Development of the Hands-free AI Speaker System Supporting Hands-on Science Laboratory Class: A Rapid Prototyping

https://doi.org/10.3991/ijet.v18i01.34843

Gyeong-Geon Lee, Minji Choi, Taesoo An, Seonyeong Mun, Hun-Gi Hong(✉)
Seoul National University, Seoul, Republic of Korea
hghong@snu.ac.kr

Abstract—The recent progress of natural language processing (NLP), speech recognition, and speech generation envision using hands-free artificial intelligence (AI) speakers in classrooms to support student learning. In science education, conventional hands-on laboratory education has been considered crucial in fostering students' manipulative experimentation skills. However, touching things with gloved hands other than experimental equipment and apparatuses is strictly restricted because of the safety issue, which calls for another channel to get timely support. Therefore, we ideated that adopting hands-free AI speakers in the hands-on science laboratory classroom would support student learning. Using the rapid prototyping method, we designed and developed an AI speaker-based system that answers student queries concerning solution-making, experimental processes, and waste liquid disposal, corresponding to the initial, middle, and final phases of a laboratory class. The system was internally validated by usability tests of 9 expert panels and 18 university students and then revised. The revised system was externally validated in an analytical chemistry experiment class for 3 sessions with 13 university students. We present the result of the prototype development and internal and external validations with quantitative and qualitative data. The AI speaker system enabled students to use the auditory learning mode in the laboratory while concentrating on the experimentation with their hands in the external validation. Future research topics were suggested.

Keywords—AI in education (AIEd), hands-free AI speaker, hands-on science laboratory class, rapid prototyping, natural language processing (NLP)

1 Introduction & background

The innovative educational research and practices to follow up the rapidly changing contemporary world are now supported by Artificial Intelligence (AI) technologies [1]. The Natural Language Processing (NLP) technology has explosively grown due to the deep learning algorithms [2], which are relevant to converting human speech to text (STT; or speech recognition, SR) and text to human-like speech (text to speech, TTS; or speech generation). This technological development yielded products, services, and platforms of AI-based smart speakers, such as Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home [3]; and they are more and more integrated into our daily lives.

For decades, scholars have insisted that the AI in Education (AIEd) will bring innovations in teaching and learning and suggested their framework for its roles [4–6]. Although there can be many frameworks of visions for AIEd, its epitome seems clear: the AI comes into the classroom where there are instructors and students, changes the processes of learning, reducing instructors' burdens for repetitive work and allowing them to engage with more valuable teaching behavior, and supporting student learning in cognitive, affective, or any other learning modes (based on [6–8]).

However, there seems that there has not been much empirical research on the possibilities of AI Speakers in Education (AISEd), and there is still a need for explorative studies on its utility in being an intelligent agent to answer students' questions based on natural languages. (cf. [9]) Some innovative research on AISEd has focused on the possibility of native language or English education [6, 10–11]. Nevertheless, it seems that there has been little research that incorporated AI speakers into the science classroom.

Science education has emphasized the 'hands-on' laboratories science 1980s [12]. In the usual hands-on science laboratory class, students gather weekly in the lab and conduct an experiment to get data, which is interpreted in light of scientific theories. And the important learning mode in the so-called hands-on laboratory class is kinesthetic [13] as they are expected to foster experimentation skills [14–15]. Students manipulate apparatuses and equipment such as beakers and other glassware, pipettes, pH meter, UV-VIS spectrometers, and treat reagents that need caution to avoid accidents [16]. Therefore, strict safety rules compel students to wear latex gloves, goggle, and lab coats and not allow them to touch non-experimental items with gloved hands, such as door handles, elevator buttons, smartphones, laptops and even lab notes (see [17]).

Consequently, in principle, students cannot freely access the information they need during the experiment and cannot help but ask TA frequently. Here, the 'hands-free' characteristic of the AI speaker stands out – as a user need not use their hands to operate it but verbally calls a system to get the information they require, students in a science laboratory class will also benefit from it, without mitigating the safety rules. Therefore, the need for empirical research that implements AI speakers in an authentic laboratory teaching and learning site arise.

In this study, the researchers took the design and development (D&D) research approach for the hands-free AI speaker system supporting a hands-on science laboratory class. As the product of this research is a novel instructional tool, the rapid prototyping (RP) method that acquires user feedback for revision before the final field test would be helpful. Throughout the development and evaluation of the AI speaker system, the researchers would investigate its merits based on the responses of experts in AI, educational technology, and science education, and university students. And further, the current status, limitations, and future directions for the AI speaker system in the science laboratory class will be discussed.

2 Research questions

1. How the prototype of the hands-free AI speaker system supporting a hands-on science laboratory class is developed using the RP approach?
2. How was the result of the internal validation of the prototype of the hands-free AI speaker system supporting a hands-on science laboratory class?
   2.1. What were the responses from experts?
   2.2. What were the responses from university students?
3. How was the result of the external validation of the prototype of the hands-free AI speaker system supporting a hands-on science laboratory class? (what were the responses from university students who experienced the system in the authentic science laboratory classroom?)

3 Method

3.1 Design and development research

D&D research is "the systematic study of design, development and evaluation processes with the aim of establishing an empirical basis for the creation of instructional … tools" [18]. D&D research is categorized as follows – Type I: Product & tool research and Type II: Model research [18]. As the former is relevant to this study, we focused on the specific product and yields context-specific lessons learned from developing it and analyzing the conditions for its optimal use [19].

3.2 Rapid prototyping approach

This study took an RP approach [20–21]) to develop a hands-free AI speaker system supporting hands-on laboratory classes. As the RP involves the typical stages of analysis, design, development, implementation, and evaluation (ADDIE) while emphasizing the formative evaluation and iterative process [22], we followed the generic steps.

3.3 Research field

The research field of this study was Hankuk University (pseudonym), located in Seoul, Republic of Korea. Two considerations arose from the characteristics of the research field: (1) As the STT technology of the Korean language is behind that of the English language, the performance of the AI speaker also reflects its current level. (2) Most student participants in this study were enrolled in the Department of Chemistry Education – i.e., most of them were pre-service chemistry teachers.

3.4 The ADDIE process

Need analysis and design of the AI speaker system. The researchers of this study consisted of one expert, one doctoral candidate, and three master's students of chemistry education. They all had a teacher's license for secondary chemistry. They had years of university laboratory teaching experience.

