Evaluating Student Perceptions of Conversational Agents in Mathematics Learning

Choo Peng Tan1, 2, a) and Choo Kim Tan1, 2, b)

*1Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450, Bukit Beruang, Melaka, Malaysia.*

*2Centre for Intelligent Cloud Computing, CoE for Advanced Cloud, Multimedia University, Jalan Ayer Keroh Lama, 75450, Bukit Beruang, Melaka, Malaysia.*

*a) Corresponding author: cptan@mmu.edu.my  
b) cktan@mmu.edu.my*

**Abstract.** The incorporation of conversational agent (CA) in educational has become a ubiquitous trend due to its ability to provide instant and interactive individualized support for students. This paper investigates undergraduates' perceptions of the use of CA in their mathematics learning. A mixed-method approach with convenient sampling is adopted with a 5 Likert-scales survey and interviews as instruments. This study reveals that students perceive positively towards the incorporation of CA in their mathematics, irrespective of their gender or academic performance level. This study may provide insights regarding the relevance of these factors in educational CAs and to all stakeholders to optimize CAs in enhancing undergraduates (mathematics) learning experiences.

# introduction

The rapid growth of artificial intelligence (AI) has brought transformative changes in many industries including education. Traditional teaching methods often result in student disengagement [1]. Thus, there is need for innovative pedagogies that integrate ICT, such as e-learning and educational software [2]. Conversational agents (CAs), also known as chatbots or virtual assistants, one of the AI tools, have gained prominence and being widely used in various sectors, including education. CA is an automated AI computer program designed to simulate human-like interactions with human users to provide instant responses and personalized support in assisting various tasks through text or voice-based communication [3]. Meanwhile, there has been a notable increase in student access to new technologies [4]. This technological growth benefits both students and educators by leveraging information and communications technology (ICT) to accelerate learning and improve engagement.

CAs have gained prominence due to their ability to engage students in interactive learning environments, offering real-time or immediate feedback [5], assist with learning activities [6], foster a more enjoyable and personalized learning environment [7], and, hence, enhance e-learning satisfaction, and reduce dropout rates at which [8] emphasize that CAs support scaffolding and interactivity, significantly increase students' enjoyment of learning while alleviating their anxiety [9]. CAs used in mathematics education can help explain concepts, solve problems, and provide hints tailored to the learner's progress [6], [10]. All these, in turn, promote a positive attitude towards students learning.

Studies have shown that when technologies are perceived as easy to use and useful, their adoption is more likely, which eventually benefits both users and service providers. These factors have been widely studied in various technological innovations, but their relevance in the domain of educational CAs remains underexplored [11], [12]. Understanding students' perceptions of CAs' usefulness and ease of use is important for designers, educators and all stakeholders to optimize the tools and ensure their effective adoption and success [13] and enhance user engagement and overall satisfaction [12].

This paper is to explore students' perceptions of the usefulness and ease of use of CA that incorporate in their mathematics learning. The research questions are:

RQ1: What are students’ overall perception towards the conversational agent in their learning mathematics?

RQ2: Are there significant differences in perception based on gender?

RQ3: Are there significant differences in perception based on student’s performance level?

Given the expanding role of AI in education, such research is pivotal for ensuring that CAs can be harnessed effectively to enhance the educational experience.

# Literature Review

Various technologies have been integrated into education that has led to significant shifts in teaching and learning methodologies. CAs are the most recent AI tools used in various industries including education which as the most promising innovations. Educational CAs are AI-driven systems designed to mimic the knowledge acquisition and interaction style of human teachers [14], [15].

Students generally hold positive views towards CAs and satisfy with CAs due to their accessibility and the ability to provide instant, personalized feedback, and contextually relevant and easily understandable responses [16]. CAs are also seen as tools that foster self-directed learning [17]. For example, [18] demonstrated that a CA system significantly improved students’ learning outcomes, particularly in fostering mathematical argumentation skills. Students who used CAs in mathematics learning have reported that CAs clarify difficult concepts and offer personalized explanations that traditional instructional methods might not be able to provide [19]. Furthermore, the project of intelligent tutoring systems (ITS), the concept is closely related to CAs, by [20], demonstrated that tailored guidance based on students' progress can improve learning outcomes in mathematics.

