Research Article **Educational Development in Various Fields** **EDIVF-125**

The Effectiveness of a GenAI-Integrated Coaching System on Science Inquiries Generated by Refugee Students

Ian Kim*

Seeds of Empowerment (UNITED STATES)

Abstract

This study explores the impact of the Stanford Mobile Inquiry-based Learning Environment (SMILE) system, enhanced with Generative Artificial Intelligence (GenAI), on refugee students pursuing careers in STEM fields. Addressing the educational disparities faced by this group, this research assesses how SMILE promotes interest in science and improves inquiry skills for refugee students. Through a mixed-methods approach, refugee students interacted with SMILE, engaging in science inquiries while their progress and the evolution of their questioning abilities were monitored. The SMILE equipped with GenAI provides personalized coaching, leveraging Bloom's taxonomy to offer nuanced feedback aimed at advancing students' cognitive engagement. Initial findings reveal a significant enhancement in the complexity and depth of student inquiries, indicating a progression toward higher-order cognitive questions. This advancement underscores the efficacy of SMILE, equipped with GenAI, in fostering critical thinking and creative problem-solving among refugee students. The implications of this study extend to educational technology and pedagogy, suggesting that AI can play a pivotal role in providing equitable access of STEM education for diverse learner populations. By demonstrating the adaptability of SMILE equipped with GenAI for the unique needs of refugee students, this research contributes to the discourse on the integration of AI into educational settings to support marginalized communities in STEM disciplines.

Keywords: Stanford Mobile Inquiry-based Learning Environment (SMILE), GenAI, Inquiry-based learning, Refugee students, STEM education, Cognitive engagement.
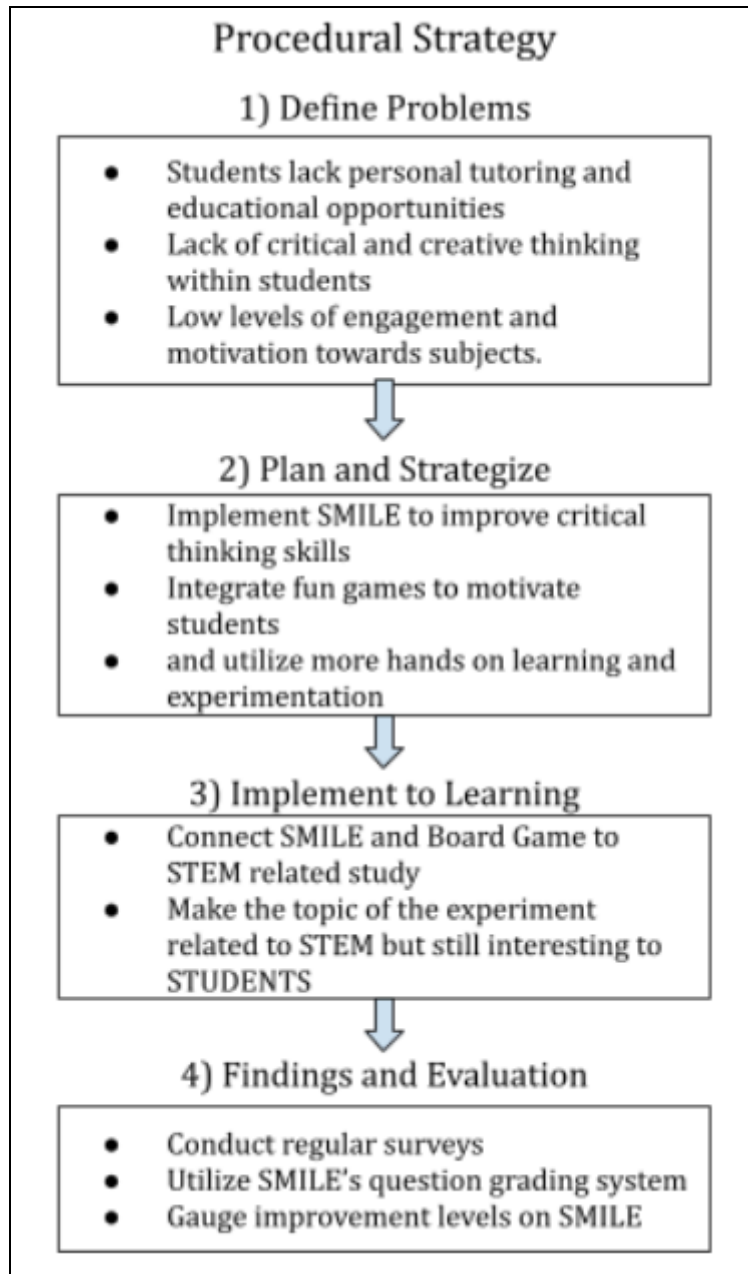
Introduction

The integration of artificial intelligence (AI) in classroom settings has been debated in recent years, especially with the development of programs and software such as ChatGPT which can make it difficult to discern between the work of students and the work of AI(Kay). Some teachers, as mentioned in Kay's paper, argue that using such programs would hinder students' critical thinking, but when utilized correctly for research purposes that enhances classroom settings, AI can be a powerful tool in assisting students through their learning by improving their engagement levels as displayed in the proceedings below. Programs such as SMILE seek to foster critical thinking and problem-solving skills for students by employing Bloom's taxonomy—a classification system for educational learning objectives organized by complexity and specificity(Armstrong). The value of asking thought-invoking questions has been researched to be very impactful on one's learning(Kozminski) and SMILE focuses on this aspect of learning to aid refugees.