As a need analysis, the researchers reviewed literature related to the characteristics of hands-on science laboratory education (e.g. [12, 14–15]) to analyze the need for the AI speaker system for hands-on laboratory classes (cf. [23]). And they preliminarily asked several pre-service chemistry teachers what functions of the AI speaker system would help them take laboratory classes. As a result, the researchers concluded that a supporting tool for hands-on science laboratory class should reduce the instructors' repeated work and promote student learning, particularly for necessary but repetitive work in the laboratory class – i.e., solution-making, checking experimental procedures and disposing of liquid waste according to safety rules. The three functions of the AI speaker system derived from those needs are presented in Table 1. And the overall conceptual flow chart for the system is illustrated in Figure 1.

Table 1. Summary of three functions of hands-free AI speaker system supporting hands-on science laboratory class

Function	Target Phase of the Laboratory Class	Need	References
(1) Solution-making helper	The beginning	– Students must calculate the amount of reagent needed to make various kinds of solutions using complex formulas.	[24–27]
(2) Experimental process helper	The middle	– Students repeatedly check that they are proceeding well according to the experimental procedure. – Students frequently question the reason and issues of a certain procedure.	[28–30]
(3) Liquid waste disposal helper	The end	– Students must dispose of liquid wastes into each safety container according to their property.	[31–33]

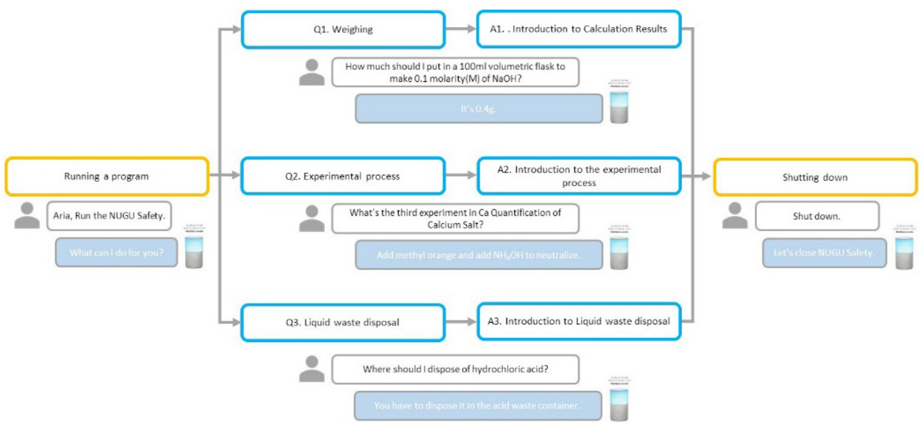

Fig. 1. Conceptual flow chart (scenario) of the AI speaker system supporting hands-on laboratory

The NUGU platform required the AI speaker to be connected online to receive a signal from the servers. Therefore, it was considered that the AI speaker should be connected to the Wi-Fi as a network environment. If the Wi-Fi speed increases, the time required to SR, NLP, and respond to user articulation would decrease. Therefore, the network environment had to be accessible to high-speed Wi-Fi.

Meanwhile, the user-AI speaker interface was designed. As most AI speaker platform provides a Q&A system with alternate articulation between the user and the speaker, this study followed the structure. The three functions were initially designed to have a single-turn conversation – i.e., the user asks a speaker a question in a sentence and the speaker answers for that. While designing and testing function 1, the need for error handling arose while the function needed to recognize three parameters from the user's speech. Therefore, the researchers took a multi-turn approach to recognize a parameter from the user's single articulation.

Development of the prototype. We developed the prototype of the AI speaker system upon the NUGU AI speaker platform serviced by SK Telecom Co., Ltd. In the platform, entity means a developer-defined category of word tokens; intent implies the user's intention while speaking to the AI speaker, inferred by a combination of entities; and action means the AI speaker's response matched with each intention. We defined the required entities, intents, and actions in the NUGU Player Builder (Figure 2). The examples of entity, intent, and action for each function are presented in Figure 3. And we trained the natural language understanding (NLU) model in the NUGU platform by inputting several word tokens for an entity and dozens of sentences that include a combination of entities for an intent. The actions of *function 2* and *function 3* were directly saved in the Play. Meanwhile, the external proxy server was connected to *function 1* in the JSON format using Python Flask to calculate the amount of reagent.

