Jurnal Pendidikan Malaysia 48(2)(2023):73–83
DOI: http://dx.doi.org/10.17576/JPEN-2023-48.73-83

The Performance Satisfaction and Usability of ChatGPT in English Language Learning Among Matriculation College Students in Malaysia

Nurafiqah Syuhada Abdul Rahim & Aidah Abdul Karim

ABSTRACT

Although the application of artificial intelligence (AI) AI techniques are being utilized in English language learning, there is a lack of research measuring ChatGPT's usability from the learners' perspective. This study examined the level of ChatGPT's performance satisfaction and usability for English language learning as experienced by matriculation college students upon a completion of a series of tasks pertaining to English language learning with ChatGPT. The study employed case study research design and involved 30 matriculation college students. Data was collected using questionnaires and analyzed using descriptive statistics such as mean, standard deviation, skewness, and kurtosis. The findings showed that there is a high level of satisfaction of ChatGPT performance in all English language learning tasks of conversation, writing, grammar and vocabulary, grammar, and vocabulary with highest satisfaction in conversation and lowest satisfaction in writing tasks. Similarly, the findings indicated that there is a high level of ChatGPT usability in five dimensions of system usability, usefulness, ease of use, ease of learning, and satisfaction, with highest score in the ease of learning and lowest score in ease of use of ChatGPT in English language learning. The findings further suggested that male students and younger students score higher on the usability of ChatGPT compared to female students and older students. The study implicated there is a need for higher education institutions to train their academic staff on the use of ChatGPT in English language learning due to a high ChatGPT performance satisfaction and usability of ChatGPT in English language learning. Additionally English educators in higher education might also considered appropriate strategies to minimize the gender age difference in using ChatGPT in English language learning.

Keywords: English language learning; ChatGPT; performance satisfaction, usability; higher education; Malaysia

INTRODUCTION

Proficiency in English can enhance job prospects, encourage professional and academic communication, and expand one's language proficiency. The structured learning of English within educational systems aids learners in flourishing skills in grammar, vocabulary, and language functions (reading, writing, speaking, listening) dexterously. Ubiquitous structured methods for English learning in the classroom include group discussions, debates, and presentations, which collectively enhance speaking, reading, writing, and listening skills. In today's globalized world, language learning is essential, and language learners now have greater opportunities associated with rising technologies (Klimova & Zaborova, 2023). Among these technologies, "Chatbots" have gained approval, with increasing attentiveness in their capability for institutional English language learning (Huang & Huang, 2022). Chatbot technology offers a conversational environment where users can interact similarly to human interaction (Jeon, 2021). Users can query chatbots for specific information, and the chatbots can provide feedback on various language aspects, such as grammar or vocabulary, upon request. This makes them suitable for English language learning (Haristiani, 2019). Jeon's (2021) empirical study demonstrated that chatbots not only promote vocabulary acquisition but also provide diagnostic information about individual learners' vocabulary learning. Moreover, research in Computer-Assisted Language Learning (CALL) suggests that rule-based, scripted voice systems are particularly effective for language learning (Godwin-Jones, 2022).

The effectiveness of chatbots in language learning is contingent on their usability, efficiency, and effectiveness (Ren et al., 2019). Usability is elucidated as "the extent to which a program can be used to achieve specific objectives with effectiveness, efficiency, and satisfaction in a particular context of use" (Peterson et al., 2019). In the recent past, ChatGPT, developed by OpenAI using the GPT-3.5 architecture, has emerged as the most advanced

conversational AI tool. GPT stands for "Generative Pre-trained Transformer," a state-of-the-art large language model trained on extensive text data, enabling it to generate human-like contextual text responses. ChatGPT understands text input and generates contextual responses based on learned patterns from its training data, though it may occasionally produce incorrect answers. ChatGPT is fine-tuned with a dialogue dataset to generate conversational responses with back-and-forth interactions between the user and the model (Fitria, 2023).

Although the application of ChatGPT is widely utilized in English language learning, there is a lack of studies that measure ChatGPT's usability from the learners' perspective. This study examined the level of ChatGPT's usability for English language learning as experienced by matriculation college students upon a completion of a series of tasks pertaining to English language learning with ChatGPT. The findings would be able to inform English educators on the best possible ways to integrate ChatGPT application in English language learning among higher education students, particularly matriculation college students.

LANGUAGE ACQUISITION AND EMERGING LANGUAGE LEARNING TECHNOLOGIES (ELLTS)

Language acquisition involves gaining the necessary knowledge and skills to communicate proficiently in a foreign or second language, such as English, thereby enabling effective interaction with native speakers. This process typically includes the development of competencies in various language domains, including writing, reading, speaking, and listening (Klimova, 2011). The importance of learning a foreign language is multifaceted, encompassing enhanced communication abilities, improved personal and professional prospects, and cognitive development. In recent years, Emerging Language Learning Technologies (ELLT) have revolutionized the language learning landscape. These innovative tools harness advanced techniques such as virtual reality (VR), artificial intelligence (AI), and natural language processing (NLP), as well as gamification, to enrich the language learning experience. ELLTs are designed to offer a more immersive, effective, and personalized approach to language acquisition (Klimova & Zaborova 2023). Table 1 provides a summary of some common ELLTs found in the literature, including their descriptions and intended purposes.