In this experiment, the SMILE program was utilized to reduce the educational gaps that refugee students face due to multiple obstacles, including language barriers, limited access to quality resources, and cultural challenges(Schorchit). Students inputted various questions into SMILE, after which the questions were evaluated and graded by AI on a scale of one to five based on how complex and thought-provoking their questions were. For example, "How many colors are in the rainbow?" is regarded as a simple factual recall question that is stated as a basic fact and graded as Level 1, but "How would the world change if the color red didn't exist in the rainbow" is graded as Level 5 since it is regarded as a thought invoking, challenging and hypothetical situation to which the answer cannot be

easily obtained. Students' activity aims to reach higher levels with every question they ask. By formulating more engaging questions, students can gain a better understanding of various science topics and an interest in STEM fields. The SMILE program effectively provided individualized coaching for each student based on the questions they formulated, and the students were able to gain a stronger grasp for formulating more engaging questions by the end of the study. Our research contributes to the discourse on integrating AI into educational settings not only for students from ordinary families but also for those from socioeconomically disadvantaged families, which thus implies that modern educational technology can assist struggling students in all aspects.

Procedural Strategy

1) Define Problems

- Students lack personal tutoring and educational opportunities
- Lack of critical and creative thinking within students
- Low levels of engagement and motivation towards subjects.

2) Plan and Strategize

- Implement SMILE to improve critical thinking skills
- Integrate fun games to motivate students
- and utilize more hands on learning and experimentation

3) Implement to Learning

- Connect SMILE and Board Game to STEM related study
- Make the topic of the experiment related to STEM but still interesting to STUDENTS

4) Findings and Evaluation

- Conduct regular surveys
- Utilize SMILE's question grading system
- Gauge improvement levels on SMILE

Methodology

Overview/Background

The total study was divided into 4 sessions spanning over 6 hours. There were a total of 6 students participating in this study. The students ranged from 9 years to 16 years old (5th to 11th Grade) and were refugees from countries including North Korea, Eritrea, and Ethiopia. Their language barrier was minimal as most students had a good foundation in English (except for North Korean students, but they were able to communicate in a mixture of English aDnd Korean). Over the span of the 4 sessions, the students were introduced to the SMILE program and also implemented it to their learning.

Introducing SMILE program (Day 1)