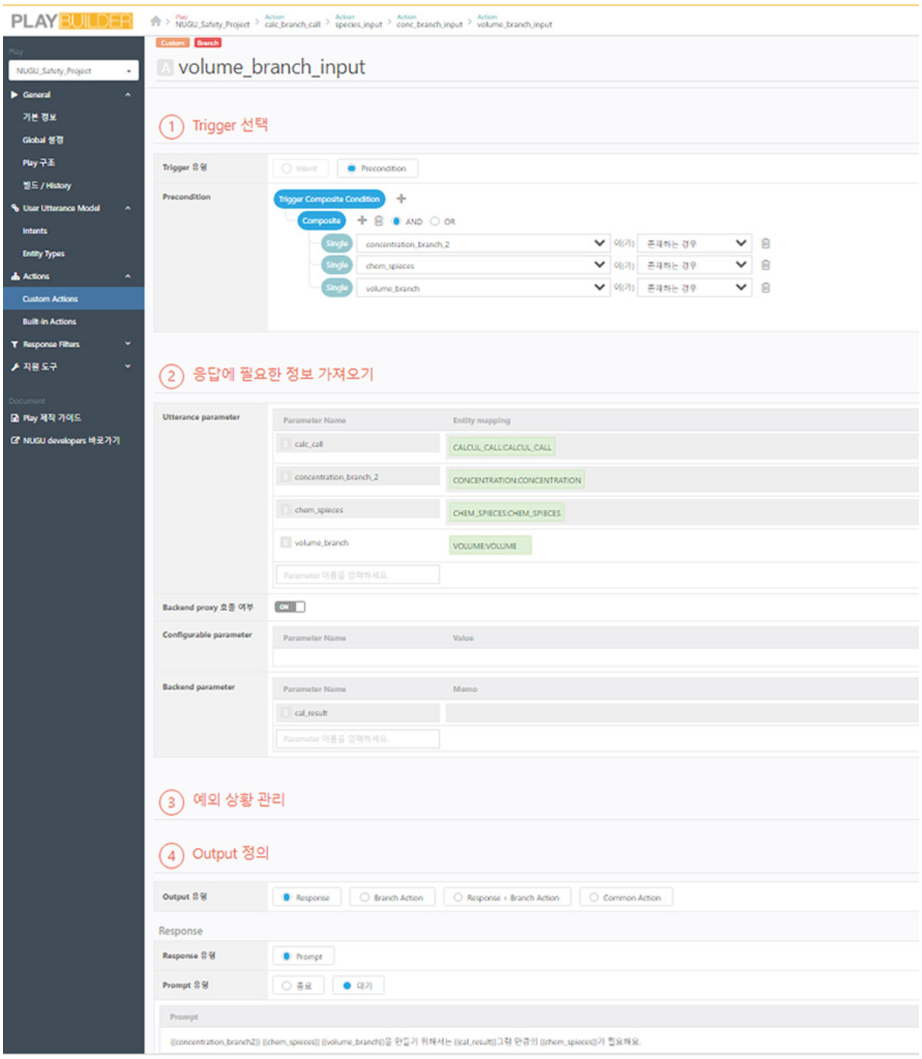

Fig. 2. Example of the development on the NUGU play builder

Function	Intent (user speech)	Entity (word token)	Action (AI speaker speech)
(1) Solution-making helper	- CALCULATION_BRANCH_CALL: "I want to **make** a solution." (multi-turn conversation starts) (1) SPECIES_INPUT: "**sodium hydroxide**" (2) CONCENTRAION_BRANCH_INPUT: **0.1M**" (3) VOLUME_BRANCH_INPUT: **100 milliliters**"	- MAKE: make, produce - CHEM_SPIECES: NaOH, Sodium hydroxide - CONCENTRATION: 0.1M, 0.1mol, 0.1 molarity - VOLUME: 100mL, 100 milliliters, one hundred milliliters - HOW_MUCH: how much, how many, how, to what amount - GRAM: gram, grams	- CALCULATION_BRANCH_CALL: "What reagent are you going to use?" (multi-turn conversation starts) (1) SPECIES_INPUT: "How much molarity will you make?" (2) CONCENTRATION_BRANCH_INPUT: "How many milliliters of volumetric flask are you going to use?" (3) VOLUME_BRANCH_INPUT: "To make **100 milliliters** of **0.1M** of **sodium hydroxide** solution, you need **{{CALCULATION_RESULT}}*** of NaOH."
(2) Experimental process helper	- PROCEDURE_ASK: "What is **the second process** of **$HClO_4$ solution standardization** experiment?"	- HCLO_TITRATION: $HClO_4$ solution standardization, perchloric acid solution standardization - PROCEDURE_TWO: second process, second step, the second process, second	- PROCEDUER_RESPONSE: "**The second process** of **$HClO_4$ solution standardization** is to dissolve KHP solution with acetic acid in a 100-milliliter volumetric flask."
	- ENDPOINT_CALL**: "How does the color change at the **endpoint** in the **titration** experiment?" - CAUTION_CALL**: "What should I **be careful** about in today's experiment?"	- ENDPOINT: endpoint, end point, equivalence point - TITRATION: titration - CAUTION: caution, points to note, be careful	- ENDPOINT_RESPONSE: "The indicator will turn purple to blue when titration is done." - CAUTION_RESPONSE: "The experiment manual has changed, so we have to put 15 milliliters of KHP solution in the process of number 2 of HClO4 solution standardization today. And be careful with acidic solutions because today's experiment is non-aqueous titration."
(3) Liquid waste disposal helper	- ACID_DISPOSAL: "Where do you throw **nitric acid**?" - BASE_DISPOSAL: "What kind of waste liquid is **sodium hydroxide**?" - ORGANIC_DISPOSAL: "Where do you throw **methyl orange**?" - INORGANIC_DISPOSAL: "What kind of waste liquid is **arsenic trioxide**?" - NEUTRAL_DISPOSAL: "Where do you throw **sodium bicarbonate**?"	- ACID: vitamin C, nitric acid, perchloric acid, chloric acid - BASE: sodium hydroxide, ammonium hydroxide, methyl violet - ORGANIC: ammonium oxalate, methyl orange, sodium acetate - INORGANIC: arsenic trioxide, potassium iodide, potassium permanganate, iodine - NEUTRAL: starch, sodium bicarbonate	- ACID_DISPOSAL: "Throw **nitric acid** in acid liquid waste container." - BASE_DISPOSAL: "**Sodium hydroxide** is alkali." - ORGANIC_DISPOSAL: "Throw **methyl orange** in organic liquid waste container." - INORGANIC_DISPOSAL: "**Arsenic trioxide** is inorganic substance." - NEUTRAL_DISPOSAL: "You can throw **sodium bicarbonate** away in the sink."

{{}}: Denotes parameter

*: The value of {{CALCULATION_RESULT}} is returned from the external server connected to the Play ; **: Added after the formative assessment of the prototype (internal validation)

Dotted arrows (-->): Signifies the flow of conversation in the multi-turn conversation. ; Arrows (→): Signifies that the recognized work token from the user's articulation is transferred to the AI speaker's articulation.

Fig. 3. The corresponding structure of user intents, entities, and actions in the AI speaker system

Internal validation as a formative evaluation. The researchers shot a video that shows a person using the AI speaker prototype with three functions in sequence. In the video, a person awakes the system and asks the speaker how to make a specific solution (*function 1*), what is the first or second step of the experimental procedure (*function 2*), and where to dispose of a couple of kinds of waste liquids (*function 3*), and finally ends the session. The video was about 2 minutes long.

The participants of the internal validation consisted of two groups: the expert review panel and the student group. The expert review panel consisted of nine experts from three fields – three from AI technology, three from educational technology, and three from science education. They were asked to read the documents related to the background and process of this study, watch the video, and respond to an individual interview, explicating their thoughts on the system. The student group consisted of 16 pre-service chemistry teachers, 1 in-service chemistry teacher, and 1 chemistry graduate student. Ten of the pre-service teachers were taking an Analytical Chemistry Experiment (ACE) class for the secondary pre-service chemistry teachers.

The items for the usability test were adopted from the PACMAD (People At the Centre of Mobile Application Development) usability model [34]. Further, we asked them about the pros and cons of the system and suggestions for the system revision.

External validation as a summative evaluation. We implemented the hands-free AI speaker system supporting a hands-on science laboratory class. The research field was the ACE class at Hankuk University mentioned above. Thirteen university students (pre-service chemistry teachers) who took the course participated in the study.

While revising the AI speaker system based on the results of the internal validation, the researchers implemented and observed students using each function of the system in a laboratory session (Table 2). The scene of students using the AI speaker-based laboratory class supporting system was videotaped.

Table 2. Functions of the hands-free AI speaker system supporting hands-on science laboratory class

Session	Content of Experiment	Tested Function
1	Production and standardization of $KMnO_4$ solution	Solution-making helper
2	Titration in the non-aqueous solvent	Experimental process helper
3	Iodiometric titration of Vitamin C	Liquid waste disposal helper

After the three sessions with the AI speaker system, students were asked to respond to the survey. The survey consisted of three parts: (1) The system usability scale (SUS) consisted of 10 items on a 5-point Likert scale [35]. (2) The researchers newly developed a survey to investigate students' Perceptions of an AI Speaker System in a Science Laboratory Classroom (PASS-SLC) that consisted of 18 items on a 5-point Likert scale. It was developed with reference to the previous literature on the technology acceptance model or usability of AI speakers or chatbots [36–39], and the learning outcomes of the science laboratory class [14–15]. (3) The survey included open-ended questions at the end of SUS and PASS-SLC, asking, "why did you think like that for these items?" And the survey asked students about the pros and cons of the AI speaker system and suggestions for future revision.

