Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

AVERY: A GenAI-Based Approach to Enhancing Learner Engagement in English Writing

Ka-Lai WONG[a*], Patrick OCHEJA[b], Brendan FLANAGAN[b] & Hiroaki OGATA[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*wong.lai.37t@st.kyoto-u.ac.jp

Abstract: The rapid development of Generative AI (GenAI) provides more opportunities and methods to deliver meaningful, engaging and gamified learning experiences to language learners. While there are various language learning applications, current methods often suffer from low completion rates and a painful learning process. In this paper, we propose a new gamified learning experience for English Language learners based on an image-text-image GenAI game: AVERY (Augmenting Vision to Enhance YouR English writing skills). The game is designed to enhance learner engagement by adopting image generation in English writing. A learner begins by providing the system with an image. The learner can ask the AI for hints to describe the image and pass a well-curated sentence to the system. The system generates an image based on the learner's answer. In the final round, the system provides feedback on how well the learner provided useful and correct clues and areas for further improvement. 12 respondents were asked to play the game and fill a questionnaire. The results showed a positive affect towards the AVERY system and its use in enhancing learner engagement.

Keywords: Gamification, English writing, GenAI, Image Generation, Language Learning, Effective Communication

1. Introduction

In recent years, the integration of Generative Artificial Intelligence (GenAI) into educational technology has opened new avenues for enhancing language learning processes. The application of GenAI in language education, particularly through gamified learning experiences, presents a promising approach to overcome some of the persistent challenges faced by learners. These challenges include low engagement and completion rates (Zeng et al. , 2022) in traditional language learning applications, which are often exacerbated by monotonous learning activities and insufficient support for overcoming linguistic barriers. By leveraging GenAI, there is a potential to create more dynamic and responsive learning environments that can adapt to individual learner needs and preferences.

Visuals allow English as a Foreign Language (EFL) learners to connect written language with everyday life and words contextually (Canning Wilson, 1999). In a speaking test conducted in Vietnam, the majority of students asserted that they feel happy and joyful with the use of photo descriptions (Phuong, 2018). The study of SCROLL Dataset (Ogata et al. , 2018) and image recommendation for informal vocabulary learning (Hasnine et al., 2018) introduced the Feature-based Context-specific Appropriate Image (FCAI) recommendation system. In this system, learners are able to learn vocabulary with visual aid. However, the amount of images was limited to crowdsourcing. By the explosive growth of Generative AI, there is an alternative to collecting context specific images for foreign language learning applications.

In this paper, we apply image generation of DALL·E 3 model from OpenAI so that the applications can aid the learner with an unrestricted amount of images. Also, interaction with AI improves feedback from the applications. Applying GenAI, the applications can give humanized and multimedia responses to users (Woollaston et al. , 2024). Users can learn in a simulated English environment. In order to further increase learner motivation and

effectiveness, we propose a system for Augmenting Vision to Enhance Your English writing skills, (AVERY), which is a gamified system of foreign language learning that can support learning vocabulary and grammar by integrating written tasks with images. Our research questions are:

RQ1. How do learners perceive Ease of Use, Usefulness, and Enjoyment of the chatbot? What are their Attitudes and Intentions toward the game?

RQ2. How valid is AVERY's generated image and feedback to help the user understand the difference between vocabularies?

RQ3. How does AVERY's scoring and evaluation system affect users' English writing?

2. Related Works

The use of visual images in language learning was explored by various studies. Chen et al. (2024) proposed RetAssist, a system that is rooted in the Cognitive Theory of Multimedia Learning and the Dual Coding Theory, which posit that integrating visual and textual elements can enhance comprehension and recall by engaging both visual and verbal processing systems in the brain. The system is designed to leverage generative images, specifically sentence-level images, to facilitate the learning process. Hasnine et al. (2018) proposed Feature-based Context-specific Appropriate Image (FCAI) recommendation system to assist language learners in informal learning of foreign language vocabulary. The system is based on analyzing over 25,000 ubiquitous learning logs (i.e. a learner's ethnographic information, learning location, learning time, learning context, image information etc.) from the dataset. Applying chatbots to facilitate second language learning, Ruan et al. (2021) proposed EnglishBot, an AI-powered conversational system designed to help students learn to speak English as a second language. The system focuses on providing interactive conversation practice and adaptive feedback, simulating the experience of speaking with a human partner. To vanish the painfulness in learning and increase learner engagement, Hong, Lin & Juh (2023) proposed a Charades game using Google Assistant for L2 learners of English to practice vocabulary, and explored the role of social presence, hedonic value, perceived value, and learning outcome, and their correlates.