Table 1: Existing studies in ELLTs

Classification	Interpretation	Aims	Available software or applications
Virtual reality tools for language learning	Virtual Reality (VR) language learning tools immerse learners in virtual environments where they can practice language skills in realistic, simulated settings. This immersive experience helps in contextual learning and enhances retention.	To provide immersive, experiential learning environments that simulate real-life scenarios, helping learners practice and improve their language skills in context.	Mondly VR: Offers immersive language lessons in virtual reality, allowing users to engage in interactive conversations and scenarios (Liu et al 2021) ENGAGE VR: Provides a platform for creating virtual classrooms and interactive language lessons. (Govindarajan and Christuraj 2023)
Chatbots and conversational AI	Chatbots and conversational AI provide interactive language practice by simulating conversations with users. These tools offer a safe, non-judgmental environment for learners to practice speaking and writing.	To offer interactive, conversational practice in a safe environment, helping learners improve their fluency, confidence, and conversational skills.	Replika: An AI chatbot that engages users in conversation to help improve their language skills. (Brown and Stent 2021) ChatGPT: Provides conversational practice and feedback on language use, helping learners refine their speaking and writing skills. (Fitria 2023)

continue ...

... cont.

Gamification-based language learning	Gamification-based language learning incorporates game design elements such as points, badges, and leaderboards into educational activities. This approach increases motivation and engagement through competition and rewards.	To enhance learner engagement and motivation by making language learning fun and competitive, encouraging consistent practice and progress.	Lingodeer: Offers gamified language lessons with interactive exercises and rewards (Huang and Huang 2022) MindSnacks: to improve foreign language skills, including the ability to recognize and spell new words. (Prathyusha 2020)
AI-powered applications for language learning	AI-based language learning apps leverage artificial intelligence to provide personalized learning experiences. These apps adapt to individual learning styles and paces, offering tailored lessons, feedback, and practice exercises.	To enhance language learning efficiency and effectiveness by providing adaptive, personalized learning experiences that cater to individual learner needs.	Duolingo: Uses AI to tailor lessons and provide personalized feedback based on user performance. (Brown et al., 2020) Babbel: Offers AI-driven courses that adapt to the user's learning speed and retention (Chen & Chen 2021)
Speech recognition-based language learning	Speech recognition-based tools use technology to analyze and provide feedback on learners' pronunciation and speaking skills. These tools help improve speaking accuracy and fluency through real-time correction and practice.	To help learners improve their speaking and pronunciation skills by providing instant, accurate feedback on their spoken language.	Pronunciation Coach: Analyzes speech and provides detailed feedback on pronunciation (Jeon, 2021) Elsa Speak: Uses speech recognition to offer personalized pronunciation feedback and exercises (Haristiani, 2019)
Augmented reality-based language learning tools	Augmented Reality (AR) language learning tools overlay digital content and interactive experiences onto the real world through devices like smartphones or AR glasses. These tools create immersive, contextual learning environments.	To enhance engagement and contextual learning by integrating interactive, immersive experiences into the real world, helping learners visualize and interact with language concepts in a practical setting.	Mondly AR: Uses AR to create interactive language learning experiences with virtual objects and scenarios. (Klimova and Zaborova, 2023) AR Flashcards: Provides vocabulary learning through 3D objects and animations overlaid on physical flashcards (Peterson and Peterson, 2019)
NLP-based language learning platforms	Natural Language Processing (NLP) based platforms use advanced algorithms to understand and generate human language. These platforms offer interactive exercises, instant feedback, and conversational practice.	To facilitate interactive and dynamic language learning by understanding and processing human language, enabling real-time feedback and conversational practice.	Rosetta Stone: Uses NLP to provide real-time speech recognition and feedback, helping users improve their pronunciation and fluency (Wang and Wang, 2020) Memrise: Utilizes NLP to enhance vocabulary learning and provide interactive language practice (Song and Song, 2022)

As indicated in Table 1, chatbots as part of AI applications have offered a great potential to assist language learning. Recently, chatbots have become prominent in language learning due to their ability to engage in one-on-one conversations with users through natural language processing (NLP) in various target languages (Jeon et al., 2023). Users utilize chatbots for daily language practice and learning activities such as conversations, generating and answering questions, and conducting vocabulary assessments (Jia et al., 2012). Research indicates that chatbots can facilitate different types of interactional exercises, encouraging learners to produce more language output in a low-anxiety environment (Chen, 2011).