The first day involved introducing students to the SMILE program and explaining how it functions based on Bloom's Taxonomy, which categorizes thinking skills from basic recall to advanced analysis and creation to measure cognitive activity. Then, the students were asked to formulate random questions of their choice and create as many questions as they could based on the individualized feedback the machine gave. The questions could be grouped into 5 levels depending on the difficulty. So, the students were asked to self-evaluate their questions and explain what level their questions would be placed in the SMILE program to spur their critical thinking and recognize the difference between complex "level 5" questions and simpler "level 1" questions. Photographs of the activity are displayed in Fig 1 to 4.
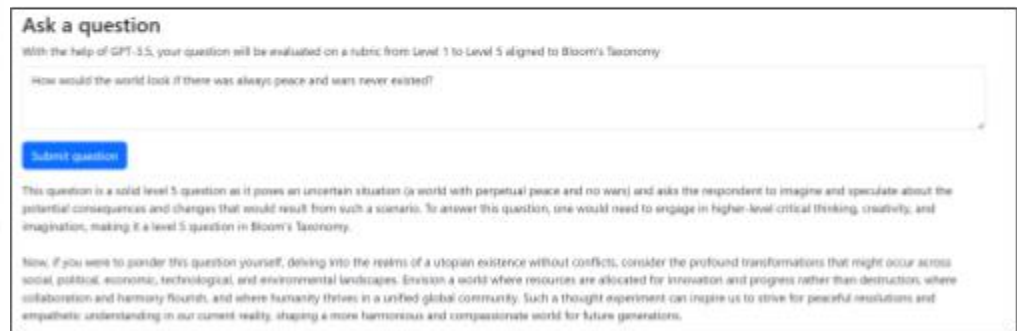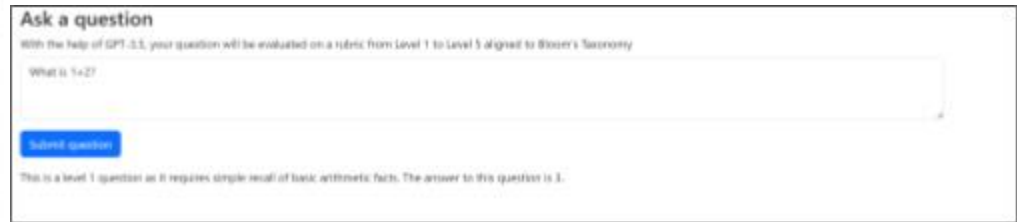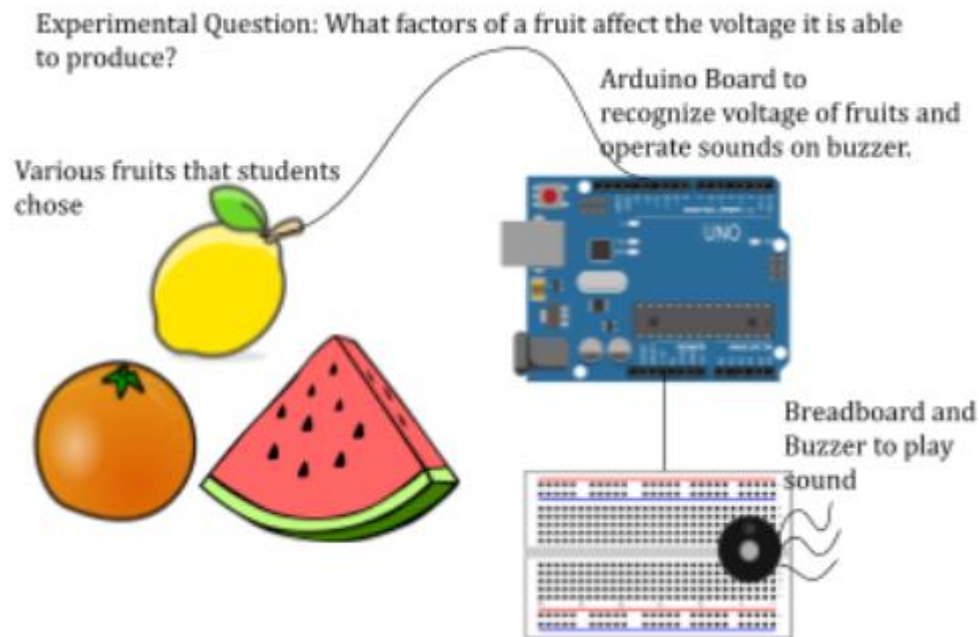

Figure 1. Left: Students Organize questions into their expected level
Figure 2. Right: Example of Level 1 vs. a Level 5 question created by students

Integrating SMILE into STEM-related study (Day 2)

Once the students were familiarized with SMILE, they were introduced to an experiment—which they would eventually perform—that was designed to connect their background of music lessons (provided by the refugee tutoring center) to STEM. The experiment consisted of connecting a circuit board to different fruits and having the circuit board produce a sound of different frequencies in accordance with the fruit it was connected to depending on the voltage that the fruit produced.

Figure 4: Code for Arduino Board used in experiment. Maps a situation where arduino board outputs different frequencies on the buzzer based off of a higher or lower voltage measured in the fruits

With the experiment in mind, students were asked to a list of key terms and variables that might be related to the experimental question (for example: Acid batteries, voltage, frequencies, etc..). Since the Arduino Board was programmed to output a different frequency based on the different voltages in the fruits, students had the goal of figuring out what factors might affect the voltage that is produced by a fruit including acidity, salinity, water content, size of fruit and more. Once the extensive list of variables was created, students were required to formulate new questions of varying difficulty utilizing the words they had just created (for example, how might the acidity of a fruit affect the voltage that the fruit battery produces) with the goal of answering the experimental question.

Figure 5: Left: Students create a list of words related to the STEM experiment
Figure 6: Right: Students create questions from the list of words

Board game (Day 3)