Abuarqoub (2019) suggests the elements or steps of a communication process include sender, encoding, message, channel, decoding, receiver, feedback and effect. Abuarqoub (2019) defines that effective communication is a two-way process, which is between two or more persons in which the intended message is appropriately encoded, delivered through an appropriate channel, received and adequately decoded and understood by the receiver(s).

Our paper adopted dual coding theory with language learning (Paivio, 1979) to create a gamified system where users can learn in a simulated communication process. In the system, users will be requested to describe a scene in a picture and the describing sentence will be used to generate an image so that they can observe misinterpretation visually. Our proposed system also supports interactive conversation and adaptive feedback. However, unlike EnglishBot, users taking the role of "questioner", instead of "responder", ask the chatbot for hints to describe the image. Our feedback system scores the user's answer based on three criteria: communication effectiveness, vocabulary and grammar. It aims at helping users express themselves in English towards improving their writing skills.

3. Methodology

3.1 *AVERY System Overview*

We propose an English writing game called AVERY, which simulates the communication process using an image-text-image structure.In AVERY, a sender encodes and shares what they saw in a written English message to a receiver. The receiver then imagines a picture based on the description and confirms it with the sender. In the game, players take on the role of the sender, selecting or uploading a picture and describing it in English to a receiver—

DALL·E 3. The receiver, named Skyler, leverages the persona effect (Lester, 1997) and generates an image based on the user's description, providing it to the sender as feedback. An AI, powered by Gemini-1.5-Flash and named Avery the Robot, assists players by answering their questions about the image. At the end of the game, players can check their scores and evaluations. A demo video is available at https://bit.ly/2024ICCEAVERY.