The most used chatbots in language learning contexts exhibit three primary features: (1) continuous support for users and students rounds the clock, unlike human partners; (2) extensive language information that human partners might lack; and (3) the ability to perform repetitive tasks such as answering frequently asked questions and providing language practice. In their systematic literature review, Jeon et al. (2023) identified English as the most prevalent language used in chatbots for language learning. The review highlighted that the most common tasks users engage in with chatbots for language learning include speaking, listening, writing, reading, vocabulary, and grammar practice, as well as conversational exercises. Most studies focus on higher education students (e.g., undergraduate and postgraduate levels), though some research also addresses primary and secondary school students, demonstrating positive impacts on students learning English as a foreign language (Yang et al., 2022). Okonkwo and Ade-Ibijola (2021) conducted a systematic review of chatbot use in education, reporting that majority of studies fall within the "Teaching and Learning" domain (66%), followed by "Research and Development" (19%), "Advisory" (4%), "Assessment" (6%), and "Administration" (5%). Key studies related to chatbots in teaching and learning are discussed by Akcora et al. (2018), Chen et al. (2020), and Hobert and Berens (2020).

Despite advancements in using emerging technologies for language learning, questions remain regarding their effectiveness, user experiences, satisfaction, and engagement. Garzon et al. (2023) conducted a meta-analysis on using mobile applications for learning English, concluding that mobile applications are a valid alternative to traditional methods such as flashcards. Their findings indicate that mobile apps are particularly effective at the bachelor's level, providing better outcomes than traditional lecture-based settings. Luo (2023) performed a systematic review of gamification tools for foreign language learning, emphasizing the need to understand their effectiveness. The study identified various positive, negative, and neutral impacts of gamification techniques and outlined factors affecting their effectiveness. Additionally, Farrokhnia et. al. (2023) conducted a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis of ChatGPT for language learning and education. Table 2 presents key points from the SWOT evaluation, revealing that, despite numerous strengths and opportunities, there are significant weaknesses and threats associated with ChatGPT that could lead to its misuse in language learning.

While current advanced studies have explored the role of ChatGPT in language learning and its implications, none have conducted an exploratory study involving the actual use of ChatGPT for this purpose. Therefore, this study aimed to measure the usability of ChatGPT for formal English language learning as experienced by matriculation college students upon a completion of a series of tasks pertaining to English language learning with ChatGPT.

Table 2: SWOT evaluation associated with ChatGPT

Factors	Internal factors	External factors
Helpful to achieve goals	Strengths:	Opportunities
	Self-improving capability	Facilitating personalised learning
	Providing personalised responses	Facilitate complex learning
	Providing real-time responses	Decreasing teaching workload
Harmful to achieve goals	Weaknesses:	Threats:
	Difficulty in evaluating the quality of responses	Lack of understanding of the context
	The risk of biases and discrimination	Threatening academic integrity
	Lack of higher-order thinking skills	Declining in higher order thinking skills.

Source: Farrokhnia et al. (2023)

RESEARCH METHODOLOGY

This study employed a case study research design and involved 30 matriculation college students in Malaysia. The respondents were recruited through official communication channels of the targeted matriculation colleges, focusing exclusively on students enrolled in matriculation courses in Malaysia. The dissemination post regarding the research activity was accessible via a link prepared by the study. The respondents' personal information, including email and current study program, was collected using an online registration form, ensuring confidentiality to protect participant privacy and anonymity. Prior to registration, participants were provided with detailed information about the research objectives, and the specific tasks they would perform during the research. To ensure privacy and confidentiality, each participant was assigned a random ID number automatically generated in the online registration form. All data were securely stored and accessible only to the study in compliance with ethical guidelines and regulations.

The usability testing was conducted in a controlled online meeting room with adequate capacity for 30 participants. Participants interacted with the ChatGPT tool using their laptops and email accounts, with a stable internet connection ensured to facilitate smooth communication and interaction. The task design aimed to evaluate ChatGPT's usability for formal English language learning, comprising four modules: (1) conversing with ChatGPT on various topics, (2) asking ChatGPT to write paragraphs in different contexts such as formal and informal writing, (3) using ChatGPT to identify, correct, and receive suggestions for grammar mistakes, and (4) practicing vocabulary learning with ChatGPT. Each participant was allocated 1 hour to complete all modules, with 15 minutes dedicated to each module. These tasks were designed to assess a range of language learning skills and understand ChatGPT's usability.

A post-task questionnaire was developed to gather participant feedback on the usability of ChatGPT and their overall experience during the 1-hour interaction. The questionnaire comprised five major sections: (1) demographic information, (2) previous English language proficiency, (3) the "Usefulness, Satisfaction, and Ease of Use" (USE) questionnaire, (4) the "System Usability Scale" (SUS) questionnaire, and (5) satisfaction of ChatGPT performance in English learning tasks of conversation, writing, grammar, and vocabulary. The study employed the USE questionnaire, a common tool for evaluating the usability of a technical system which consists of 30 statements rated on a Likert scale items from 1 (strong disagreement) to 5 (strong agreement). It assesses users' perceptions and experiences of a system's overall usability for specific tasks (Gao et al., 2018). Similarly, the SUS questionnaire, comprising ten statements rated on the same Likert scale items, provides a standardized and reliable measure for assessing the usability and satisfaction of a system as suggested by Gao et al. (2018). To gain maximum insights into the usability of ChatGPT, the study analyzed participants' responses to the SUS and USE questionnaires using descriptive statistics such as.