After crafting their own questions, students participated in a board game (#6) to display their ability to pinpoint questions of different levels. The board game was designed with the help of ChatGPT to create rules. The objective for each student was to return to the base position of the board, which could be achieved through many paths. Students were allowed to move tiles if they correctly guessed the level of a question. The number of tiles they moved per turn was dependent on the level of question they guessed (for example, if a student guessed level 5 question correctly, they could move 5 tiles. If level 3 was guessed correctly, they could move 3 tiles). If a student incorrectly guessed the level, they could not move any tiles.
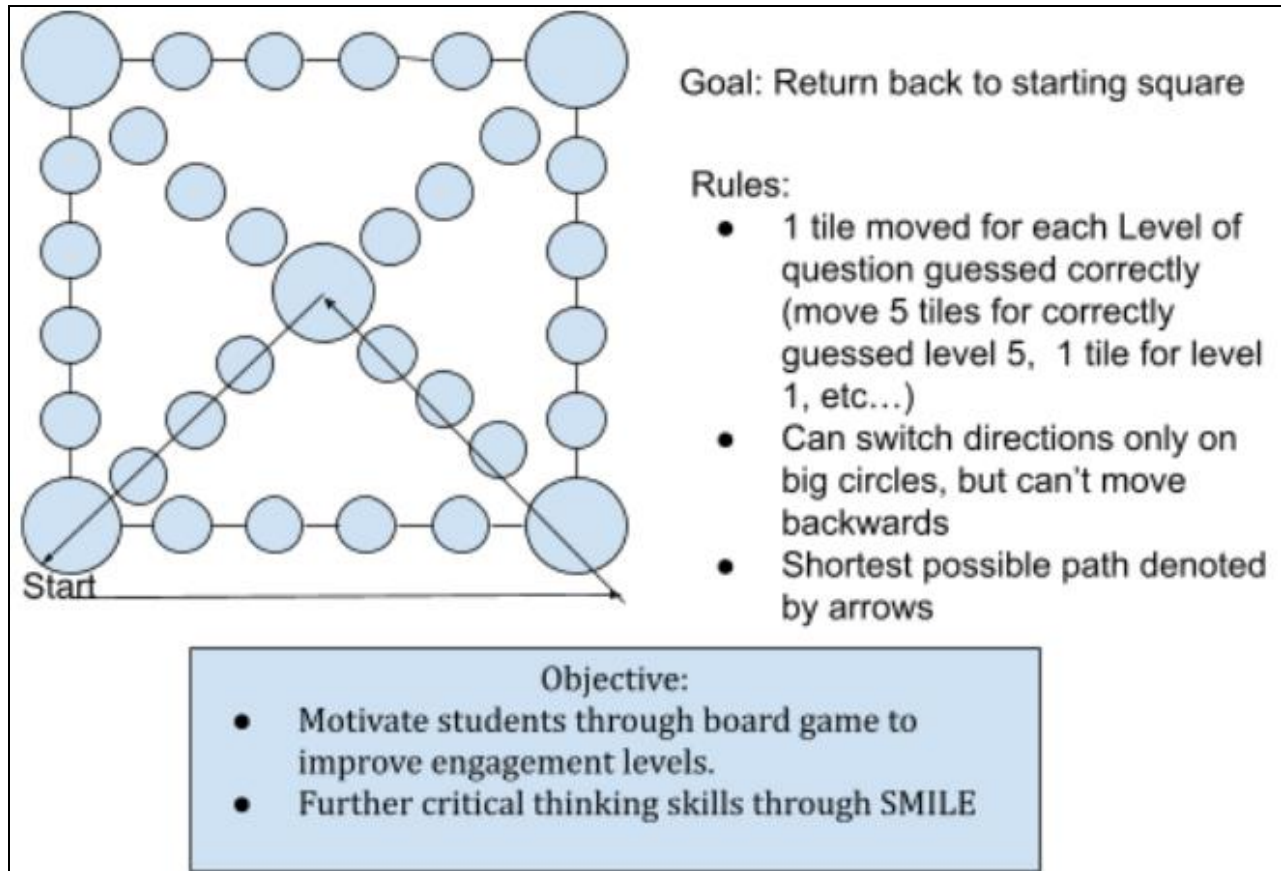
Goal: Return back to starting square

Rules:
- 1 tile moved for each Level of question guessed correctly (move 5 tiles for correctly guessed level 5, 1 tile for level 1, etc...)
- Can switch directions only on big circles, but can't move backwards
- Shortest possible path denoted by arrows

Start

Objective:
- Motivate students through board game to improve engagement levels.
- Further critical thinking skills through SMILE

Figure 7: Model of Board Game and its rules/objectives.

Experiment (Day 4)

The students were also asked to design an experiment such that the questions and factors they created from Day 2 and Day 3 could be tested independently (for example, to test the effect of fruit size, they used one small slice of watermelon and one big piece of the same watermelon so that other variables remained constant). Then, the experiment was performed, and students wrote down observations to determine what factors affected the output voltages from the fruit batteries. Students then compared their findings with their original hypotheses and questions. Photographs of the activity are displayed in Fig 7 to 9.
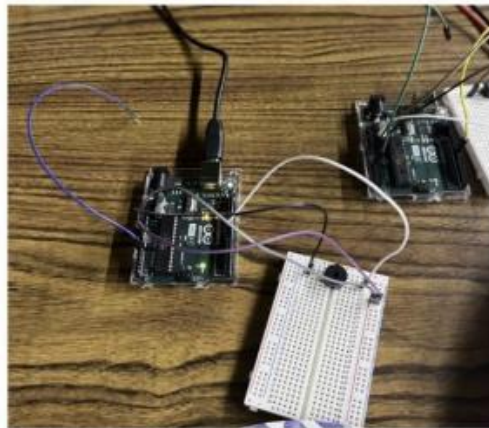