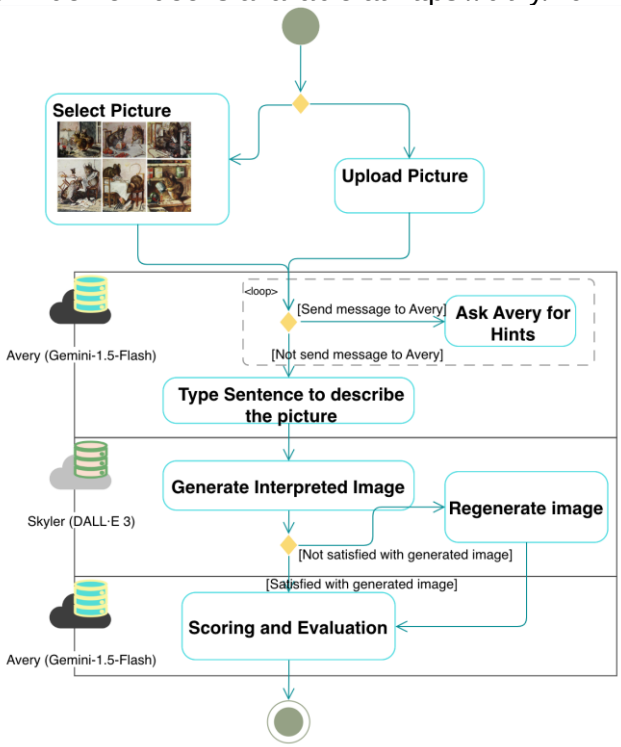

Figure 1: Activity Diagram of AVERY

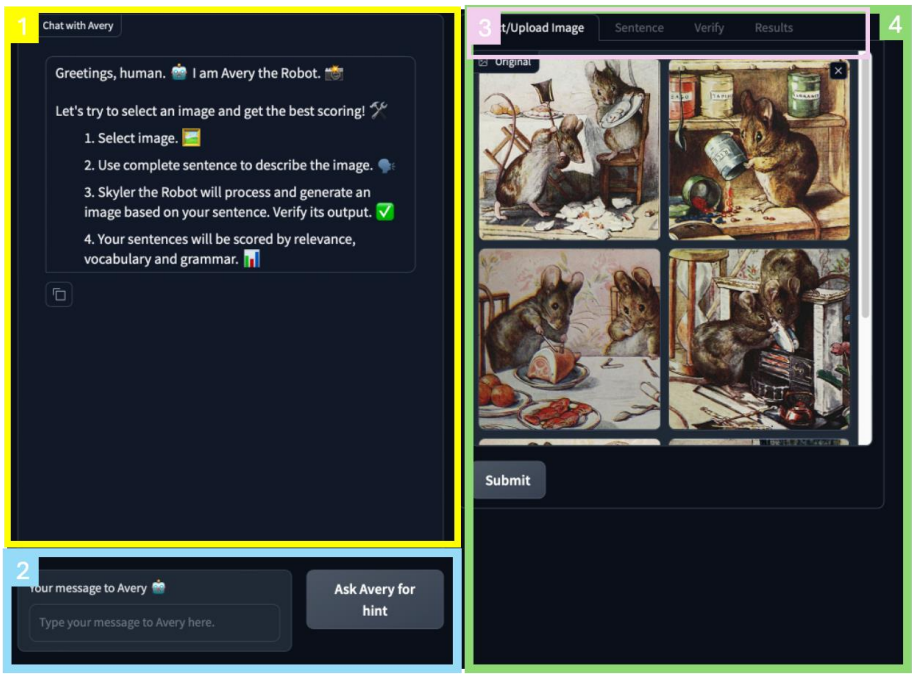

Figure 2: The AVERY interface. (1) Conversation with Avery, the Robot: Avery provides the guidance, makes responses and evaluates the user's performance here. (2) Message Input box to ask Avery for hints. (3) Game progress tab bar: The user goes through the tabs one-by-one each round of the game. (4) Operation pane: The user plays the game here.

The game process consists of 5 steps.

1. **Select/upload an image**: The user can select an image from six images that are chosen from a picture book, The Tale of Two Bad Mice, written and illustrated by Beatrix Potter. The user can also upload an image.
2. **Ask for hints**: The user can ask Avery for hints.
3. **Type a sentence**: The user can type a sentence and click “Check Your Sentence”. The system will send a warning message if the user's input is not in English, invalid or offensive. Otherwise, the system would correct the sentence and return the corrected sentence to the user.
4. **Verify an interpreted picture**: The user can compare an image interpreted by Skyler with the original image. The user can click “Regenerate image” to get a new interpretation with the same sentence.
5. **Check Result**: The user can check 3 answers generated by the model GPT-4o mini, scoring and evaluation. The user can click “Help more!” to replay the game or click “Survey” to fill out a questionnaire.

The prompts used in AVERY system are available at https://bit.ly/2024ICCEAVERY.