4 Results

4.1 The prototype of the AI speaker system supporting hands-on science laboratory class

Figure 4 represents the one-shot demonstration video of the prototype of the AI speaker system developed in this study, which is about 2 minutes long. At first, a researcher awakened the system by calling 'Aria' (the designated name of the agent in the NUGU platform) (Figure 4a and b). After the Play was loaded, a researcher requested *function 1* to *function 3* in sequence, which the speaker successfully performed (Figure 4c–v). Note that, during the multi-turn conversation in *function 1*, the speaker handled an error recognizing the VOLUME entity (Figure 3), asking the user again (Figure 4h–k). After the user finished testing all the functions, he commanded to shut down, which was realized (Figure 4w–x). The response time of the AI speaker for each articulation of the user took a couple of seconds.

4.2 Internal validation

The survey scores from the expert review panel and university students are presented in Table 3.

Table 3. The PACMAD quantitative survey result of internal validation (1–4 scale)

Category	Item	Expert (N = 9)	Student (N = 18)
User	Effectiveness	3.55 (.53)	3.29 (.69)
	Efficiency	3.44 (.73)	2.88 (.86)
	Average	3.5 (.56)	3.09 (.59)
Task	Satisfaction	3.22 (.67)	3.18 (.57)
	Learnability	3.67 (.71)	3.65 (.61)
	Average	3.44 (.58)	3.41 (.57)
Content	Memorability	3.89 (.33)	3.59 (.62)
	Error	3 (.5)	2.76 (.75)
	Cognitive load	3.78 (.5)	3.44 (.63)
	Average	3.56 (.29)	3.26 (.44)
Overall		3.51 (.3)	3.25 (.43)

Expert review panel. Experts evaluated the effectiveness of the prototype at 3.55 and efficiency 3.44, which were averaged to 3.5 in the user category. They scored the satisfaction of it at 3.22 and learnability 3.67, which were averaged to 3.44 in the task category. Finally, they scored the memorability of it at 3.89, error 3, and cognitive load 3.78, which were averaged to 3.56. The overall score of the quantitative survey from expert review panel was 3.51, which is quite positive (> 2.5) on the 4-point Likert scale.

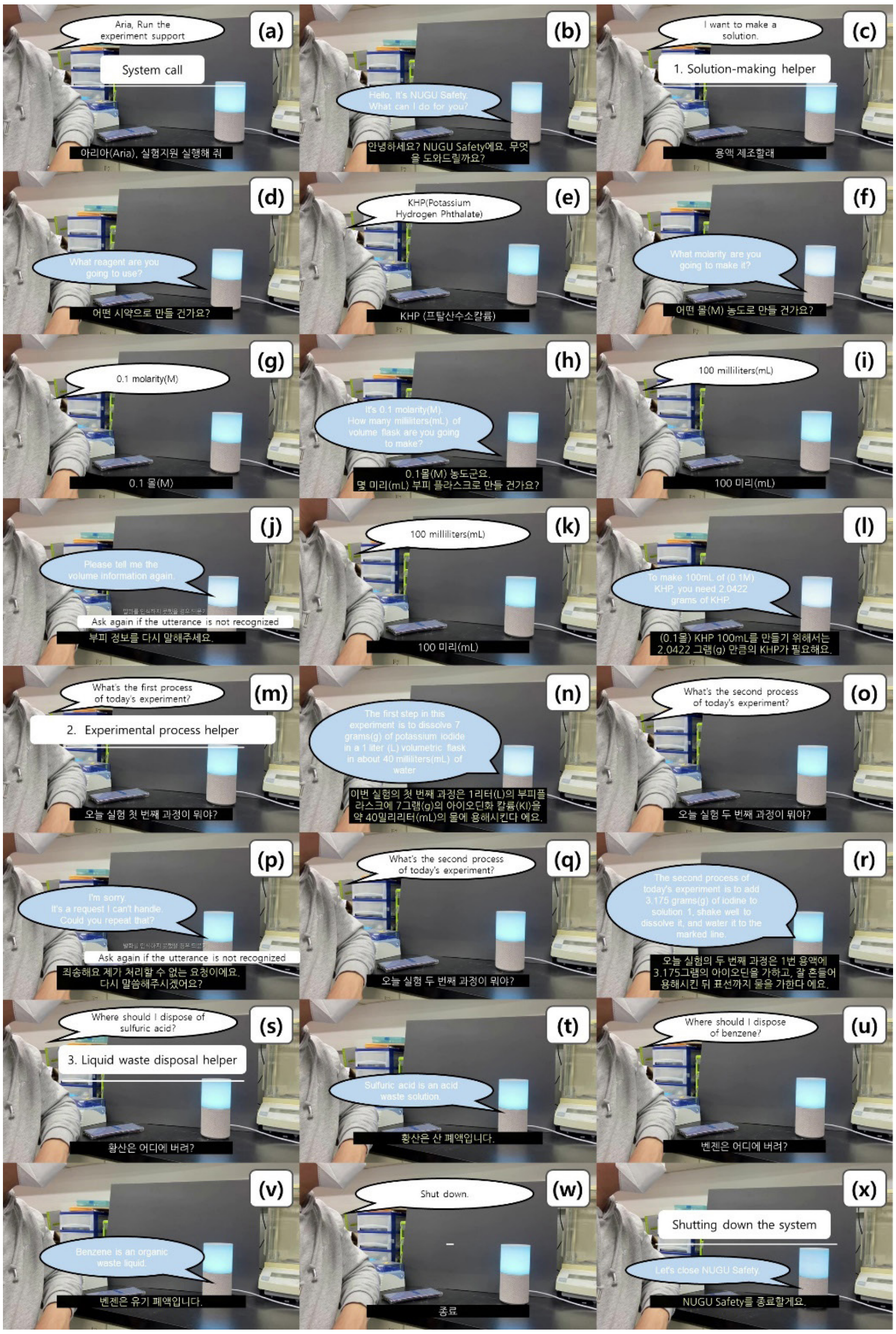

Fig. 4. Demonstration video of the prototype of a hands-free AI speaker system supporting hands-on science laboratory class

University students. University students evaluated the effectiveness of the prototype at 3.29 and efficiency 2.88, which were averaged to 3.09 in the user category. They scored the satisfaction of it at 3.18 and learnability 3.65, which were averaged to 3.41 in the task category. Finally, they scored the memorability of it at 3.59, error 2.76 and cognitive load 3.44, which were averaged to 3.26. The overall score of the quantitative survey from university students was 3.25, which is positive (> 2.5) on the 4-point Likert scale.