Students’ perceptions towards the educational CAs are important factors to ensure the implementation of educational CAs is effective. Their perceptions are strongly influenced by the perceived usefulness and ease of use of these tools. [21] mentioned that students are more likely to use the tools regularly when they find technology is useful in achieving their learning goals and easy to navigate. Furthermore, the tools which align with students’ expectations of improving learning outcomes and providing a user-friendly experience foster greater engagement and long-term use [13], [22].

Often, males having higher self-efficacy and tend to exhibit more positive attitude towards technology than females. Thus, gender could be a significant factor influencing students’ perceptions of CAs. Males are likely to be more comfortable with technology; hence, they often show more favorable attitudes towards technology learning tools [23]. Contrarily, females might show more higher anxiety level or cautions attitudes towards technology, especially when the technology is perceived as impersonal or unsupportive [24]. Similar trends could be observed in the context of mathematical CAs education at which male students tend to show favorable experiences with CAs, especially in problem-solving or technical tasks [25]. Conversely, female students tended to be more critical of factors such as emotional responsiveness and communication style [26]. To reduce gender-based discrepancies in perceptions and enhance engagement for all students, [27] suggests that gender-neutral designs that allow the customization of voice, tone, and responsiveness, could be implemented.

Besides, different performance levels of students may engage with educational CAs with different attitudes, interests, and behaviors because of their varying confidence levels, prior mathematics experiences, and learning needs [22]. The relationship between student performance level and technology use is complex. Students of high-performance perceived positively towards CAs as they appreciate CAs as valuable tools in refining their understandings of concepts [15]. Moreover, they motivate by the advanced problem sets by CAs that stimulate their intellectual curiosity [28].

Conversely, students of low performance may struggle with the personalized nature of CAs [13] or more likely to have a negative experience when CAs provided limited help [10], especially when dealing with basic or foundational concepts. Research by [24] found that students of low performance often felt unsupported by CAs, especially the subjects like mathematics, where they require more guidance to succeed. Without adequate scaffolding, these students may feel overwhelmed, disengaged, and dissatisfied with their learning experience.

Personalized feedback, without doubt, could foster student engagement and improve learning outcomes [29]. In mathematics education, CAs play a crucial role in supporting both foundational skills and more complex by offering repetitive practice and immediate feedback [19]. The personalized learning experiences provided by CAs are particularly beneficial for students who struggle with abstract concepts at which they offer tailored hints and explanations at students’ own pace [29].

Based on those studies mentioned above, educational CAs designers may consider a gender-neutral design to reduce the gender-based discrepancies in perceptions. While students of low performance require more guidance and support to be fully benefited from CAs, students of high-performance strive from the complexity and challenges offered by CAs. With these, CAs could continuously grow and be optimized to use as a potential learning tool in mathematics education. Although there are extensive studies on the adoption of educational technologies in general, research focusing on students' perceptions towards CAs remains scanty. This study is to address this gap by investigating students' perceptions towards the use CAs in mathematics learning, particularly focus on how gender and academic performance levels may influence these perceptions.

# Methodology

This study adopts a quasi- experimental and mixed-method (both quantitative and qualitative research approach), utilizing a survey-based and interviews design to collect data on students' perceptions of CAs in educational settings. The survey, the questionnaire on students' perceptions of conversational learning agents (SPCA), is an online tool adapted from previous studies [30 – 32]. It evaluates students' views on the usefulness and ease of use of CA. It has two sections: Section A gathers demographic data, and Section B includes 25 statements where students rate their perceptions on a 5-point Likert scale, with higher values indicating stronger agreement.

The CA is designed and developed grounded in Vygotsky Theory to scaffold the students in their mathematics learning. The sample, which is the convenient sample of the instructor, includes 30 university IT students who interacted with CA for a trimester. Participants self-registered the group based on their preferable timetable a trimester before the subject is offered.

Although the sample size in this study is small (*n* = 30), the use of independent samples *t*-tests to examine differences in students’ perceptions based on gender and performance level is statistically justified. However, “*30 data points should provide enough information to make a statistically sound conclusion about a population*” [33]. As recommended in recent literature [34], small-sample *t*-tests can yield valid results when assumptions are satisfied, and effect sizes and confidence intervals are reported alongside *p*-values.