Figure 7. Left: Students create a circuit board
Figure 8. Middle: Completed Circuit board by students
Figure 9. Right: Students perform experiments and write observations

Results

Introducing the SMILE program (Day 1)

Charts 1 and 2 show students' performance on the day. The number of level 1 questions is the largest, and there are few questions for levels 2 to 4. This means that students cannot control the difficulty of the questions as they want. Chart 2 also shows that the number of correct guesses for the difficulty of the question is low (35%). Additionally, most of the questions asked were either level 1 or level 5 and this was because the students tended to ask more hypothetical or fact based questions rather than more specific and insightful questions to specific topics.
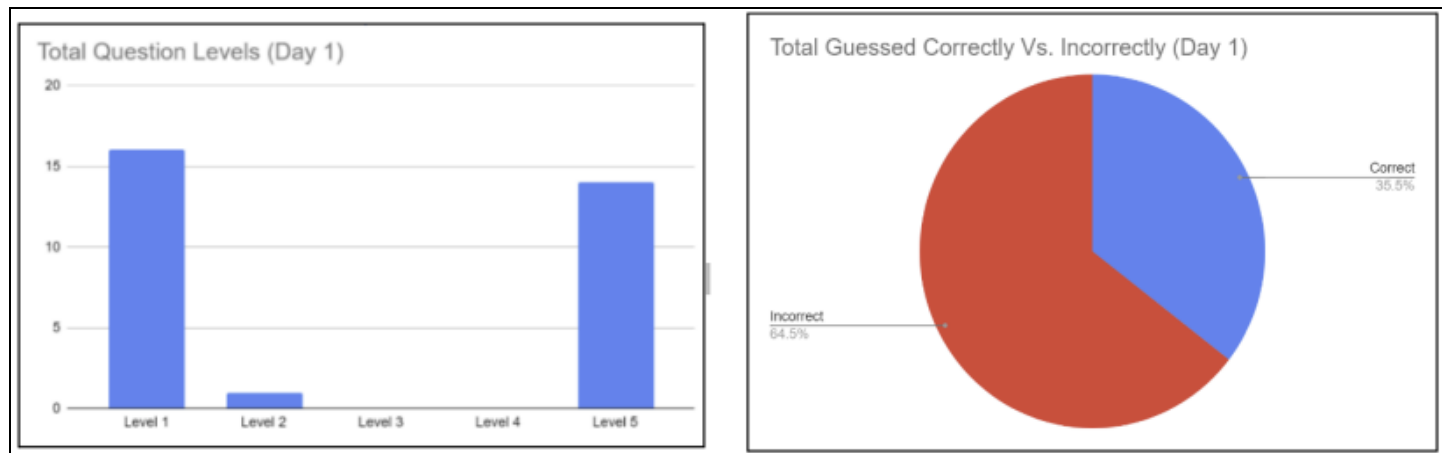
Chart 1: Number of questions for different difficulty levels on day 1
Chart 2: Percentage of correct guess (35%) and Incorrect guess (65%) on day 1

SMILE program in STEM-related study (Day 2)

After the introduction on day 1, students' performance improved on day 2, as can be seen in charts 3 and 4. The number of questions correctly guessed increased to 48%. Additionally, the students asked fewer level 1 questions and a greater concentration of upper-level (3, 4, and 5) questions, showing the improvement in students' cognitive thinking.

What did you feel was the most difficult part out of all the sessions we've had?

5 responses

Guessing the question level accurately because I disagreed with the AI sometimes. For example, when I asked the question about what changes the energy a fruit produces, I expected the question to be a level 4 or 5 because there could be many factors that change the answer, but the program only rated it a level 2.

Forming interesting questions with certain keywords was hard because you had to think about specific scenarios that don't usually occur. But I felt that it got easier as I made more and more questions.

Trying to guess the question levels for the Board Game. I thought many questions could either be between level 2 or 3 and it wasn't super clear how they were differently scored on SMILE.

The most difficult part was trying to get level 5 questions that weren't to general like "what if there was no sound."

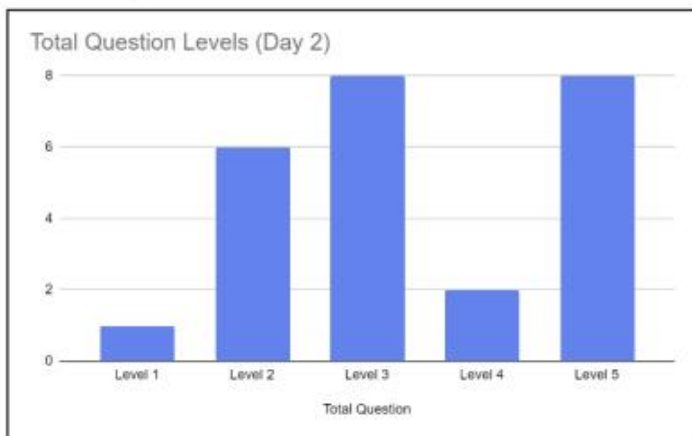I had trouble finding factors that would affect the voltage and energy that a fruit would produce.