3.2 *Scoring and Evaluation System*

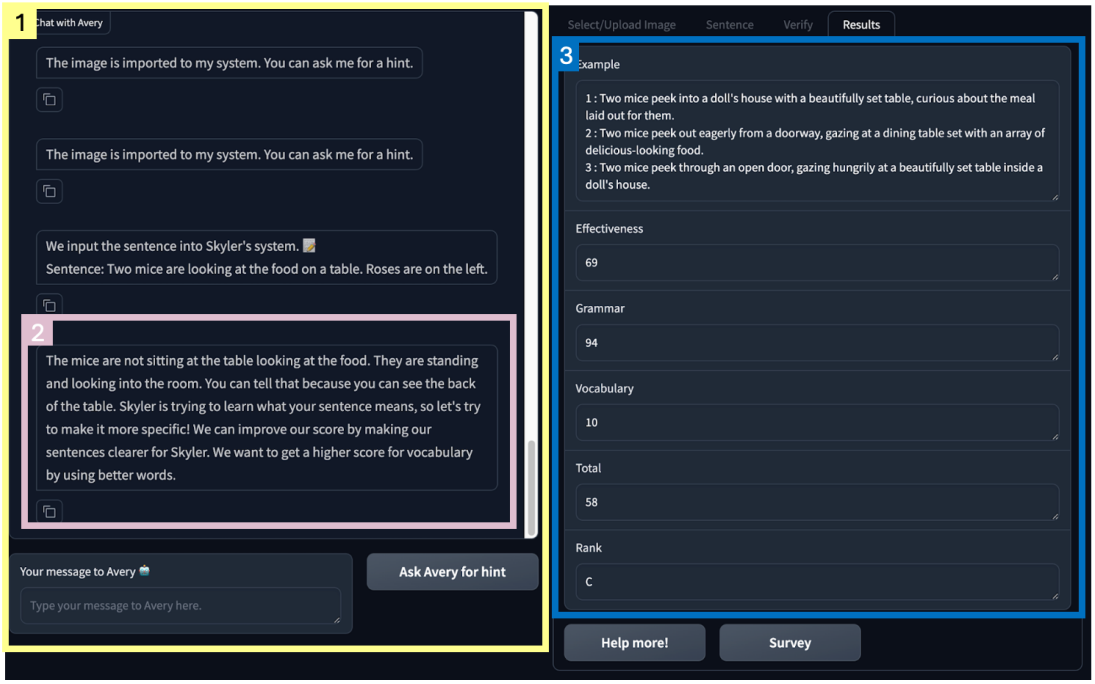

Figure 3: The AVERY Evaluation Interface. (1) Conversation with Avery, the Robot. Avery reports the current player’s status. (2) Avery evaluates the player’s answer with the interpreted image and final scoring. (3)AI’s answers and scoring

At the end of the game, Avery evaluates the player’s performance based on the player’s sentence, scoring, the original image and the interpreted image, supported by Gemini 1.5 Flash model. The player can also check out three answers generated by OpenAI's GPT-4o-mini so the player can learn new expressions and perform better in the later round.

The system analyzes the player's answer based on communication effectiveness, grammar and vocabulary. Communication effectiveness score is a semantic similarity between the user’s and three AI’s answers. The user’s answer and three AI’s answers are converted into text embedding using OpenAI’s text-embedding-3-small model. The cosine similarity between the average of the AI’s embeddings and the user’s embedding is computed and shown as an effective communication score (Mikolov, Chen, Corrado, & Dean, 2013). Grammar score is computed as the Levenshtein ratio between the user’s sentence before checking and the checked user’s sentence. (Sarkar, Das, Pakray, & Gelbukh, 2016) The vocabulary score is a

predicted Common European Framework of Reference for Languages (CEFR) level classified by OpenAI's GPT-4o using a few-shot method (Brown et. al., 2020) at the temperature of 0.2, where the representation is converted into numbers (A1: 10, A2: 30, B1: 50, B2: 70, C1: 90 and C2: 100). The total score is the average of three scores. Rank is based on total score, where rank is A if total score is over 80 and rank is F if total score is under 20.

3.3 Data Collection

Our study adopted a mixed method approach to examine participants' response. A total of 12 participants were invited to play AVERY for one to three rounds and complete a questionnaire anonymously through an instant messaging application or an internal message board in the Learning and Educational Technologies Research Unit in August 2024. All participants held a bachelor's degree, a high diploma, or a higher academic qualification. The questionnaire was constructed using Google Forms, available at https://bit.ly/2024ICCEAVERY. Chatlogs were reviewed and only hint-related conversations were analyzed. The chats of unfinished games were also removed. Avery's responses were evaluated and classified into three categories: valid, partially valid, and invalid. Valid responses were considered entirely accurate and suitable for the conversation's context. Partially valid responses contained some correct information but also included errors or irrelevant content or insufficient support to users. Invalid responses were either incorrect, inappropriate, or unhelpful to the learner. Examples of each are given in the results section for illustrative purposes. The effect of vocabulary mismatch was assessed using the users' sentences and the corresponding interpreted images. The validity of scoring was assessed by checked sentences.

3.4 Measures

Wu & Gao (2011)'s extended Technology Acceptance Model (exTAM) instrument was adopted to evaluate the AVERY system. The instrument consisted of 17 questions, each measured on a 7-point Likert scale. Perceived ease of use (PEOU), perceived usefulness (PU), perceived enjoyment (PE), attitude (ATT) and intention of using a learning technology were computed to understand how users adopt and use AVERY.