Suggestions for the prototype revision. In the interview, experts pointed out that the SR performance should be improved. Also, they suggested that more student questions should be recognizable to the speaker and further proposed that the scenario should be more elaborated for each function. Practically, they ideated that a brief introduction to what the speaker can provide to students is needed. Students similarly proposed processing more questions in the system, and the scenario should be more elaborated.

Consequently, the researchers were able to revise the AI speaker system before the external validation process, as follows: (1) The system awakening was simplified; (2) The number of error handling was increased for *function 1*; (2) Another entity-intent-action was added to *function 2* – providing additional tips not presented in the experiment manual (Figure 3); (3) Another entity was added to *function 3* to handle questions such as "where should I dispose of a solution after titrating I2?"; (4) More sentences were put to train the NLU model more precisely; and (5) The manual for using the AI speaker system was prepared.

4.3 External validation

Figure 5 shows the implemented AI speakers in the ACE course and the manuals provided to students.

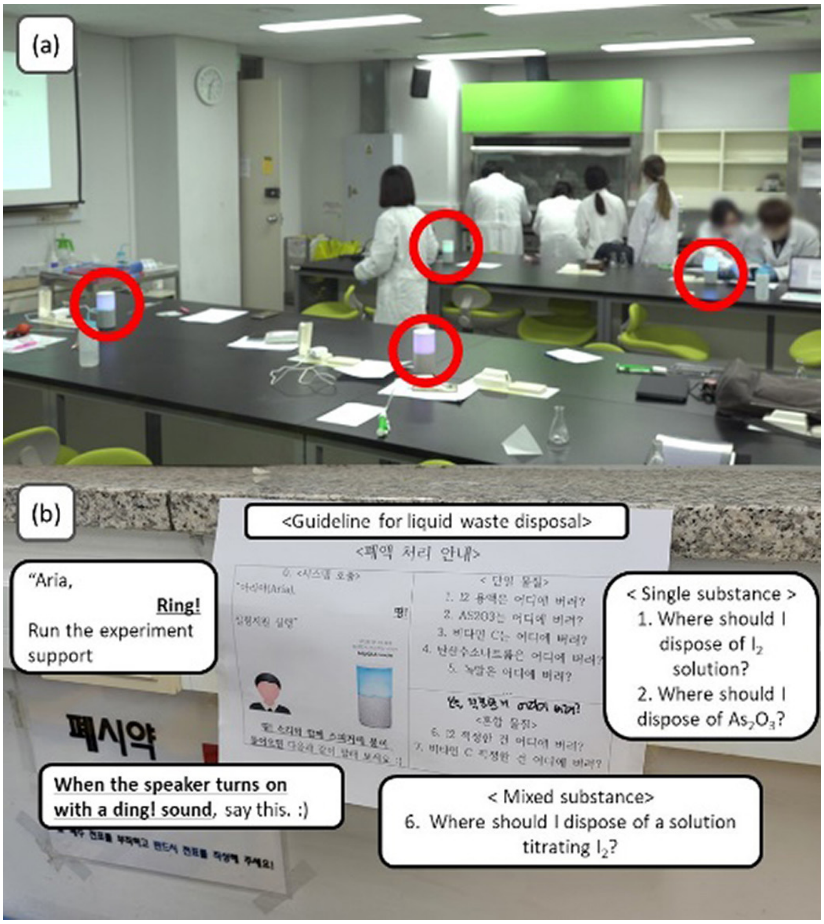

Fig. 5. AI speaker system installed in the science laboratory classroom (red circle) (a) and an example of the manual provided beside the speaker (b)

***Function 1* – solution-making helper.** At the beginning phase of the experiment, students asked the AI speaker system the amount of reagent needed to make a solution (Figure 6). First, a student called the system (Figure 6a) and requested a solution-making helper (Figure 6b). As the speaker did not catch what he had said, he bent toward the speaker to let it hear his voice clearer, which succeeded in calling *function 1* (Figure 6c). Then, the student and AI speaker made a conversation about the reagent, concentration, and volume of the solution to make a potassium permanganate solution, and consequently, the AI speaker gave the student answer to the question (Figure 6d–f). As the AI speaker was installed at the fume hood where the reagent needed to make a solution was kept, the student could make a solution according to the direction of the AI speaker.

Fig. 6. The use of AI speaker *function 1* – solution-making helper

***Function 2* – experimental process helper.** At the middle phase of the experiment, students asked the AI speaker system what they should notice in the procedure. At the 'titration in the non-aqueous solvent' experiment session (Table 2), a student called the system (Figure 7a) before asking the speaker about the indicator's color change at the titration endpoint. However, at first, the AI speaker system did not detect student's saying (Figure 7b). Although the student tried again, the speaker malfunctioned and played pop music, which made the other students laugh (Figure 7c and d). It was the last time she succeeded in calling the system and got the information – the indicator changes its color from purple to blue (Figure 7e and f).

Fig. 7. The use of AI speaker *function 2* – experimental process helper

***Function 3* – liquid waste disposal helper.** At the end phase of the experiment, students asked the AI speaker system in which container (acid, base, organic, or inorganic) to dispose of certain liquid waste. In Figure 8a, a student asked a colleague where to dispose of Arsenic trioxide (As_2O_3), and in response, she suggested that they would ask it to the AI speaker system (Figure 8a). She called the system, and it responded by asking her what waste liquid she wanted to dispose of (Figure 8b). When she answered the AI speaker concerning Arsenic trioxide, the speaker gave her appropriate information – i.e., it is an inorganic waste liquid. Then she wondered, making an exclamation ("Oh~") (Figure 8d).

Fig. 8. The use of *function 3* – liquid waste disposal helper

System Usability Scale (SUS). The overall SUS score scaled to 0–100 was 63.85 (Table 4). Bangor et al. [35] reported that the mean SUS score from about 3,500 surveys from 273 studies was 69.5. Therefore, the SUS score in this study is lower than the average. Consequently, this score lies in the range of "good" SUS score with "marginal" acceptability [35]. Also, items 1 and 10 satisfied benchmark scores for average SUS studies (≥ 3.39, ≤ 2.09, respectively), and item 7 satisfied that for industrial sense (≥ 4.19) [40]. These items show the strengths of the system developed in this study.