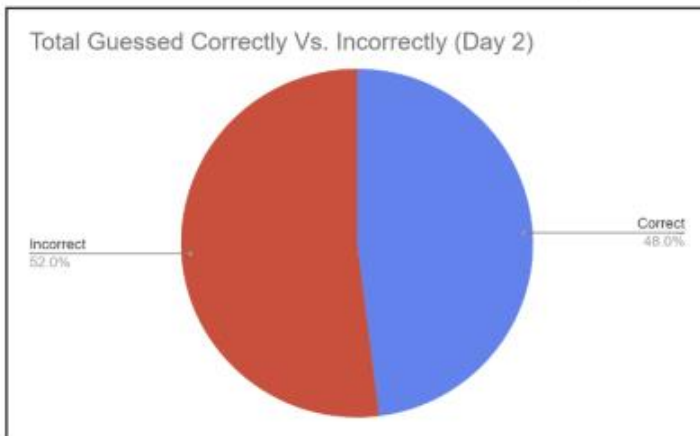**Chart 3:** Number of questions for different difficulty levels on day 1
Chart 4: Percentage of correct guess (48%) and Incorrect guess (52%) on day 2

What students learned

At the end of Day 3, we conducted an anonymous survey for the students to evaluate the program. The questionnaires asked how much the students enjoyed the program, how difficult it was, what they liked the most about it, and how it could be improved. Students' responses to these questionnaires are shown in Figs. 10 to 12
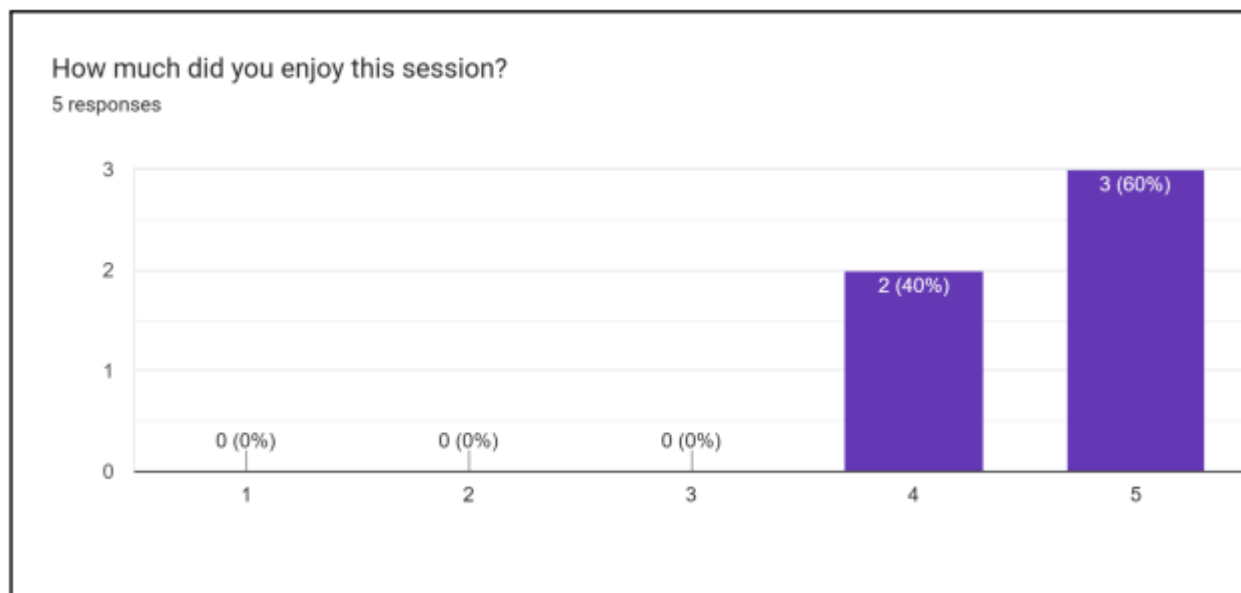
Figure 10: Survey results showed how much the participating students enjoyed the program on a scale of 1-5, with 5 being enjoyed very much and 1 not at all. Overall, the program was accepted positively.
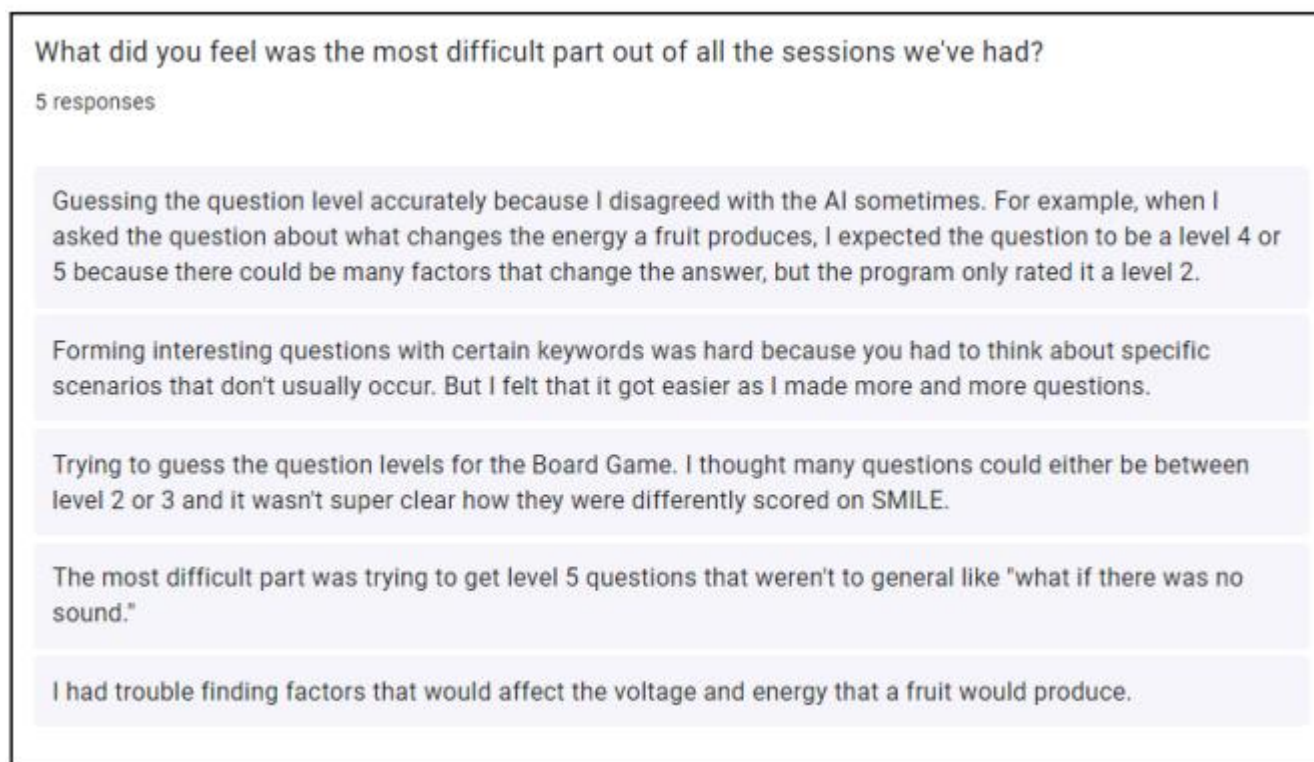