RESEARCH FINDINGS

This section explains findings of the study that have been organized into profile of the respondents, ChatGPT's performance satisfaction across English learning tasks, usability of ChatGPT and usability of ChatGPT across gender.

PROFILE OF THE RESPONDENTS

Many of the respondents are male (63.3%), with only 36.7% are female and aged 18 years old (50%) and 19 years old (50%). The respondents are in their first (50%) and second (50%) year of study with 80 % of the respondents enrolling in social science programs while 20% of them enrolling in pure science programs. Figure 1 depicted descriptive statistics for four components of English language skills, i.e., speaking, listening, reading, and writing of the 30 research respondents. The mean score of speaking skills is 4.2 that indicates a relatively high level of self-assessed speaking proficiency among respondents. The standard deviation of 0.5 suggests that there is a moderate variability in speaking skills scores. The minimum score of 3 and maximum score of 5 of speaking skills show a range of speaking abilities within the group. The negative skewness (-0.2) indicates a slight skew towards higher scores, while the kurtosis (0.1) suggests a distribution close to normal. On the other hand, the mean score of listening skills is 4.0 which indicates that respondents generally rate their listening skills highly. The standard deviation of 0.6 points to a moderate spread in the data of speaking skills. With scores ranging from 3 to 5, there is some variability in the listening skills of the respondents. The negative skewness (-0.3) indicates a slight skew towards higher scores of listening skills, and the kurtosis (0.2) indicates a relatively normal distribution of listening skills.

Statistics	Speaking	Listening	Reading	Writing
Mean	4.2	4.0	4.1	3.9
SD	0.5	0.6	0.4	0.5
Min	3	3	3	2
Max	5	5	5	5
Skewness	-0.2	-0.3	-0.1	-0.4
Kurtosis	0.1	0.2	0.0	0.3

Note. N = 30; Max = Maximum; Min = Minimum; SD = Standard Deviation

Figure 1: English language skills rating of the research respondents

Figure 1 also indicated that the respondents' mean score of reading skills is 4.1 that reflects a high self-assessment of reading skills among the respondents. The low standard deviation (0.4) suggests that there is less variability in reading skills scores compared to other skills. The scores range for reading skills is from 3 to 5, showing some diversity in reading skills of the respondents. The skewness (-0.1) and kurtosis (0.0) values indicate a nearly normal distribution of scores in reading skills. Lastly, figure 1 also suggested that the respondents' mean score writing skills is 3.9 which is slightly lower than the other skills, suggesting participants perceive their writing skills as slightly less proficient. The standard deviation of 0.5 indicates moderate variability in writing skills score, and the wider range of scores (2 to 5) suggests significant diversity in respondents' perceived writing skills. The negative skewness (-0.4) also shows a skew towards higher scores, and the kurtosis (0.3) suggests a slightly peaked distribution in respondents' perceived writing skills. Overall, participants rate their speaking, listening, and reading skills highly, with writing being slightly lower.

CHATGPT'S PERFORMANCE SATISFACTION ACROSS ENGLISH LEARNING TASKS

Figure 2 also provided a detailed statistical summary of respondents' satisfaction ChatGPT performance in various English learning tasks— conversation, writing, grammar and vocabulary, grammar, and vocabulary. As shown in figure 2, the average satisfaction of ChatGPT performance in conversation tasks is 4.15, indicating a generally high level of satisfaction among participants. With a standard deviation value of 0.49, there is a moderate variability in respondents' satisfaction rating in conversation task. The scores ranged from 3.0 to 5.1, showing that while most respondents rated the conversation tasks highly satisfactory, there were a few lower ratings. The negative skewness (-0.34) and kurtosis (-0.20) values indicate that there is a slight skew towards higher ratings with a flatter distribution than a normal curve.

On the other hand, the satisfaction rating of ChatGPT performance in writing tasks average score is 3.79, slightly lower than satisfaction in conversation tasks. With a value

Statistics	Conversation	Writing	Grammar	Vocabulary
Mean	4.15	3.79	4.03	4.12
SD	0.49	0.55	0.40	0.51
Min	3.0	2.5	3.2	3.0
Max	5.1	4.9	4.8	5.2
Skewness	-0.34	-0.37	-0.13	-0.9
Kurtosis	-0.20	-0.38	-0.65	-0.48

Note. N = 30; Max = Maximum; Min = Minimum; SD = Standard Deviation

Figure 2: Respondents' satisfaction rating of ChatGPT performance in four English language tasks

of 0.55 for standard deviation, there is a moderate variability in satisfaction of ChatGPT performance in writing tasks. The scores ranged from 2.5 to 4.9, showing a wider spread and some lower ratings compared to satisfaction rating in other tasks. The negative skewness (-0.37) and kurtosis (-0.38) values suggest that there is a skew towards higher ratings but with a flatter distribution.