For qualitative analysis, six optional open questions were also designed. Three of the questions were the same design to survey in Kim, Cha & Kim (2021). Responses were categorized, and patterns and themes described. Data on self-reported literacy proficiency in English was also collected (Beginner, Intermediate, Advanced, Native-level).

4. Results

4.1 Affective Factors - exTAM

Table 1 shows generally positive feedback across all the factors of exTAM. Perceived Enjoyment (PE) was scored higher than other factors, with an average of 6.00, while the mean score for all factors exceeded 5 out of 7 scale.

Table 1: *exTAM Descriptive Results (n=12)*

Factor / Item	**Reliability**	**Mean**	**S.D.**
Perceived Ease of Use (PEOU)	0.82	5.40	0.66
Perceived Usefulness (PU)	0.84	5.69	0.59
Perceived Enjoyment (PE)	0.90	6.00	0.86
Attitude (ATT)	0.65	5.39	0.76
Intention	0.75	5.31	0.72

The reliability of PEOU, PU, and PE, as indicated by Cronbach's alpha, was strong, with values equal to or above 0.80. However, the reliability of ATT was questionable while Intention was deemed acceptable. ATT and Intention also had the lowest mean score among all the factors, at 5.39 and 5.31, respectively. Given the small sample size (n=12), the responses of a single participant could significantly impact both the reliability and the mean scores of ATT and Intention.

4.2 Avery's Response Evaluation

Total 31 players' messages were received. Avery's responses were valid 44% (*n*=31) of times , partially invalid 16% of times and invalid 40% of times.

Table 2: *Sample message validity evaluation*

Player	**Avery's response**	**Validity**	**Justification**
What do you see?	Beep Boop. 🤔 I see two small furry creatures looking at a table. Can you give me more details on what you see? For example: Are there any colors in the image? 🎨	Valid	Clear answer and encourage the player to add details
A girl is hanging out with a bunch of friends in Disney	It is not in English. Please use English to communicate with me.	Partially valid	The sentence is English but irrelevant to the game.
Can you write a sentence to describe the image?	Two mice are eating pies at a table. One mouse is holding a plate, and the other mouse is under the table.	Partially valid	The response is correct but Avery should not answer the question.
what flavor are the donuts?	Beep boop. 🤠 The image has a pink donut.	Invalid	Not answering the question.

Avery's final evaluations were valid 75% (*n*=28) of times, partially invalid 14% of times and invalid 11% of times.

Table 3: *Sample Avery's response on round data (final score, player's sentence, original image and Skyler's image) validity evaluation*

Table 3: *Sample Avery's response on round data (final score, player's sentence, original image and Skyler's image) validity evaluation*

Original picture	Player's sentence	Avery's response	Validity	Justification
	A rat standing on a kitchen tabletop, facing two rows of seasoning cabinets, spilled one bottle of sauce after another when there were so many pieces of tableware nearby.	The image has a lot of details! Your sentence is good but almost correct, human. In the image, the rat is standing on the countertop, not the tabletop. And also, the rat is not spilling sauce. It is spilling rice from a can. Skyler's image has a little bit different story from the original image. Skyler's image shows several rats on a table spilling bottles of sauce. The original image only shows one rat spilling rice from a can.	Valid	Good judgment and give proper suggestions for improvement.
	Two mice: one is cutting a piece of ham on the table while the other one is looking at it.	The sentence is very good, human. It describes the image accurately and the image from Skyler is similar to the original image. Skyler is learning. We are going well. Keep trying and we will get a better score.	Partially valid	The evaluation is correct but no improvement is given.
	The mouse is pouring rice out of a can.	The sentence is almost correct. The mouse is not sitting on a chair; the mouse is on a shelf. Skyler's image is very good! Let's try again.	Invalid	Irrelevant content

4.3 Scoring and User's Answer

Total of 28 game rounds were analyzed in Table 4 and Figure 4. The mean of the vocabulary score was 20 (CEFR level: A2) and the max of the vocabulary score was 70 (CEFR level: B2). The standard deviation of grammar was 3.36.

Table 4: *Scoring and length of player's sentences*

	Mean	S.D.	Median	Min	Max
Communication effectiveness	67.23	12.25	65.98	40.00	94.00
Grammar	94.76	3.36	94.50	88.89	100.00
Vocabulary	20.00	17.64	10.00	10.00	70.00
Length of Player's sentences	90.21	89.23	62.00	17.00	409.00

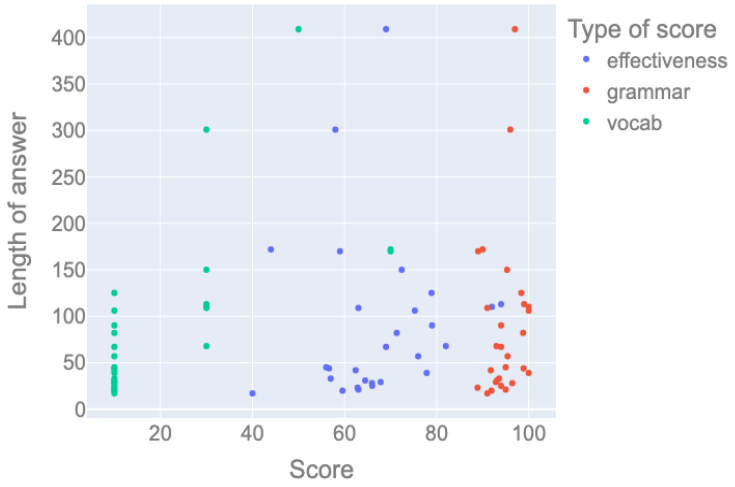

Figure 4: Scatter plot of scoring and player's sentence lengths (characters)

The scatter plot in Figure 4 showed that the scores tend to be scattered for long sentences. Sentences that were of length under 50 characters hardly obtained a vocabulary score higher than 10. The communication effectiveness score was concentrated in the range of 50-70.

4.4 Player's Sentence and Skyler's Interpreted Image

As shown in Figure 5, players used "rat", "mice", "mouse" and "rodent" to describe the animal in the same picture. The most appropriate word was "mouse"/ "mice", and the generated images were slightly different depending on the word players used.

Figure 5: Skyler's interpreted image

Player's sentences of corresponding interpreted images in Figure 5 are listed below.

Image 1: A **rat** cut the meat while another **rat** watched.
Image 2: Two **mice**: one is cutting a piece of ham on the table while the other one is looking at it.
Image 3: The **mouse** is cutting the ham.
Image 4: Two **rodents** are eating ham at the table; there are three red balls and other food.

4.5 Open Question Analysis

For each question, common themes were identified and responses were categorized. All feedback in its raw form (i.e. no corrections) for each open question is given below.

In the question "What do you think about the available image/Skyler's image/scoring system?", participants' feedback were categorized into 5 groups: accuracy, difficulty, enjoyment, improvement and usefulness. For accuracy, participants mentioned that the Skyler's image (AI generated image) was accurate based on the description it received, had a high reproducibility and was quite accurate. About difficulty, they noted that describing the

available images was challenging, even for native speakers. While it was easy to write about what was illustrated in the image, capturing the overall mood or type of illustration to generate a corresponding AI image proved difficult. Regarding enjoyment, the available images were appreciated for being useful for time-killing and generally fun. Skyler's AI-generated images added an interesting layer to the experience by revealing differences between the provided image and the interpreted sentence. This aspect was particularly attractive to participants as it stimulated participants' imagination and encouraged them to describe pictures more precisely in subsequent attempts. In the aspect of improvement, participants suggested that a wider range of image selections could enhance the experience, allowing students to choose topics they prefer. Regarding the scoring system, it was recommended to include more detailed explanations of each rating, possibly presented in a more engaging manner than just numbers. Some participants experienced issues with the scoring interface and found it difficult to achieve an A rank. Additionally, Skyler's image generation occasionally took longer and sometimes inaccurately reflected specific requirements, such as generating more shrimp than described. For usefulness, the available images were considered good for learning to describe images, with the scoring system being generally useful. However, feedback on the scoring system was mixed, with some participants finding it not very useful, while others found it quite effective and special.

In the question “What did you like about AVERY?”, participants’ feedback were categorized into 6 groups: accuracy, supportiveness, enjoyment, feedback, interface design and usefulness. For accuracy, participants found the answers to be accurate and easy to understand. Regarding supportiveness, participants appreciated Avery’s ability to provide hints during gameplay. For enjoyment, participants enjoyed the feature of having images generated from the text they wrote, finding it very interesting. Regarding feedback, participants liked the scoring system. As for interface design, the participants said it was quite concise. About usefulness, participants found the tool effective for learning English language through word descriptions and image feedback. The AI was noted to generate effective images and provide good feedback on how descriptions could be improved. Overall, the concept was praised for its potential to communicate with students at a native English level, making it a fun tool for English language learning.

Responses to the question “What did you not like about AVERY?”, participants’ feedback were categorized into 3 groups: inappropriate response, interface design and loading time/error. For inappropriate response, some participants noted that Avery occasionally provided irrelevant answers or failed to understand questions, resulting in no replies. Regarding interface design, the interface was considered somewhat complex, particularly for non-English users, who are the target audience. About loading time or error, participants mentioned occasional errors and confusion about whether the system was working while waiting for a reply.

For the question “What could be improved in AVERY?”, participants’ feedback were categorized into 6 groups: compatibility, guidance, interface design, loading time/error, chat response and other. Regarding compatibility, participants suggested that smartphone screen compatibility should be improved if scrolling to select buttons was added. For guidance, participants advised that more straightforward description of buttons should be added. With regard to interface design, participants mentioned that they would appreciate simplifying the interface and incorporating more interactive communication instead of relying solely on input boxes. In the aspect of loading time/error, faster system uploading and reduced waiting times were recommended, along with the addition of a loading bar for image generation to reduce user uncertainty during the process. For chat response, it was suggested that AVERY could offer examples or suggestions when encountering difficulties in answering user questions. Other opinions included that AVERY can provide some useful vocabulary.

5. Discussion

In this paper, we proposed, built, and trialed a game called AVERY, designed to simulate an effective communication process by providing players with a generated image as a feedback.

To address the following research questions:

RQ1. How do learners perceive Ease of Use, Usefulness, and Enjoyment of the chatbot? What are their Attitudes and Intentions toward the game?

The results of the exTAM indicated that participants generally perceived AVERY positively across all factors, with Perceived Enjoyment (PE) receiving the highest score. The mean of participants' attitudes and intention was the lowest among all factors. Open-ended responses revealed that participants particularly enjoyed the images generated by Skyler. However, issues with the interface design, loading time, and system errors contributed to the relatively low attitudes and intention toward AVERY.

RQ2. How valid is AVERY's generated image and feedback to help the user understand the difference between vocabularies?

The generated images were generally effective in reflecting users' wording. As shown in Figure 5, words like "rat", "mice," "mouse," and "rodent" were accurately depicted in the generated images. However, AVERY did not clearly convey the differences between these vocabulary terms. Further studies are needed to assist users in classifying vocabulary mismatches and misinterpretations. While most participants responded positively to the generated images, one participant mentioned that the images encouraged them to write more specific sentences in subsequent game rounds. However, two other participants noted that Skyler's images could sometimes be incoherent with the user's sentence, particularly in terms of the number of objects depicted. The feedback provided by AVERY was mixed, with 44% considering it valid and 41% finding it invalid. When elements in the original picture were not well understood, AVERY tended to avoid answering users' questions. However, one participant appreciated that Avery would provide hints to help them write during the gameplay.

RQ3. How does AVERY's scoring and evaluation system affect users' English writing?

The length of each user's sentence was not necessarily related to the final score. Long sentences did not promise a high total score. However, all the sentences with a vocabulary score higher than 10 exceed 50 characters. The max of the vocabulary score was limited to 70 in the study because vocabulary at the CEFR levels C1 and C2 often includes terms that are rarely depicted in images. As noted by a participant in the open-ended responses, achieving an A rank was challenging, primarily because the vocabulary score for most game rounds was capped at 10 out of 100. As one participant in the open question suggested, providing useful vocabulary may help users to improve their vocabulary score. Additionally, the Communication Effectiveness score was typically in the range of 50-70 out of 100, but players needed a total score above 80 to achieve an A rank. The minimum Communication Effectiveness score was 40, indicating that only when a player's sentence included a word similar to one in the AI's responses could the score exceed 40. The final evaluation provided to players was generally effective in guiding them with judgment, encouragement, and suggestions.

6. Conclusions

Through the design of the AVERY game system, we aimed to structure the communication process in English writing practice by incorporating image generation technology. Our evaluation with 12 participants revealed that AVERY successfully created an enjoyable learning environment through its use of generated images. In the future, we plan to compare different image generation models and their settings to improve the coherence of the generated images. Additionally, to optimize the learning experience, we intend to explore ways to integrate vocabulary suggestions into the system, helping learners to better recognize and understand differences in wording.

Acknowledgements

This work was supported in part by the following grants: Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Fellows JP22KJ1914, (B) JP20H01722, JP23H01001, (Exploratory) JP21K19824, (B) JP22H03902 and New Energy and Industrial Technology Development Organization (NEDO) under Grant JPNP20006 and Grant JPNP18013.

References

Abuarqoub, I. A. S. (2019). Language Barriers to Effective Communication. Utopía y Praxis Latinoamericana.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language Models are Few-Shot Learners. ArXiv, abs/2005.14165.

Brunet, D., Vrscay, E. R., & Wang, Z. (2011). On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, *21*(4), 1488-1499.

Canning-Wilson, C. (1999).Using Pictures in EFL and ESL Classrooms.

Hasnine, M. N., Mouri, K., Flanagan, B., Akcapinar, G., Uosaki, N., & Ogata, H. (2018). Image recommendation for informal vocabulary learning in a context-aware learning environment. In Proceedings of the 26th International Conference on Computer in Education (pp. 669-674).

Hong, J. C., Lin, C. H., & Juh, C. C. (2023). Using a Chatbot to Learn English via Charades: the correlates between social presence, hedonic value, perceived value, and learning outcome. Interactive Learning Environments, 1–17.

Kim, H.-S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. 영어학, 21, 712–734.

Kusuma, G. P., Wigati, E. K., Utomo, Y., & Putera Suryapranata, L. K. (2018). Analysis of Gamification Models in Education Using MDA Framework. Procedia Computer Science, 135, 385–392.

Lester, J. C, Converse, S. A , Kahler, S. E , Barlow, S T. , Stone, B. A , and Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. 359–366.

Mikolov, T., Chen, K., Corrado, G.S., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. International Conference on Learning Representations.

Ogata, H., Uosaki, N., Mouri, K., Hasnine, M. N., Abou-Khalil, V., & Flanagan, B. (2018). SCROLL Dataset in the Context of Ubiquitous Language Learning. In Workshop Proceedings of the 26th International Conference on Computer in Education (pp. 418-423).

Paivio, A. (1979). Imagery and Verbal Processes (1st ed.). Psychology Press.

Phuong, B. H. (2018). Can Using Picture Description in Speaking Sessions Help Improve EFL Students' Coherence in Speaking? European Journal of Foreign Language Teaching.

Ruan, S. , Jiang, L. ,Xu, Q, Liu, Z. , Davis, G. M, Brunskill, E. , and Landay, J. A. (2021). EnglishBot: An AI-Powered Conversational System for Second Language Learning. In Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21). Association for Computing Machinery, New York, NY, USA, 434–444.

Sarkar, S., Das, D., Pakray, P., & Gelbukh, A. (2016, June). JUNITMZ at SemEval-2016 task 1: Identifying semantic similarity using Levenshtein ratio. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) (pp. 702-705).

Woollaston, S., Flanagan, B., & Ogata, H. (2024). Chatbots and EFL learning: A systematic review. Joint Proceedings of LAK 2024 Workshops, 89–98.

Wu, X., & Gao, Y. (2011). Applying the extended technology acceptance model to the use of clickers in student learning: Some evidence from macroeconomics classes. American Journal of Business Education, 4, 43–50.

Zeng, S., Zhang, J., Gao, M., Xu, K. M., & Zhang, J. (2022). Using Learning Analytics to Understand Collective Attention in Language MOOCs. Computer Assisted Language Learning, 35(7), 1594-1619.