Figure 11: Survey results show how difficult the program was. Most students answered that guessing the question level of difficulty was the most challenging part, and some also commented on the efficacy of AI.
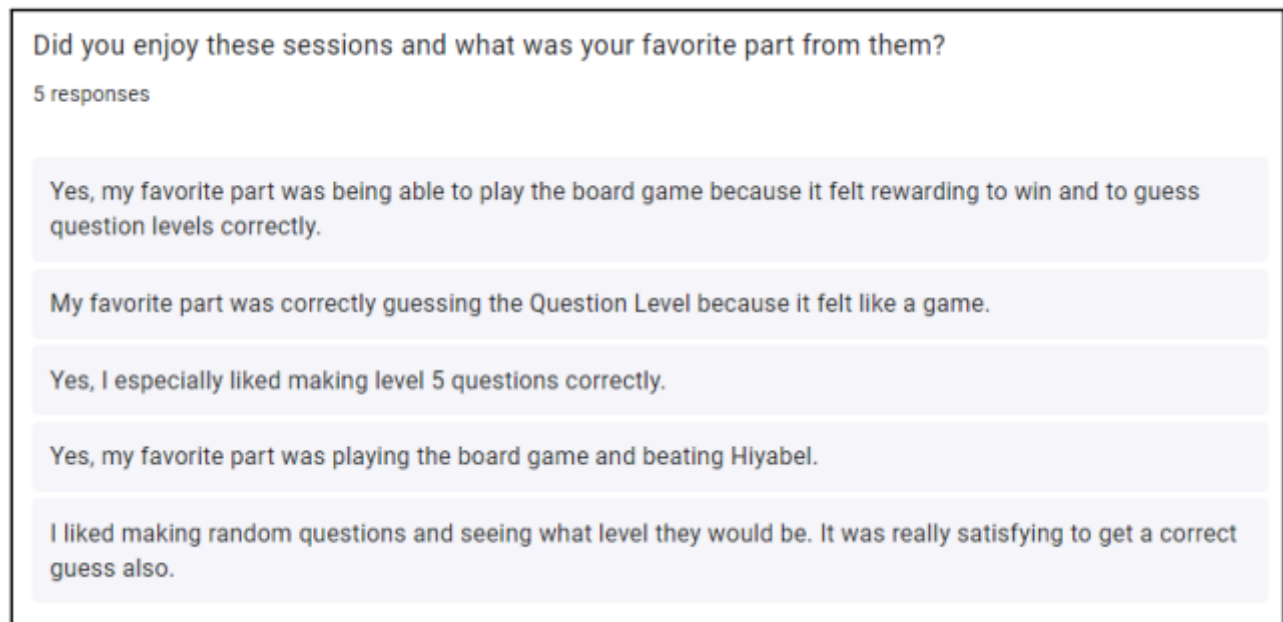
> Did you enjoy these sessions and what was your favorite part from them?
>
> 5 responses
>
> Yes, my favorite part was being able to play the board game because it felt rewarding to win and to guess question levels correctly.
>
> My favorite part was correctly guessing the Question Level because it felt like a game.
>
> Yes, I especially liked making level 5 questions correctly.
>
> Yes, my favorite part was playing the board game and beating Hiyabel.
>
> I liked making random questions and seeing what level they would be. It was really satisfying to get a correct guess also.