Figure 2 also showed that the average score for satisfaction rating of ChatGPT performance for grammar task is 4.03 that suggests a high satisfaction in the grammar task. With the value of 0.40 for standard deviation, respondents' satisfaction on grammar tasks shows less variability compared to other tasks. Grammar tasks' scores

ranged from 3.2 to 4.8, indicating a consistently high rating. The minimal skewness (-0.13) and moderate kurtosis (-0.65) suggests that a normal distribution of satisfaction ratings on grammar tasks. Lastly, as shown in figure 2, the average score for satisfaction rating of ChatGPT performance for vocabulary tasks is 4.12, close to the satisfaction rating for conversation tasks. With a value of 0.51 of standard deviation, there is a moderate variability in satisfaction rating scores for vocabulary tasks. The satisfaction scores ranged from 3.0 to 5.2, very similar to conversation tasks' satisfaction rating with some extreme high ratings. Higher negative skewness (-0.9) and moderate kurtosis (-0.48), indicating a pronounced skew towards higher ratings.

THE USABILITY OF CHATGPT

Data on the usability of ChatGPT was analyzed to understand the system's performance across five dimensions of system usability, usefulness, ease of use, ease of learning, and satisfaction. Figure 3 illustrates descriptive statistics of the five dimensions of ChatGPT usability which include the mean, standard deviation (SD), minimum (Min), maximum (Max), skewness, and kurtosis values for each of these dimensions.

Statistics	System Usability	Usefulness	Ease of Use	Ease of Learning	Satisfaction
Mean	40.53	42.01	39.95	42.81	41.26
SD	0.49	0.41	0.59	0.48	0.41
Min	3.0	3.2	2.9	3.1	3.1
Max	5.1	5.0	5.2	5.3	5.0
Skewness	-0.34	-0.35	-0.29	-0.27	-0.30
Kurtosis	-0.20	-0.31	-0.19	-0.23	-0.28

Note. N = 30; Max = Maximum; Min = Minimum; SD = Standard Deviation

Figure 3: The usability of ChatGPT across system usability, usefulness, ease of use, ease of learning, and satisfaction

As illustrated in Figure 3, the average system usability score of 40.53 out of 50 indicates a relatively high usability rating from the respondents. The standard deviation is low (0.49), suggesting that the responses were closely clustered around the mean. The skewness and kurtosis values indicate a slight leftward skew and a relatively flat distribution respectively. The mean score for usefulness is 42.01 out of 50, reflecting a positive perception of ChatGPT's usefulness among the respondents. The standard deviation of 0.41 implies that there was some variability in responses, but most were near the average. The negative skewness suggests that more responses were above the mean, while the kurtosis value indicates a relatively normal distribution.

Figure 3 also showed that the ease of use received a high average score of 39.95 out of 50, signifying that the respondents found the system is easy to use. The standard deviation is 0.59, indicating some spread in the responses. The skewness and kurtosis values suggest a slightly negatively skewed and relatively normal distribution. The ease of learning aspect had an average score of 42.81 out of 50, indicating that participants generally found ChatGPT easy to learn. The standard deviation of 0.48 shows moderate variability in responses. The skewness and kurtosis values suggest a slight negative skew and a relatively normal distribution. The mean satisfaction score of 41.26 out of 50 indicates that participants were generally satisfied with ChatGPT. The standard deviation is 0.41, reflecting some variation in the responses. The skewness and kurtosis values indicate a slightly negatively skewed distribution and a relatively normal distribution. The descriptive statistics highlight the overall positive feedback from participants regarding the usability of ChatGPT across various dimensions. The mean scores for all aspects are relatively high, and the standard deviations are low, suggesting that most participants had similar positive experiences. The skewness and kurtosis values indicate that the distributions of responses are generally normal, with slight negative skews. These results suggest that ChatGPT is a useful, easy-to-use, and satisfying tool for English language learning, and that participants found it easy to learn how to use the system effectively.

THE USABILITY OF CHATGPT ACROSS GENDER

Figure 4 shows the distribution of responses for five dimensions of ChatGPT usability, i.e. system usability, usefulness, ease of use, ease of learning, and satisfaction across gender. As shown in figure 4, the system usability scores for both genders are very close, with male respondents' rating is slightly higher (40.63) compared to

females (40.36). This minor difference suggests that both genders have a similar perception on the system's usability, indicating a general positive perception on ChatGPT's interface and functionality in assisting English language learning. Similarly in terms of usefulness, male respondents provide slightly higher ratings (42.84) compared to female respondents (42.18). The small difference implies that both genders find ChatGPT to be a useful tool for English language learning, although males perceive a marginally higher utility.