The study is completed in three phases. In the first phase, first, the research objectives are established, and the sample is determined. Then, the research methodology is chosen. At the same time, the lessons content and CA are designed and developed, and the research instrument (questionnaire) is adopted. The permission from the faculty dean and instructor are obtained. In the second phase, the field study is conducted for a duration of a trimester (14 weeks). The students are briefed, and consent is obtained at the beginning of the intervention. Then, they are taught by the instructor in physical classroom and are provided by the developed CA to be used outside of classroom. The structured-adopted questionnaire is given after they have completed the learning with CA. The survey is conducted electronically to ensure accessibility and ease of participation. Also, the interview is conducted to collect qualitative data. In the third phase, the data collected from the questionnaire is analyzed using SPSS. Descriptive statistics and inferential statistics (at 0.05 significance level) are used.

# Results

## Participants Demography

The sample is the convenient sample of the instructor at which the students self-registered to the group based on their preferred timetable, a trimester before the subject is offered. The participants consist of 46.67% female and 53.33% male who’s aged 19 – 24 years old with a mean age of 20.07 years old. Approximately 40% of respondents reported using CAs before but the usage was not in education with the most common use cases being travelling inquiries, and hobbies. Meanwhile, the students are categorized based on their mathematics achievement prior to joining the degree program, as shown in Table 1, at which low are for grade C and C+, average are for B–, B, and B+, and high are for A–, A, and A+.

|  |  |  |
| --- | --- | --- |
| TABLE 1. Performance levels distributions | | |
| Performance Level | Frequency | Percentage (%) |
| Low | 5 | 16.67 |
| Average | 10 | 33.33 |
| High | 15 | 50.00 |
| Total | 30 | 100.00 |

## Validity and Reliability of Instrument

The instrument (questionnaire) is sent for validation by five experienced lecturers in an established local university. The language in the instrument is also vetted by the English lecturer. Furthermore, reliability and validity are calculated and generated by SPSS using Cronbach alpha coefficient (α) and Exploratory Factor Analysis (EPA). The α generated is .946, implies that the instrument has internal consistency [35] and is reliable [36] in measuring students’ perception on the use of CA in their mathematics learning. In the Exploratory Factor Analysis (EPA), an item with factor loading of at least .500 is retained.

## Normality Test

The normality test of the data is conducted through the skewness and kurtosis analysis. Skewness measures the symmetry of a distribution, while kurtosis measures the degree to which the data cluster in the tails of a distribution [37]. The values of the coefficient of skewness is .221 and coefficient of kurtosis is –.740. Both the values are within absolute value of 1, considered as excellent value and the data are reasonably symmetric and normally distributed [35].

## Descriptive Statistics

The analysis of SPCA was conducted to evaluate the students’ perceptions of CA. The overall mean score is 4.1400 with a standard deviation of .4832, indicating a generally high level of positive perception towards the use of the conversational agent in mathematics learning, with moderate consistency in responses.

## t-tests

t-tests (including ANOVA) are used to compare the means of the independent groups (gender and performance levels) to determine if there is a statistically significant difference among them.

First, the t -test was conducted to determine if there are significant differences in students' perceptions of using a conversational agent (CA) in mathematics learning based on gender. The null hypothesis (H1) is as follows:

H1: No significant differences in perception based on gender.

Table 2 summarizes the t -test results.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **TABLE 2**. *t*-test based on gender | | | | | | | |
| **Gender** | **Mean** | **SD** | **MD (SD)** | **95% confidence interval of the mean difference** | | ***t*-value** | ***p*-value** |
|  |  |  |  | **Lower** | **Upper** |  |  |
| Female | 4.0800 | .1767 | -.1125 | -.4785 | .2535 | -.630 | .534 |
| Male | 4.1925 | .2919 | (.1787) |  |  |  |  |

The *p*-value of .534 is greater than the significance level. This means we fail to reject the null hypothesis, indicating that there are no significant differences in students' perceptions based on gender.

The Cohen’s d is -.2304 indicates the effect size is small.

Next, the ANOVA test was conducted to discover if there are significant differences in students' perceptions of using CA in mathematics learning based on student’s performance level. The null hypothesis (H2) is as follows:

H2: No significant differences in perception based on student’s performance level.

Table 3 summarizes the ANOVA results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TABLE 3**. ANOVA based on performance level | | | | | |
|  | **Sum of Squares** | ***df*** | **Mean Square** | ***F*** | **Sig.** |
| Between Groups | .152 | 2 | .076 | .310 | .736 |
| Within Groups | 6.618 | 27 | .245 |  |  |
| Total | 6.770 | 29 |  |  |  |

The p-value of .736 is greater than the significance level. This means we fail to reject the null hypothesis, indicating that there are no significant differences in students' perceptions based on students’ performance level.

The η2 is 0.4832 indicates the effect size is very large.

# Discussion

This study is to understand undergraduates' perceptions of using a CA in their mathematics learning. The results revealed several key findings.

First, the overall mean score of 4.1400 indicates that students were supportive of using CA as a learning tool, despite it being their first experience using it in learning. The results of this study demonstrate that students' perceptions of the usefulness and ease of use of educational conversational agents significantly influence their intention to use these tools. This is consistent with previous research in the broader field of educational technology adoption, which has shown that students are more likely to engage with tools that they find both useful and easy to use [13].

The students in this study, who used a digital instructor for one-on-one, self-paced learning, expressed satisfaction with the human-like conversations provided by the CA [38], [39] during the interview session too. They preferred this personalized approach to learning as they could learn at their own pace and preferable time through their own devices. Therefore, students perceive CA as enjoyable, a sentiment supported by [40] and [41] at which personalized CAs could tailor interactions to individual learning preferences. Additionally, students feedback that the responds are instant, and it is a "tireless" learning assistant, available at any time to support their learning, even when the content was repeated [42].

Second, the results revealed the lack of significant differences in males and females’ perceptions toward CA which is consistent with broader educational research. The use of technology in education tends to benefit all students equally [43]. This result suggests that CAs could be a valuable tool in promoting gender equity in mathematics education by providing equal opportunities for engagement and support.

Third, the ANOVA results indicated that students of different performance levels perceive the same in using CA in their learning. This implies that CAs may offer a consistent level of support and engagement, and provide a uniform learning experience to all students, regardless of their academic performance levels. This is especially important in mathematics education as anxiety and negative attitudes can hinder students’ performance. CAs can reduce mathematics anxiety and improve students' attitudes towards the subject [44].

In addition, this study and the design of CA in this study is grounded on Vygotsky’s Theory that scaffolds students in their Zone of Proximal Development. Meanwhile, students have the option to learn the problem solving with hints or step-by-step guidelines and solutions, which can significantly improve their learning experience [20]. Students’ motivation and engagement tend to increase when CAs offer clear or scaffolded support to help the students to progress at their own pace [45]. These might be the factors that contribute to the uniform perception of students with different performance levels.

# Conclusion

Overall, the students of this study favor using CA as their mathematics learning tool regardless of their gender or academic performance level. The findings may provide insights to guide in the design and implementation of CA to meet students’ needs and promote adoption in other subjects. Future research should explore the incorporation of CAs in larger samples, to other subjects, and study its long-term impact and potential strategies to optimize its effectiveness.

# Acknowledgments

We sincerely thank all the participants for their invaluable contributions in this study.

# References

1. M. C. Pascoe, S. E. Hetrick, and A. G. Parker, “The impact of stress on students in secondary school and higher education,” International Journal of Adolescence and Youth **25**, 104–112 (2020).3
2. V. Nikolić, D. Petković, N. Denić, M. Milovančević, and S. Gavrilović, “Appraisal and review of e-learning and ICT systems in teaching process,” Physica A: Statistical Mechanics and its Applications **513**, 456–464 (2019).
3. A. Følstad and P. B. Brandtzaeg, “Chatbots and conversational agents: A bibliometric analysis of research trends,” Computers in Human Behavior **121**, 106–119 (2021).
4. A. Schleicher, *PISA 2018: Insights and Interpretations* (OECD, 2019)*.*
5. D. E. Gonda, J. Luo, Y. L. Wong, and C. U. Lei, “Evaluation of developing educational chatbots based on the seven principles for good teaching,” in 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (IEEE, Wollongong, Australia, 2018), pp. 446–453.
6. D. Joubert and A. Al-Khouri, “Exploring the role of conversational agents in enhancing student learning and engagement,” Journal of Educational Technology **42**, 45–58 (2023).
7. R. Bogue, D. Smith, and C. Tully, “Personalized learning with conversational agents: A case study,” International Journal of Artificial Intelligence in Education **30**, 500–517 (2020).
8. I. Dekker, E. M. de Jong, M. C. Schippers, M. de Bruijn-Smolders, A. Alexiou, and B. Giesbers, “Optimizing students’ mental health and academic performance: AI-enhanced life crafting,” Frontiers in Psychology **11**, 1–15 (2020).
9. T.-C. Hsu, H.-L. Huang, G.-J. Hwang, and M.-S. Chen, “Effects of incorporating an expert decision-making mechanism into chatbots on students’ achievement, enjoyment, and anxiety,” Educational Technology & Society **26**, 218–231 (2023)
10. S. Chai and P. Wong, “Conversational agents in STEM education: Enhancing learning experiences through AI-driven interactions,” International Journal of Artificial Intelligence in Education **33**, 125–139 (2023).
11. S. Kelly, S.-A. Kaye, and O. Oviedo-Trespalacios, “What factors contribute to the acceptance of artificial intelligence? A systematic review,” Telematics and Informatics **77**, 101925 (2023)
12. L. Labadze, M. Grigolia, and L. Machaidze, “Role of AI chatbots in education: Systematic literature review,” International Journal of Educational Technology in Higher Education **20**, 56 (2023)
13. W. Li and J. Hsieh, “Students’ acceptance of chatbots in higher education: An extended technology acceptance model,” Computers in Human Behavior **119**, 106–119 (2022).
14. S. Han and M. K. Lee, “FAQ chatbot and inclusive learning in massive open online courses,” Computers and Education **179**, 104395 (2022)
15. J. Q. Pérez, T. Daradoumis, and J. M. M. Puig, “Rediscovering the use of chatbots in education: A systematic literature review,” Computer Applications in Engineering Education **28**, 1549–1565 (2020).
16. G.-J. Hwang and N.-S. Chen, “Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions,” Educational Technology & Society **26**, 1–18 (2023)
17. A. Kukulska-Hulme, C. Bossu, K. Charitonos, T. Coughlan, A. Deacon, N. Deane, R. Ferguson, C. Herodotou, C.-W. Huang, T. Mayisela, I. Rets, J. Sargent, E. Scanlon, J. Small, S. Walji, M. Weller, and D. Whitelock, *Innovating Pedagogy 2023: Open University Innovation Report 11* (The Open University, Milton Keynes, 2023).
18. X. Wu, S. Li, H.-T. Wu, Z. Tao, and Y. Fang, “Does retrieval-augmented generation introduce unfairness in large language models? Evaluating fairness in retrieval-augmented generation systems,” in Proceedings of the 31st International Conference on Computational Linguistics (Association for Computational Linguistics, Abu Dhabi, UAE, 2025), pp. 10021–10036.
19. Z. Cai, Q. Wang, and L. Chen, “Enhancing learning with conversational agents: A study on personalized explanations and concept clarification,” Journal of Educational Technology **45**, 123–137 (2022).
20. J. M. Spector and R. Huang, *Educational Technology: A Primer for the 21st Century* (Springer, 2022).
21. P.-N. Chou and C.-H. Chen, “The impact of technology acceptance on learning outcomes: A study on the use of educational technology in higher education,” Journal of Educational Technology Development and Exchange **15**, 45–60 (2022).
22. T. Buchanan, X. Liu, and J. Lee, “The role of educational chatbots in enhancing student engagement and learning outcomes,” Educational Technology & Society **23**, 45–56 (2020).
23. Y. Zhao, X. Liu, and J. Zhang, “Gender differences in attitudes towards AI-based learning tools: A comparative study,” Computers and Education **168**, 104213 (2021).
24. Y. Liu and X. Li, “Gender differences in perceptions of AI-based educational tools: A comparative analysis,” Computers and Education **172**, 104321 (2023).
25. J. Lee, S. Kim, and H. Park, “Gender differences in the use of conversational agents for problem-solving tasks,” Journal of Educational Technology **38**, 215–230 (2022).
26. H. Park and J. Kim, “AI capabilities and competitive advantage: Evidence from technology industries,” Strategic Management Journal **44**, 922–947 (2023).
27. D. Rao, X. Miao, Z. Jiang, and R. Li, “STANKER: Stacking network based on level-grained attention-masked BERT for rumor detection on social media,” in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Association for Computational Linguistics, Punta Cana, 2021), pp. 3347–3363.
28. Q. Zhou, J. Ma, Q. Liu, C. Wu, Z. Yang, T. Yang, Q. Chen, Y. Yue, and J. Shang, “Traditional Chinese medicine formula, Sanwujiao granule, attenuates ischemic stroke by promoting angiogenesis through early administration,” Journal of Ethnopharmacology **321**, 117418 (2023).
29. L. Huang, X. Feng, W. Ma, Y. Gu, W. Zhong, X. Feng, W. Yu, P. Peng, D. Tang, D. Tu, and B. Qin, “Learning fine-grained grounded citations for attributed large language models,” in Findings of the Association for Computational Linguistics: ACL 2024 (Association for Computational Linguistics, Bangkok, Thailand, 2024), pp. 14095–14113.
30. R. G. P. Galluccio, “Animated pedagogical agents as Spanish student performance, motivation, and appearance, and type of activity on language instructors: Effect of accent, perception of agent,” Ph.D. dissertation, The Florida State University (2008).
31. W. Yeo, “Reducing anxiety level in mathematics learning using pedagogical agent,” Ph.D. dissertation, Multimedia University, Malaysia (2016).
32. Y. M. Yusoff, Z. Muhammad, M. S. M. Zahari, E. S. Pasah, and E. Robert, “Individual differences, perceived ease of use, and perceived usefulness in the E-library usage,” Computer and Information Science **2**, 76–83 (2009).
33. P. Reagen, “The importance of identifying the right sample size for business improvement,” (Leanscape, 2023).
34. J. C. F. de Winter, “Using the student’s t-test with extremely small sample sizes,” Practical Assessment, Research & Evaluation **18**, 1–12 (2013).
35. D. George and P. Mallery, *IBM SPSS Statistics 29 Step by Step: A Simple Guide and Reference*, 18th ed. (Taylor & Francis, 2024).
36. J. R. Fraenkel and N. E. Wallen, *How to Design and Evaluate Research in Education*, 10th ed. (McGraw-Hill Education, 2020).
37. A. P. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. (SAGE Publications, London, 2018).
38. I. Scarpellini and Y. Lim, “Role-based design of conversational agents: Approach and tools,” in HCI International 2020 – Late Breaking Posters (Springer, 2020), pp. 366–375.
39. E. Fleisch, C. Franz, and A. Herrmann, *The Digital Pill: What Everyone Should Know About the Future of Our Healthcare System* (Emerald Publishing, 2021).
40. C. Strohmann, “Stereoselective synthesis of cyclobutanes by contraction of pyrrolidines,” Journal of the American Chemical Society **143**, 18864–18870 (2021).
41. S. E. Nißen, K. Wolski, L. Cho, et al., “Lipoprotein(a) levels in a global population with established atherosclerotic cardiovascular disease,” Open Heart **9**, e002060 (2022).
42. W. Huang, K. F. Hew, and L. K. Fryer, “Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning,” Journal of Computer Assisted Learning **38**, 237–257 (2022).
43. OECD, *Gender, Education and Skills: The Persistence of Gender Gaps in Education and Skills* (OECD Publishing, 2023).
44. X. Li and Y. Xing, “The impact of conversational agents on mathematics anxiety and attitudes,” Journal of Educational Technology **45**, 123–135 (2022).
45. J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, “ILVR: Conditioning method for denoising diffusion probabilistic models,” arXiv:2108.02938 (2021)