Figure 12: Survey results showing students' favorite parts of the program. In general, students found playing the board game, crafted with the help of ChatGPT, to be enjoyable. Many students also found the SMILE program to be like a game that helped them think as they tried to guess the question levels correctly.

Discussion and Conclusions

This study aims to seal the education gap for refugees utilizing AI, considering the effectiveness of a GenAI-integrated coaching system on science inquiries. In the past few years, over one million refugees have been born around the world. The global pandemic (COVID-19) and the disruption to education impacted all students but has mainly affected refugees, who were already facing significant educational barriers [8]. Today's progress in educational advances for refugees, students, and teachers has led us to witness how such technology and digital learning are rapidly advancing, especially AI. In this experiment, we used the SMILE program integrated with AI to close the educational disparities that refugee students face due to their worlds. Our firsthand experience with the technology explains the ways in which AI can be incorporated into classroom settings to achieve such a goal. Before the experiment, students were unaware of the SMILE program and Bloom's taxonomy. This can be seen through their tendency to form lower-level questions on Day 1 and often overestimate their expected levels for their questions. It was more difficult for many students to find questions requiring them to think critically, but the SMILE program feedback helped them adapt quickly. When returning on the next day to create questions related to a scientific experiment, students could pinpoint what level they believed their questions to be, and the questions were at a higher level on average. Leveraged with a board game created with the help of ChatGPT, the students showed high levels of engagement and enthusiasm within the subject being taught. By the final day, when the students conducted their actual science experiment, they displayed high levels of critical thinking and engagement with the subject. Once students could ask questions they were truly curious about through the SMILE program, their enthusiasm towards the science behind voltage and sound waves was at far higher levels than from the start of the experiment. This was shown by the fact that the agreement on the difficulty of questions assessed by the students and the GenAi was 32% on Day 1, but after the students carried out the experiment on Day 2, it increased to 48 %. Also, our survey conducted after the experiment to evaluate our program confirmed that the SMILE program integrated with GenAI was effective and could readily be adoptable for students from marginalized families.

Ultimately, we hope to adopt SMILE within all learning environments including schools of all levels to help solve the educational gap found within disadvantaged refugee students. The findings of this paper show that SMILE can be effective for achieving this goal in and outside of the classroom setting.

References

1. American Progress. (n.d.). The future of testing in education: Artificial intelligence. https://www.americanprogress.org/article/future-testing-education-artificial-intelligence/.
2. CIS Spain. (n.d.). The benefits of AI in education. https://www.cis-spain.com/en/blog/the-benefits-of-ai-in-education/.
3. Kay, M. R. (2023). Can AI lead a classroom discussion? Educational Leadership, 81(3), 76–77. https://research.ebsco.com/linkprocessor/plink?id=015386c3-eb7e-3936-9fc5-1308534bfb00.
4. Santhanam, R., & Schniederjans, M. J. (1991). Artificial intelligence: Implications for teaching decision science.

Interfaces, 21(5), 63–69. http://www.jstor.org/stable/25061533.

5. UNESCO. (2021). AI in education: A critical review. https://unesdoc.unesco.org/ark:/48223/pf0000368303

6. Schorchit, N. (2017, March 21). Despite inclusive policies, refugee children face major obstacles to education. NEA. https://www.nea.org/nea-today/all-news-articles/despite-inclusive-policies-refugee-children-face-ma jor-obstacles-education

7. PubMed. (2020). https://pubmed.ncbi.nlm.nih.gov/32776345/

8. Lasko, E. (2023, May 19). The impact of covid-19 on refugee education. Refugees International. https://www.refugeesinternational.org/the-impact-of-covid-19-on-refugee-education/

9. Kozminski, K. G. (2013, March). The value of asking questions. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3596240/

10. Armstrong, P. (1970, June 10). Bloom's taxonomy. Vanderbilt University. https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy