily Bing, the company's flagship browser- and Bing's possible use in education is emphatized in Microsoft communication. The impact undoubtedly of the new tool has been strong, the final product is awaited.

The impact will involve all sectors of our society, but for now the most direct seems to be the training and cheating effect. ChatGPT, at the moment, is not certified and does not cite sources where it gets its information (Yu, 2023). A central and strategic question arises: what will be the impact on educational systems, specifically in the teaching of science disciplines and mathematics in particular? Can teaching not take into account the new perspectives and challenges that AI poses, is one prepared for this "challenge"? Our focus will be on a particular but key aspect of educational processes, i.e. *classroom assessment.* Ethical issues are involved in classroom assessment at several levels (Green et al., 2007); The EU recently drafted an interesting document "Ethical guidelines for educators on the use of artificial intelligence (AI) and data in teaching and learning" (European Commission, 2022).

In the past, artificial intelligent projects like "Eliza" (Weizenbaum, 1966), "Student" or "GPS" (Russel, 2010) attempted to produce agents which can solve general mathematical problems. One soon realized that this goal was too ambitious, especially when "problems" means "open ended tasks" and in particular "argumentative tasks". Not only because of technical hurdles, but also because one had to realize that many problems are computationally hard and cannot be settled in reasonable time. While this slowed the momentum of AI developments in education in the last decade, it did not stop it, and the emergence of new tools is likely to start a new era.

Teachers' identity is looked at as a system of beliefs (Ferretti et al., 2021) and these beliefs affect classroom practices and values. Mathematics education research has traditionally looked at a

teacher's identity as a system of beliefs. Teachers' beliefs have become a focus of research in mathematics education on the agreed assumption that beliefs affect school practices and behaviors. The idea of system thinking is based on the assumption that the concept of belief can be better understood in the relation to other constructs, such as attitude, affect and values, where the different parts are not separable from each other and are situated in a specific context within any person or group.

The issue of defining beliefs has been rephrased, and beliefs are now seen as part of a larger set of constructs regarding teacher identity and classroom interaction, where teachers are "sensible systems that act in a coherent way" (Hannula et al., 2016). For the purpose of our study, we would like to outline a specific and operational definition of teachers' beliefs and convictions, which embeds their identity and thereby scrutinizes the interplay between identity and assessment.

An unavoidable starting point when we want to identify beliefs and convictions related to mathematics teaching and learning is the notion of epistemology, as it has been discussed by Brousseau (Brousseau, 2006). We therefore need to cast teachers' beliefs and convictions in epistemology and the related notions of cultural knowledge and personal knowledge. An epistemological conception is a set of convictions and pieces of knowledge that tell us what the knowledge of individuals and groups of individuals is: its functioning, its forms of validation, its acquisition and its influence on teaching and learning. Brousseau introduces the concept of school epistemology to outline the combination of convictions, beliefs that circulate in the school system regarding the nature of knowledge, methodology, assessment, teaching and learning and the social and political role of mathematics. The noosphere is the joining link between the school system (the teachers' choices included) and the broader social environment it belongs to (Godino, 1993).

The purpose of this paper is to describe an experiment whose goal is to bring out teachers' beliefs, and attitudes (McLeod, 1992) in an assessment situation in which AI is involved. Assessment is definitely an important element in defining teacher identity, and the role of AI in assessment processes is a central issue. In addition, comparing teachers' behaviors and attitudes in these assessment situations allows us to investigate their beliefs and attitudes toward AI.

This may help in exploring how far students' use of AI can invalidate assessment. In particular, we use ChatGPT to investigate whether teachers recognize the use of AI, whether this could be useful to bring out beliefs and practices of teachers. Comparing teachers' behaviors and attitudes in these assessment situations allows us to investigate their beliefs and attitudes toward both AI and assessment. Our general problem may be stated as follows: how AI's answers to a mathematical task interface with teachers' assessment practices and beliefs?

METHODOLOGY OF THE STUDY AND ITALIAN CONTEXT

The study involved 23 middle school teachers and a task on prime numbers from the Italian national standardized assessments (INVALSI). The tests are administered every year at census level and student results are provided to each school institution. Results and questions of the INVALSI tests are considered as a resource also for researchers in the field of mathematics education (Garuti & Martignone, 2015; Bolondi & Ferretti, 2021) and are used in national and international research (e.g. Spagnolo et al., 2021).

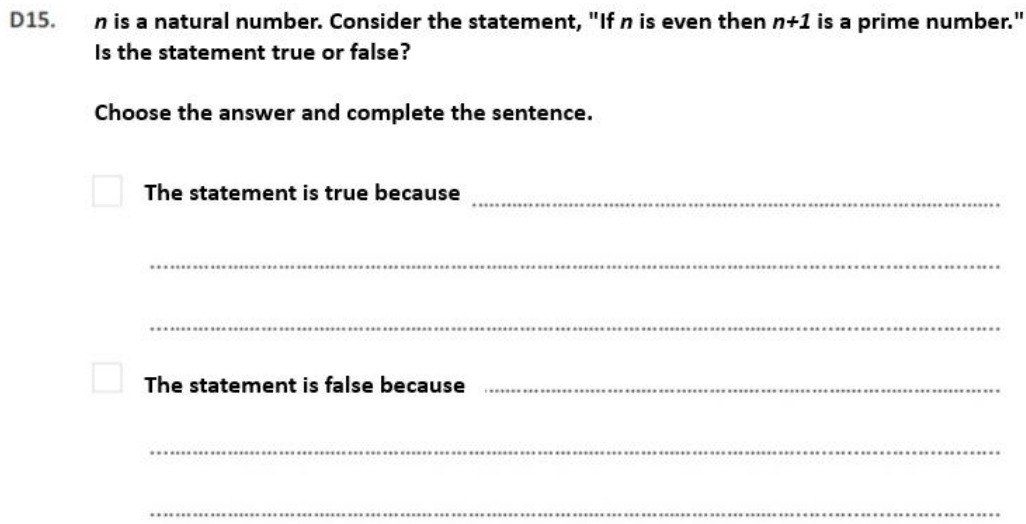
The SNV Framework is designed taking into consideration the Italian National Guidelines, in which argumentation is considered between the competence goals for every school grade (Garuti & Martignone, 2019). For this reason it is possible to find mathematical tasks that require the recognition or production of a justification, as the tasks we selected for our research design.

For this research, results at the national level were used as a benchmark of difficulty and significance at national level; specifically, nationwide in Italy, 37,1% of grade 8 students answer correctly,

41,4% of grade 8 students get it wrong, and 21,6% of grade 8 students do not answer. The task required production of argumentation to support a statement or its denial. Teachers had to assess 5 answers to the task, taking in consideration accuracy of content, accuracy of language, coherence to the task, coherence to the claim, completeness in calculation, and completeness in argumentation. After their assessment, it was disclosed to them that one of the protocols had been produced by AI ChatGPT-3 in answer to the task. None of the protocols (not even the one produced by ChatGPT-3) contained an answer that was considered correct according to standardized assessment standards. After that, teachers tried to identify the answer produced by ChatGPT and to give explicit reasons for their choice. In a second phase, teachers were involved in focus groups and individual interviews to compare criteria and ways in which they had assessed and addressed choices.

In particular, the tasks chosen in this study are represented in Figure 1; the choice of INVALSI task has been made because it allows us to have some extra information that provides a further background for our study, such as the performance of Italian students (which is related to task peculiarities). In our study, we focused on grade 8 INVALSI task, which represent the last year of middle school in Italy.

We identified an INVALSI task (Figure 1) for grade 8 and selected some of the typical answers that grade 8 students provide.


D15. ***n* is a natural number. Consider the statement, "If *n* is even then *n+1* is a prime number." Is the statement true or false?**

Choose the answer and complete the sentence.

☐ **The statement is true because** .....

.....

.....

☐ **The statement is false because** .....

.....

.....

Figure 1. Task D15, Math INVALSI Test Grade 8, 2016, www.gestinv.it.

The same task was proposed to ChatGPT-3 and the response was as follows.

> ChatGPT-3: The statement is false because, for example, when n=2, we have that n is even, but n+1=3 which is a prime number. However, if we take n=4, then n is even, but n+1=5 which is a prime number.
>
> So, it is not possible to say that "if n is even then n+1 is a prime number" in general.

We asked 23 middle school teachers to assess all selected answers (including that of ChatGPT-3). The assessment was done according to the following criteria: accuracy of content, accuracy of language, coherence to the task, coherence to the claim, completeness in calculation, completeness in argumentation. Then it was revealed to the teachers that one of the solutions was made by ChatGPT-3, asking them to recognize it.

DISCUSSION AND PRELIMINARY RESULTS

The data of the national Italian sample are gathered, analized and released by INVALSI. Data can be found at www.gestinv.it. Following, we report the national results of the task shown in Figure 1.

This task aims to assess the ability to use of some typical forms of mathematical thinking (conjecture, argue, verify, generalize); in particular supporting their convictions by bringing appropriate examples and counterexamples and using argumentation to justify their choices.

From a quantitative point of view, the percentage of correct answers of the Italian national sample is 37,1%, while the percentage of incorrect answers is 41.4% and of unanswered questions is 21.6%.

Below we show 4 categories of student answers (Figure 2).

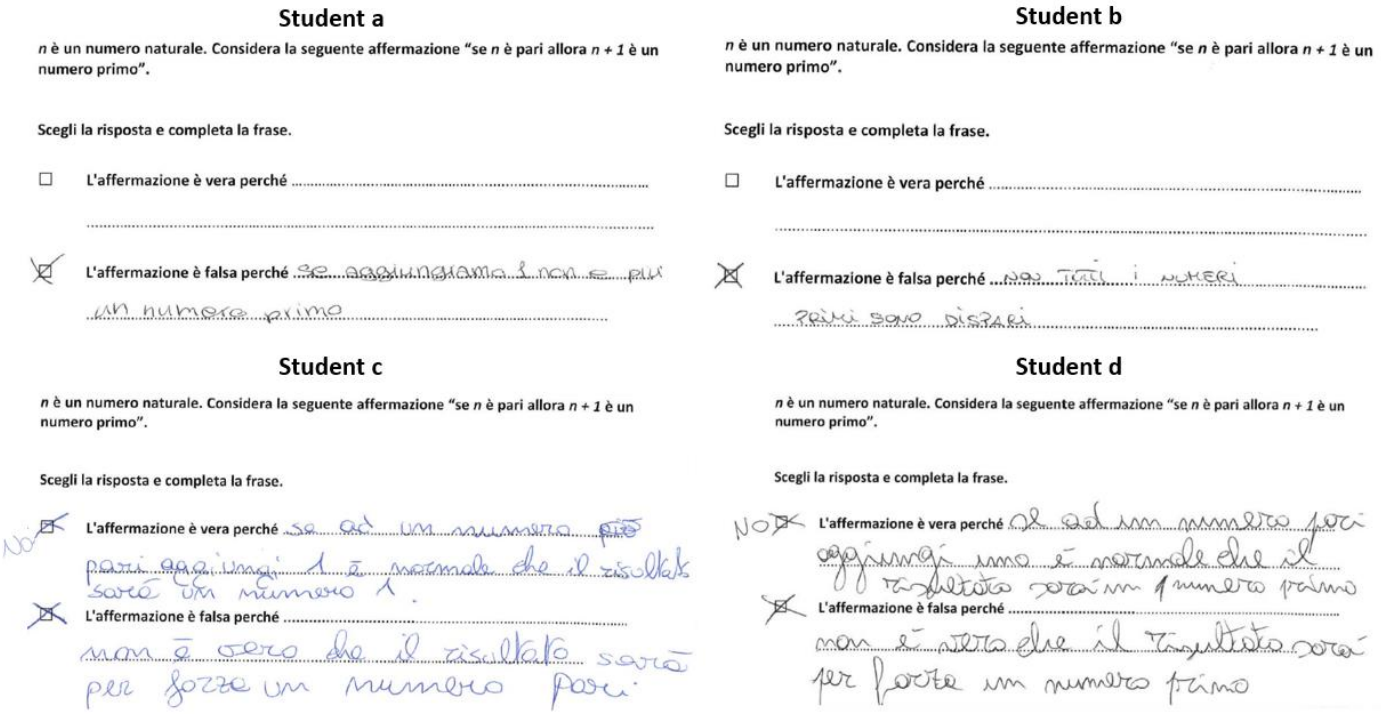


Figure 2. Examples of student answers.

The transcription of the 4 responses (Figure 2) to the task in Figure 1 follows.

Student a: The statement is false because if we add 1 it is no longer a prime number.

Student b: The statement is false because not all prime numbers are odd.

Student c: The statement is true because if to an even number you add 1 it is normal that the result will be a number 1. [Answer deleted]

The statement is false because it is not true that the result will necessarily be an even number.

Student d: The statement is true because if to an even number you add one it is normal that the result will be a prime number. [Answer deleted]

The statement is false because it is not true that the result will necessarily be a prime number.

At the same task ChatGPT-3 answers by giving examples (not counterexamples) and without maintaining consistency with the request of the task and its own claim.

All 5 protocols have been assessed by 23 middle school teachers. After assessing them, we asked them to identify the ChatGPT one. Only 19% of teachers recognized the protocol by ChatGPT-3, while the most chosen protocol was that of Student d (that is "The statement is false because it is not true that the result will necessarily be a prime number") . The percentages are shown in Figure 3: 0.1% of teachers chose Student a answer, 19% chose Student b answer, 14,3% chose Student c answer, 47,6% chose Student d answer, and 19% identified the protocol produced by ChatGPT-3.

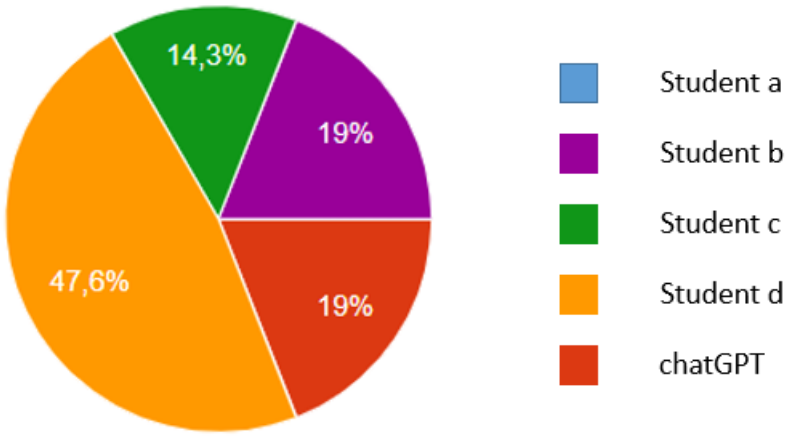


Figure 3. Percentages of teachers' answers to the question: "One of the 5 protocols was created by ChatGPT. In your opinion, which one?".

Then, focus group and interviews have been conducted with the purpose of letting teachers share and make explicit reflections about motivations of their choices. Teacher motivations of their justifications have been analyzed from a qualitative point of view.

We provide examples of answers given by teachers explaining how they try to identified the protocol that they thought had produced ChatGPT-3:

Teacher 1: I chose by exclusion, because no answer seemed to me that could be from ChatGPT.

Teacher 2: I thought it was a trick question and there was no ChatGPT answer.

Teacher 3: I expected ChatGPT should necessarily answer correctly, so I chose the answer that seemed the less wrong.

From this it emerges that interviews allow us to make explicit some aspects of teacher's thinking about AI. For example, these motivations, bring out some teachers' beliefs, such as the idea that AI always gives a correct answer. Results represented in Figure 3 together with in-depth interviews allowed us to hypothesize the characterization of some aspects of teachers' perceptions of AI.

Specifically, 19 out of 23 of teachers interviewed explicitly stated that they had been in trouble trying to recognize ChatGPT's answer because they expected a *correct* one. The idea that AI always replies correctly is a misconception that can affect us as we work with it (whether we use AI to do an assignment or assess something AI has done).

At the end of the interviews, teachers tested ChatGPT-3. Another element that surprised some of the teachers was that the ChatGPT's answer changed radically if the teacher changed formulation of the question slightly, and ChatGPT's answer also changed if the same question was asked twice in the same way. One teacher during the interviews claimed "[…] if every time we ask the same question to ChatGPT it rephrases the answer it might be really difficult for teachers to identify students who use it for cheating!".

Qualitative analysis has clarified some aspects related to the comprehensibility of teaching situations, and some of these can also be used in teacher professional development. We believe that the analyses, put in relation with the motivations of the teachers, will bring out some teachers' beliefs, such as the idea that AI always gives a correct answer.

CONCLUSION

The irruption of AI into school practice creates new situations that can be considered stress tests that can be used by researchers (Wardat et al., 2023). One of the purposes is to gather information regarding these new tools in mathematics education (Rahman, & Watanobe, 2023). It also becomes important to move toward responsible AI in education, aware of the challenges and implications for research and practice (Mhlanga, 2023).

Questionnaire answers and interviews with teachers highlight some of the criteria they use when assessing student argumentation in mathematics. In particular our analysis shows teachers' beliefs about the role of AI in mathematics education, particularly the idea that AI always gives a correct answer.

It is clear that good and specific questions must be asked in order to obtain more and more significant answers to gain knowledge about the object of the research. If the question is generic, the answer is generic, often so much so that it is incorrect (Megahed et al., 2023).

This suggests that the fears surrounding the use of chaptGPT by students, and which have led some U.S. colleges and universities to ban its use, are overblown: if the output of chaptGPT is complex, articulate, and robust, there is accurate, detailed research behind it, grounded in good, surgical questions.

This is one of the fundamental goals of teaching action, and that is, to help students ask good questions in order to actually access the knowledge that is only theoretically available to everyone now.

So two sticking points for students: learning to ask good questions and checking the reliability of the information obtained. It seems to us that this is enough to ensure the need for the presence of competent teachers capable of helping students grow as knowledgeable, critical and ethical subjects. AI can become a formidable ally of the teacher in his or her many activities, the figure of the teacher assumes, in this new context, a strategic role even at a social level. This is a necessary and different approach to knowledge.

Finally, we believe that this study also allows to highlight some teachers' beliefs, such as the idea that AI always answers correctly.

REFERENCES

Aydın, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT generated literature review: Digital twin in healthcare. *Available at SSRN 4308687*.

Bolondi, G., & Ferretti, F. (2021). Quantifying Solid Findings in Mathematics Education: Loss of Meaning for Algebraic Symbols. *International Journal of Innovation in Science and Mathematics Education*. 29. 10.30722/IJISME.29.01.001.

Brousseau, G. (2006). Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990 (Vol. 19). Springer Science & Business Media.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8,* 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Commissione europea, Direzione generale dell'Istruzione, della gioventù, dello sport e della cultura, (2022). *Orientamenti etici per gli educatori sull'uso dell'intelligenza artificiale (IA) e dei dati nell'insegnamento e nell'apprendimento*, Ufficio delle pubblicazioni dell'Unione europea. https://data.europa.eu/doi/10.2766/438

Ferretti, F., Santi, G. R. P., Del Zozzo, A., Garzetti, M., & Bolondi, G. (2021). Assessment practices and beliefs: Teachers' perspectives on assessment during long distance learning. *Education Sciences, 11*(6), 264.

Garrido, A. (2012). AI and mathematical education. *Education Sciences*, *2*(1), 22–32.

Garuti, R., & Martignone, F. (2015). The SNV (INVALSI) experience. Teaching and learning mathematics: resources and obstacles. *Proceedings of CIEAEM*, pp. 95–98.

Garuti, R., & Martignone, F. (2019). Assessment and argumentation: an analysis of mathematics standardized items. In U. T. Jankvist, M. Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Eleventh Congress of the European Society for Research in Mathematics Education,* Freudenthal Group & Freudenthal Institute, Utrecht University and ERME, pp. 4075–4082.

Godino, J. D. (1993). La metáfora ecológica en el estudio de la noosfera matemática. *Quadrante, 2*(2), 69–79.

Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999–1011.

Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., Goldin, G. A. (2016). *Attitudes, beliefs, motivation and identity in mathematics education: An overview of the field and future directions*. Springer Nature.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Knill, O., Carlsson, J., Chi, A., & Lezama, M. (2004). An artificial intelligence experiment in college math education. *Preprint available at http://www. math. harvard. edu/~ knill/preprints/sofia. pdf*.

McLeod, D. (1992). Research on affect in mathematics education: a reconceptualization. In D. Grows (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 575–596). McMillan Publishing Company.

Megahed, F. M., Chen, Y. J., Ferris, J. A., Knoth, S., & Jones-Farmer, L. A. (2023). How generative ai models such as chatgpt can be (mis) used in spc practice, education, and research? an exploratory study. *Quality Engineering*, 1–29.

Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. *Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning*.

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–9). IEEE.

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 57–83.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Spagnolo, C., Capone, R., & Gambini, A. (2021). Where do students focus their attention on solving mathematical tasks? An eye tracker explorative study. Inprasitha, M., Changsri, N., & Boonsena, N. (Eds). (2021). *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (Vol.4). Khon Kaen, Thailand: PME, pp. 89–96.

Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, 1181712.

Xu, W., Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *IJ STEM Ed* **9**, 59 (2022). https://doi.org/10.1186/s40594-022-00377-5

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

ISBN 978-960-466-338-5