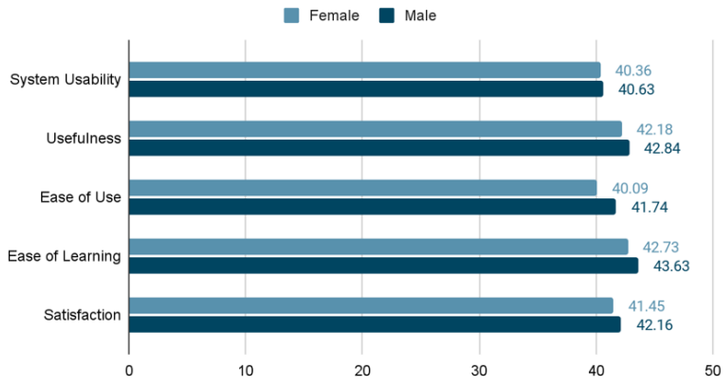

Figure 4: Dimensions of ChatGPT usability across gender

Figure 4 shows a more noticeable difference in the ease-of-use scores; with male respondents' rating it higher (41.74) than females (40.09). This suggests that male respondents might find ChatGPT slightly more user-friendly. However, the overall high scores indicate that the system is generally considered easy to use by both genders. For ease of learning, male respondents rated ChatGPT higher (43.63) compared to females (42.73). This difference, while minor, indicates that males might find it slightly easier to learn how to use the system. The high scores from both genders reflect the system's intuitive design and effective user guidance. Lastly the satisfaction scores are also higher for males (42.16) compared to females (41.45). This suggests that males are slightly more satisfied with the system's performance. Nevertheless, both genders exhibit high satisfaction levels, demonstrating overall approval of ChatGPT's capabilities for English language learning.

THE USABILITY OF CHATGPT ACROSS AGE

Figure 5 presents the distribution of responses across five dimensions of ChatGPT of system usability, usefulness, ease of use, ease of learning, and satisfaction across

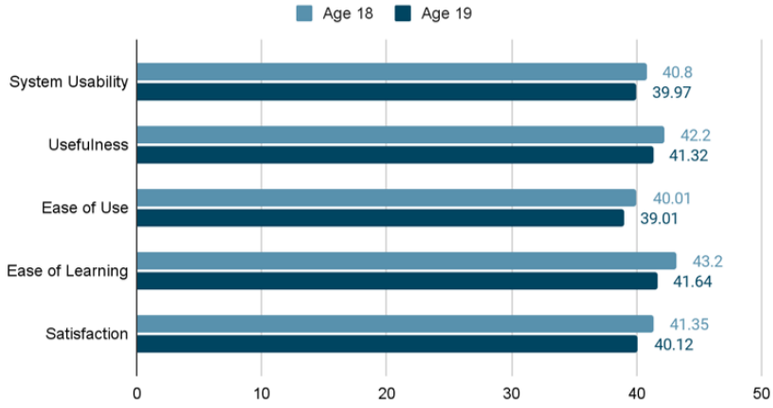

Figure 5: Usability of ChatGPT across age

respondents' age. The system usability scores indicate a slight preference among 18-year-olds (40.8) compared to 19-year-olds (39.97). The proximity of the scores suggests that both age groups find the system generally usable, with minor differences in perception. In terms of usefulness, participants aged 18 rated ChatGPT higher (42.2) than those aged 19 (41.32). This indicates that the younger age group perceives the system as slightly more beneficial for their English language learning needs.

Figure 5 also illustrates that ease-of-use scores show that 18-year-olds find ChatGPT marginally easier to use (40.01) compared to 19-year-olds (39.01). While the difference is small, it reflects a consistent trend of younger participants rating the system more favorably. For ease of learning, the scores again favor the 18-year-old group (43.2) over the 19-year-old group (41.64). This suggests that younger participants find it easier to learn how to use the system effectively, which could be attributed to their adaptability or familiarity with similar technologies. Satisfaction ratings are higher among 18-year-olds (41.35) compared to 19-year-olds (40.12). This indicates a slightly greater overall satisfaction with ChatGPT among the younger participants, reinforcing the trend observed across other metrics. The age-group-based analysis of SUS and USE questionnaire responses reveals that 18-year-old participants generally rate ChatGPT higher across all five aspects compared to their 19-year-old counterparts. These findings suggest that the usability, usefulness, ease of use, ease of learning, and satisfaction with ChatGPT are perceived more positively by the younger age group. This trend may highlight differences in user experience and adaptability between the two age groups, providing valuable insights for optimizing ChatGPT's usability for diverse age demographics.

DISCUSSION AND CONLUSION

This study aimed to evaluate the usability and effectiveness of ChatGPT as a tool for English language learning among pre-university students in Malaysia. The analysis of the System Usability Scale (SUS) and the Usefulness, Satisfaction, and Ease of Use (USE) questionnaires provided insights into students' English language learning experience. The findings showed that there is a high level of satisfaction of ChatGPT performance in all English language learning tasks of conversation, writing, grammar and vocabulary, grammar, and vocabulary with highest satisfaction in conversation and lowest satisfaction in writing tasks. The findings of this study are in line with the existing literature on the usability and effectiveness of AI tools in education. For instance, Chen & Chen (2021) and Liu et al. (2021) found that AI tools enhance user satisfaction and are beneficial for educational purposes, which corroborates the positive feedback received in this study.

Similarly, the findings indicated there is a positive perception of ChatGPT's usability among students in higher education across system usability, usefulness, ease of use, ease of learning, and satisfaction of ChatGPT in English language learning. This finding is consistent with previous research by Chen and Chen (2021) which highlighted the ease of use and user satisfaction associated with AI tools in educational settings. The findings further suggested that the male respondents rate all five dimensions of ChatGPT slightly higher than female respondents, indicating that male respondents showed a higher overall approval in all five dimensions of ChatGPT usability. This finding is aligned with findings by Venkatesh and Morris's (2000) that highlight gender differences in technology adoption, and thus highlights the need to address gender-specific usability concerns in using AI tools such as ChatGPT in English language learning. Additionally, the study found that younger students have rated ChatGPT usability higher compared to older students suggesting that younger users might adapt more easily to new technologies which is consistent with Prensky (2001) concept of digital natives.

In conclusion, this study also contributes new insights by specifically addressing the use of ChatGPT for English language learning among pre-university students in Malaysia. The nuanced understanding of demographic variations underscores the need for tailored strategies to maximize the benefits of AI tools in diverse educational settings. Future research should explore the underlying reasons for these demographic differences and develop strategies to address them, ensuring equitable benefits from technological advancements in language learning. The findings of this study confirm that ChatGPT is a highly usable and effective tool for English language learning, with demographic factors such as gender might influence user perceptions. These insights provide a foundation for future research and practical applications aimed at enhancing the implementation of AI tools in education, catering to the diverse needs of learners.

REFERENCES

Alenizi, M. A. K., Mohamed, A. M., & Shaaban, T. S. (2023). It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education. *Innoeduca. International Journal of Technology and Educational Innovation* 9(2): 5-23. https://doi.org/10.1186/s41239-023-00425-2

Akcora, D.E., A. Belli, M. Berardi, S. Casola, N. Di Blas, S. Falletta, A. Faraotti, L. Lodi, D.N. Diaz, P. Paolini,

F. Renzi and F. Vannella. 2018. Conversational Support for Education. In Penstein Rosé, C., Martínez-Maldonado, R., Hoppe, H.U., Luckin, R., Mavrikis, M., Porayska-Pomsta, K., McLaren, B., & du Boulay, B. (eds.) *Artificial Intelligence in Education, Lecture Notes in Computer Science*, pp. 14–19. Springer International Publishing: Cham

Becher, T., & Trowler, P.R. 2001. *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. McGraw-Hill Education (UK).

Bender, S.M. 2024. Awareness of Artificial Intelligence as an Essential Digital Literacy: ChatGPT and Gen-AI in the Classroom. *Changing English*. Advance online publication. https://doi.org/10.1080/1358684X.2024.2309995

Brown, P., & Stent, A. 2021. ChatGPT: Large-scale generative pre-training for conversation. *arXiv preprint arXiv:2101.03467*.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. & Amodei, D. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33: 1877-1901.

Chen, J. 2011. The effects of education compatibility and technological expectancy on e-learning acceptance. Computer & Education 57(2): 1501-1511

Chen, J., & Chen, L. 2021. Personalized learning with AI: A case study of ChatGPT in ESL education. *Journal of Educational Technology Development and Exchange* 14(2): 110-125.

Chen, H.-L., Vicki Widarso, G., & Sutrisno, H. 2020. A ChatBot for Learning Chinese: Learning Achievement and Technology Acceptance. *Journal of Educational Computing Research* 58(6): 1161-1189. https://doi.org/10.1177/0735633120929622

Du, J., & Daniel, B. K. 2024. A Systematic Review of AI-Powered Chatbots in EFL Speaking Practice: Transforming Language Education. *Computers and Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2024.100230

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

Fitria, T.N. 2023. The Impact of Conversational AI on Language Learning: A Case Study of ChatGPT. *Journal of Interactive Learning Research* 34(1): 120-145.

Garzon, J., Lampropoulos, G. & Burgos, D. 2023. Effects of Mobile Learning in English Language Learning: A Meta-Analysis and Research Synthesis. *Continuing and Emerging Research Trends in Technology-Enhanced Learning* (Special Issue). Available online at https://www.mdpi.com/2079-9292/12/7/1595

Godwin-Jones, R. 2022. Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning & Technology 26(2): 5–24. http://doi.org/10125/73474

Govindarajan, R., & Christuraj, A. 2023. Integrating technology in PBL: A pathway to effective learning. J*ournal of Educational Research and Technology* 19(4): 200-215.

Govindarajan, R., & Christuraj, G. 2023. Opportunities and challenges of using ChatGPT in the ELT scenario of UTAS, Nizwa, Oman. *The Journal of Teaching English for Specific and Academic Purposes* 11(3): 593-605. https://doi.org/10.22190/JTESAP230529046G

Haristiani, D. 2019. The Use of Speech Recognition Technology in EFL Classes: An Overview. *TESOL Journal* 10(2): 345-360.

Hieu, H.H. & Thao, L.T. 2024. Exploring the Impact of AI in Language Education: Vietnamese EFL Teachers' Views on Using ChatGPT for Fairy Tale Retelling Tasks. *International Journal of Learning, Teaching and Educational Research* 23(3): 486-503. https://doi.org/10.26803/ijlter.23.3.24

Hobert, S. and F. Berens. 2020. Small Talk Conversations and the Long-Term Use of Chatbots in Educational Settings – Experiences from a Field Study. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11970 LNCS, 260–272.

Huang, L., & Huang, S. 2022. Gamification in Language Learning: A Review of Empirical Studies. *Journal of Educational Technology* 29(3): 75-89.

Jeon, M. 2021. Speech Recognition Technology in Language Learning: A Review of Current Trends. *Language Learning & Technology* 25(2): 12-25.

Jeon, J.; Lee, S. & Choe, H. 2023. Beyond ChatGPT: A Conceptual Framework and Systematic Review of Speech-Recognition Chatbots for Language Learning. *Computers &Education* 206: 104898

Jia, J., Chen, Y., Ding, Z. & Ruan, M. 2012. Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education* 58(1): 1315

Klimova, B. F. 2011. Evaluating writing in English as a second Language. *Procedia Social and Behavioral Sciences* 28 (2011): 390-394

Klimova, B., & Zaborova, E. 2023. Augmented Reality in Language Learning: A Review of Applications and Efficacy. *Journal of Educational Computing Research* 58(4): 987-1012.

Liu, R., Zhang, J., Wang, L., & Shi, X. 2021. Artificial intelligence applications in education: A systematic literature review and future research directions. *Computers & Education* 168: 104180.

Liu, G. L., Darvin, R., & Ma, C. 2024. Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2024.2310288

Luo, Z. The Effectiveness of Gamified Tools for Foreign Language Learning (FLL): A Systematic Review. *Behav. Sci. 2023* 13: 331. https://doi.org/10.3390/bs13040331

Meniado, J.C. 2023. The Impact of ChatGPT on English Language Teaching, Learning, and Assessment: A Rapid Review of Literature. *Arab World English Journal* 14(4): 3-18. https://dx.doi.org/10.24093/awej/vol14no4.1

Okonkwo, C.W. & Ade-Ibijola,A. 2021. Chatbots applications in education: A systematic review, *Computers and Education: Artificial Intelligence* 2(2021): 100033,

Peterson, P. L., & Peterson, M. W. 2019. Usability and Effectiveness of Augmented Reality Tools in Language Learning. *International Journal of Educational Technology* 38(3): 210-225.

Prathyusha, N. 2020. Role of gamification in language learning. *International Journal of Research and Analytical Reviews* 7(2): 577-583

Presky, M. 2001. Digital Natives, Digital Immigrants. *In On the Horizon,* Vol 9 (5). MCB University Press. Available online at https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf

Ren, R., Castro, J. W., Acuña, S. T., & de Lara, J. 2019. Evaluation Techniques for Chatbot Usability: A Systematic Mapping Study. *International Journal of Software Engineering and Knowledge Engineering* 29(11):1673–1702. doi:10.1142/S021819401940016

Song, J., & Song, M. 2022. The Role of AI in Enhancing Language Skills in ESL Learners. *Journal of Language Education Research* 15(2): 120-135.

Venkatesh, V., & Morris, M. G. 2000. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 24(1): 115-139.

Wang, C., & Wang, S. 2020. Teachers' perceptions and experiences of AI applications in education. *Educational Technology Research and Development* 68(5): 2469-2493.

Yang, H., Kim, H., Lee, J. & Shin, D. 2022. Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL* 34(3): 327–343. https://doi.org/10.1017/S0958344022000039

Zadorozhnyy, A., & Lai, W. Y. W. 2024. ChatGPT and L2 Written Communication: A Game-Changer or Just Another Tool? *Languages* 9(1): 5 https://doi.org/10.3390/languages9010005

Nurafiqah Syuhada Abdul Rahim
Fakulti Pendidikan
Universiti Kebangsaan Malaysia

Aidah Abdul Karim, PhD
Fakulti Pendidikan
Universiti Kebangsaan Malaysia