Table 4. The SUS quantitative survey result of external validation (N = 13) (1–5 scale)

Item No.	Question	Mean [SD]
1	I think that I would like to use this product frequently.[p]	3.69 (.85)*
2	I found the product unnecessarily complex.[n]	2.54 (1.05)
3	I thought the product was easy to use.[p]	3.62 (.77)
4	I think that I would need the support of a technical person to be able to use this product.[n]	2.77 (1.17)
5	I found the various functions in the product were well integrated.[p]	3.46 (.78)

(Continued)

Table 4. The SUS quantitative survey result of external validation (N = 13) (1–5 scale) *(Continued)*

Item No.	Question	Mean [SD]
6	I thought there was too much inconsistency in this product.[n]	2.62 (1.04)
7	I imagine that most people would learn to use this product very quickly.[p]	4.31 (.75)**
8	I found the product very awkward to use.[n]	2.54 (1.05)
9	I felt very confident using the product.[p]	3 (.82)
10	I needed to learn a lot of things before I could get going with this product.[n]	2.08 (1.12)*
Overall SUS score		63.85

Notes: p: positive statements, n: negative statements; *: satisfies an average item benchmark; **: satisfies a common industrial item benchmark

In an open-ended question aligned with the SUS, students responded that the AI speaker system was "easy to use," "easy to learn," "convenient," "simple," "intuitive," or "not complex nor difficult," which elaborates their response to the item no. 3, 7, and particularly 10. Therefore, it seems that the perceived ease of usage led to students' willingness to use the AI speaker system (item 1).

Meanwhile, students responded that there are some weaknesses in the system, which are related. (1) The inflexible structure (plot) of the "command" or conversation "as written in the manual" and the number of questions they should articulate was inconvenient (seven students). (2) The SR accuracy of the AI speaker was problematic, which led to the incorrect reaction of the system (e.g., not responding, playing pop music).

Perceptions of an AI Speaker System in a Science Laboratory Classroom (PASS-SLC). The overall average score of PASS-SLC items was 3.35 (Table 5). Students responded that they could ask the speaker how they usually speak (M = 3.46) – however, the speaker did not understand their intentions well (M = 2.46). Notably, the answers from the AI speaker was easy to understand (M = 4.31), with correct information (M = 3.85), consistently (M = 4.15), and reliably M = (4.08). However, the system was not convenient compared to calculators or smartphones (M = 2.69). Although they responded that the laboratory class became safer due to the AI speaker system (M = 3.62), it was not comfortable to use the system compared to asking friends or TA (M = 2.62). They perceived that using the system did not help them gain more scientific knowledge (M = 2.23). However, the same question for the scientific skills scored higher (M = 2.85), and the question for the attitudes toward science class got a positive response (M = 3.46). It seems that it was not due to the change in the laboratory class process (M = 2.77), the roles of instructors' (M = 2.92) or students' (M = 3), nor the reduced performance time (M = 2.77). Rather, their fun (M = 4.46) and enjoyable (M = 4.54) user experience with the AI speaker system probably led to their increased perception of attitude toward science class.

Table 5. The PASS-SLC quantitative survey result of external validation (N = 13) (1–5 scale)

Item No.	Question	Mean [SD]
1	I was able to ask the speaker in the way I usually speak.	3.46 (0.88)
2	The speaker understood the intention of my question well.	2.46 (0.66)
3	The answer of the speaker was easy to understand.	4.31 (0.63)
4	The speaker told the correct information needed for the laboratory class.	3.85 (0.99)
5	The information the speaker told was consistent.	4.15 (0.99)
6	The information the speaker told was reliable.	4.08 (0.95)
7	The system was more convenient for getting the information I wanted than using calculators or smartphones.	2.69 (1.03)
8	The laboratory class became safer than before through this system.	3.62 (0.87)
9	Using the system was more comfortable than asking friends or TA.	2.62 (1.26)
10	As a result of using the system, I could gain more scientific knowledge.	2.23 (0.73)
11	As a result of using the system, I could gain more scientific skills.	2.85 (1.14)
12	As a result of using the system, I could gain more attitudes toward science class.	3.46 (1.2)
13	The process of laboratory classes using the system became different from before.	2.77 (1.24)
14	The role of the instructor has changed in the laboratory class using the system.	2.92 (1.04)
15	The role of the students has changed in the laboratory class using the system.	3 (0.91)
16	The laboratory performance time has been reduced when using the system.	2.77 (1.09)
17	The user experience of the system was fun.	4.46 (0.66)
18	The user experience of the system was enjoyable.	4.54 (0.52)
Overall average of PASS-SLC items		3.35 (1.17)

In an open-ended question aligned with the PASS-SLC, students responded that rather than asking the system, "asking TA" or "using a calculator" to get information would be more "convenient," "fast," and "correct" (items 7 and 9). Also, four students problematized the structured way of asking the speaker as "complex" and "took time" (item 16). The reason for this was the SR accuracy (four students). And three responded that the AI speaker system would be useful when the SR accuracy increases.

Pros, cons, and suggestions for revision. Students pointed out the pros of the AI speaker system implemented in the classroom as follows: First, five out of thirteen acknowledged the high "accessibility" of the system, which could be used "easily" "whenever" they wanted, and this feature drew students' interest (two students) (items 1, 12, 17, and 18 in the PASS-SLC; Table 5). Second, three students appreciated that the system had freed their hands, lest they repeatedly put on and off the gloves to touch things other than experimental, which supports keeping safety (one student). Further, three students responded that the workload of the TA has been reduced due to the AI speaker system.

5 Discussion

5.1 Research question 1

The ADDIE process of developing and validating the hands-free AI speakers supporting hands-on science laboratory classes shown throughout this study answers research question 1.

It should be noted that the available technology largely shaped the characteristics of the prototype. First, the AI speaker system developed in this study is 'explainable.' As explainability is becoming increasingly important for ethics in AIEd [7–8], the characteristic of AISEd should be appreciated and further pursued. Meanwhile, one of the reasons for non-satisfying SR accuracy is attributed to the current state of the Korean NLP (cf. [41]), which is lower than that of the English. Thus, we can anticipate different responses if the equivalent research was conducted in English-speaking situations or any other context, which calls for further study [5].

5.2 Research question 2

The responses of experts and university students had commonalities and differences in the PACMAD items. The categories that scored highest and lowest were similar. For example, the learnability, memorability, and cognitive load categories took the first to third ranks in both groups (Table 4). The issue of explainability mentioned above can be connected to the learnability and memorability in the PACMAD. Also, reducing the cognitive load is the main contribution of AIEd [7–8]. These signify that the AI speaker system satisfied those criteria, possibly leading to students' concentration on hands-on experimentation. Further, the experts and students said the error is the most concern in the PACMAD in parallel, providing corresponding suggestions for revising the prototype. These showed the possibilities and current status of the AI speaker system in the science laboratory.

However, there were some differences between the perceptions of experts and students. The experts' PACMAD scores were higher in every category and thus overall items compared to those of students. Particularly, the efficiency showed the most gap between experts (3.44) and students (2.88) (Table 4), which was backed up by the open-ended questions. For example, experts said, "an AI speaker is efficient because it answers to the query immediately," "hands must be used in the laboratory environment in general, but one can get an answer for his/her question by just an articulation," "if it can be realized, it would be definitely helpful for the safe experimentation and would reduce time," and "it adds to efficiency because one can ask for information and get an answer." On the contrary, students said that "it takes a too long time to get an answer," "one may feel jammed up because it takes a long time asking and getting an answer," and "I think it is very good and positive, but it seems to have a long way to go."

Reflecting on the differences, it seems that the experts' background experiences made them evaluate the prototype positively. Meanwhile, students may have imagined the realistic situation the AI speaker is implementing in their laboratory classroom, making them not give high scores. Future research that scrutinizes why there are disparities in the perceptions of introducing AI (speaker) into a classroom between educational experts and students would be meaningful.

5.3 Research question 3

First, the mobile characteristics of the AI speaker enabled its optimal use according to each function – from the fume hood, laboratory table, or near the liquid waste container. Second, it is manifest that the unsafe situations in that students touch non-experimental things with their gloved hands reduced significantly. It inherently leads to the third strength, the incorporation of another learning mode – i.e., auditory one – into the science laboratory class to support students' experimentation. Much literature has shown that visual (V), kinesthetic (K), and reading/writing (R) learning modes are important in science laboratories in the hands-on inquiry, minds-on inquiry, and lab report writing [12–15]. However, almost no research seems to have pointed out that the auditory (A) learning mode can be utilized in a science laboratory class, which complements the VARK multi-modal learning [42]. This implies that the implementation of the AI speaker in the lab raised a significant, noteworthy theoretical thesis for the science education field.

The interpretation of the SUS score needs caution. According to Bangor et al. [43], while the mean SUS score of interactive voice response (IVR) systems was 73.84 (N = 401), that of a combinatory Web/IVR system was 59.45 (N = 50). Further, recent studies on the usability of Amazon Alexa showed a score of 63.69 (N = 61) [39]. These indicate that the SUS score of this study (63.85) is intermediate as an AI speaker. Therefore, according to the SUS score, the product of this study shows promises and suggests future research implications of AI speakers supporting hands-on science laboratories, although the immediate industrial or practical implications might be few.

The result of the PASS-SLC survey shows that the most strength of the AI speaker supporting the science laboratory class can be its affordance to draw students' interest in a science laboratory class. In the PASS-SLC, students responded that they did not directly gain more knowledge or skills through the system – however, if we expect the affective domain to lead to active engagement in laboratory activities, the development of knowledge and skills might follow in a long-term sense.

Manifestly, the most urgent problem for the suggestions for future revision is improving the AI speaker's SR accuracy (in Korean) to catch students' natural language articulations. It would enable more functions and a flexible conversation structure of AISEd optimized in various situations, reducing resources spent on error handling.

However, the instructional designers shall not just rely on the advancement of technologies but contemplate the possible remedies for problems listening from the learners. University students, who were pre-service chemistry teachers, suggested both increasing SR accuracy and allowing flexible conversations. At first glance, it seems contradictory and has a trade-off relationship. However, students suggested the keyword-based conversation for the AI speaker system in the external validation. It seems possible to catch two birds with one stone, even with the current technologies. Research focusing on the type of conversation (natural language-based versus keyword-based) in AISEd should follow.

6 Conclusion

Using the RP method, we designed and developed a hands-free AI speaker-based system that answers student queries concerning solution-making, experimental processes, and waste liquid disposal, corresponding to the initial, middle, and final phases of a hands-on laboratory class. The system was internally validated by usability tests of 9 expert panels and 18 university students and then revised. The revised system was externally validated in an analytical chemistry experiment class for 3 sessions with 13 university students. We presented the result of the prototype development and internal and external validations with quantitative and qualitative data. The AI speaker system enabled students to use the auditory learning mode in the laboratory while concentrating on the experimentation with their hands in the external validation. It can be considered as opening a channel for the auditory learning mode in science laboratory class to the previous visual, reading/writing, and kinesthetic modes to complement the VARK multi-modal learning.

One of the novelties of this case study lies in its research field – a science laboratory classroom filled with Korean-speaking learners. Ironically this leads to the need for more generalizable instructional knowledge (see [18]). Although this study carefully designed, developed, and implemented an AI speaker system for the science laboratory class, it does not provide comprehensive principles, guidelines, or models for future designers of AI speakers used in a classroom. Therefore, research that presents design principles for or model of AISEd in authentic teaching and learning sites should follow.

7 Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1058192).

8 References

[1] J. Huang, G. Shen, and X. Ren, "Connotation analysis and paradigm shift of teaching design under artificial intelligence technology," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 5, pp. 73–86, 2021. https://doi.org/10.3991/ijet.v16i05.20287

[2] S. Russel and P. Norvig, Eds., *Artificial Intelligence: A Modern Approach*. London: Pearson, 2021.

[3] V. Kepuska and G. Bohouta, "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)," In proc. IEEE annual computing and communication workshop and conference '08, 2018, pp. 99–103. https://doi.org/10.1109/CCWC.2018.8301638

[4] B. K. Prahani, I. A. Rizki, B. Jatmiko, N. Suprapto, and T. Amelia. "Artificial intelligence in education research during the last ten years: A review and bibliometric study," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 8, pp. 169–188, 2022. https://doi.org/10.3991/ijet.v17i08.29833

[5] N. Pinkwart, "Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future," *International Journal of Artificial Intelligence in Education*, vol. 26, no.1, pp. 771–783, 2016. https://doi.org/10.1007/s40593-016-0099-7

[6] X. Chen, D. Zou, H. Xie, and G. Cheng, "Twenty years of personalized language learning," *Educational Technology & Society*, vol. 24, no. 1, pp. 205–222, 2021.
[7] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson, 2016.
[8] V. Srinivasan, "AI & learning: A preferred future," *Computers and Education: Artificial Intelligence*, vol. 3, 100062, 2022. https://doi.org/10.1016/j.caeai.2022.100062
[9] S. B. Lovato, A. M. Piper, and E. A. Wartella, "Hey Google, do unicorns exist? Conversational agents as a path to answers to children's questions," In Proc. ACM International Conference on Interaction Design and Children '18, 2019, pp. 301–313. https://doi.org/10.1145/3311927.3323150
[10] S. Chung and B. K. Woo, "Using consumer perceptions of a voice-activated speaker device as an educational tool," *JMIR Medical Education*, vol. 6, no. 1, e17336, 2020. https://doi.org/10.2196/17336
[11] K. Pietroszek, "Providing language instructor with artificial intelligence assistant," *International Journal of Emerging Technologies in Learning*, vol. 2, no. 4, pp. 61–65, 2007.
[12] A. Hofstein and V. N. Lunetta, "The laboratory in science education: Foundations for the twenty-first century," *Science Education*, vol. 88, no. 1, pp. 28–54, 2004. https://doi.org/10.1002/sce.10106
[13] L. B. Flick, "The meanings of hands-on science," *Journal of Science Teacher Education*, vol. 4, no. 1, pp. 1–8, 1993. https://doi.org/10.1007/BF02628851
[14] D. S. Domin, "A review of laboratory instruction styles," *Journal of Chemical Education*, vol. 76, no. 4, pp. 543–547, 1999. https://doi.org/10.1021/ed076p543
[15] N. Reid and I. Shah, "The role of laboratory work in university chemistry," *Chemistry Education Research and Practice*, vol. 8, no. 2, pp. 172–185, 2007. https://doi.org/10.1039/B5RP90026C
[16] A. D. Ménard and J. F. Trant, "A review and critique of academic lab safety research," *Nature chemistry*, vol. 12, no. 1, pp. 17–25, 2020. https://doi.org/10.1038/s41557-019-0375-x
[17] D. L. Sewell, "Laboratory safety practices associated with potential agents of biocrime or bioterrorism," *Journal of Clinical Microbiology*, vol. 41, no. 7, pp. 2801–2809, 2003. https://doi.org/10.1128/JCM.41.7.2801-2809.2003
[18] R. C. Richey and J. D. Klein, *Design and Development Research: Methods, Strategies, and Issues*. Thames, Oxfordshire: Routledge, 2007.
[19] R.C. Richey, J. Klein, and W. Nelson, "Developmental research: Studies of instructional design and development," in *Handbook of Research for Educational Communications and Technology*, D. Jonassen, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2004, pp. 1099–1130.
[20] S. D. Tripp and B. Bichelmeyer, "Rapid prototyping: An alternative instructional design strategy," *Educational Technology Research and Development*, vol. 38, no. 1, pp. 31–44, 1990. https://doi.org/10.1007/BF02298246
[21] C. Lim, Y. Song, S. Hong, and C. Park, "A study on the applications and improvement of the rapid prototyping to instructional systems design (RPIsd) model," *Journal of Educational Technology*, vol. 36, no. 3, pp. 589–617, 2020. [in Korean] https://doi.org/10.17232/KSET.36.3.589
[22] T. S. Jones and R. C. Richey, "Rapid prototyping methodology in action: A developmental study," *Educational Technology Research and Development*, vol. 48, no. 2, pp. 63–80, 2000. https://doi.org/10.1007/BF02313401
[23] Argonne National Lab (ANL), *AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science*. Argonne, IL: United States, 2020.
[24] F. M. Dunnivant, D. M., Simon, and S. Willson, "The making of a solution: A simple but poorly understood concept in general chemistry," *The Chemical Educator*, vol. 7, no. 4, pp. 207–210, 2002. https://doi.org/10.1007/s00897020581a

[25] V. R. Ralph and S. E. Lewis, "Chemistry topics posing incommensurate difficulty to students with low math aptitude scores," *Chemistry Education Research and Practice*, vol. 19, no. 3, pp. 867–884, 2018. https://doi.org/10.1039/C8RP00115D
[26] P. L. Gentili, "Small steps towards the development of chemical artificial intelligent systems," *RSC Advances*, vol. 3, no. 48, pp. 25523–25549, 2013. https://doi.org/10.1039/c3ra44657c
[27] L. B. Ayres, F. J. Gomez, J. R. Linton, M. F. Silva, and C. D. Garcia, "Taking the leap between analytical chemistry and artificial intelligence: A tutorial review," *Analytica Chimica Acta*, 1161, 338403, 2021. https://doi.org/10.1016/j.aca.2021.338403
[28] U. G. Jack, "Analysis of senior secondary school students' experienced difficulty in science process skills acquisition in chemistry," PhD. Thesis, Delta State University, Abraka, 2012.
[29] D. F. Treagust, K. C. D. Tan, N. K. Goh, and L. S. Chia, "Major sources of difficulty in students' understanding of basic inorganic qualitative analysis," *Journal of Chemical Education*, vol. 81, no. 5, p. 725, 2004. https://doi.org/10.1021/ed081p725
[30] M. Stieff, S. M. Werner, B. Fink, and D. Meador, "Online prelaboratory videos improve student performance in the general chemistry laboratory," *Journal of Chemical Education*, vol. 95, no. 8, pp. 1260–1266, 2018. https://doi.org/10.1021/acs.jchemed.8b00109
[31] N. Choudhary, R. Bharti, and R. Sharma, "Role of artificial intelligence in chemistry," *Materials Today: Proceedings*, vol. 48, no. 5, pp. 1527–1533, 2021. https://doi.org/10.1016/j.matpr.2021.09.428
[32] Z. J. Baum, X. Yu, P. Y. Ayala, Y. Zhao, S. P. Watkins, and Q. Zhou, "Artificial intelligence in chemistry: Current trends and future directions," *Journal of Chemical Information and Modeling*, vol. 61, no. 7, pp. 3197–3212, 2021. https://doi.org/10.1021/acs.jcim.1c00619
[33] L. He, L. Bai, D. D. Dionysiou, Z. Wei, R. Spinney, C. Chu, ... and R. Xiao, "Applications of computational chemistry, artificial intelligence, and machine learning in aquatic chemistry research," *Chemical Engineering Journal*, vol. 426, 131810, 2021. https://doi.org/10.1016/j.cej.2021.131810
[34] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: Literature review and rationale for a new usability model," *Journal of Interaction Science*, vol. 1, no. 1, pp. 1–16, 2013. https://doi.org/10.1186/2194-0827-1-1
[35] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.
[36] K. Sohn and O. Kwon, "Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products," *Telematics and Informatics*, 47, 101324, 2020. https://doi.org/10.1016/j.tele.2019.101324
[37] J.-m. Lee, M. Jung, J. Lee, Y.-e. Kim, and C. An, "Consumer perception and adiption intention of artificial intelligent speaker: Nun-users perspective," *Journal of Consumer Studies*, vol. 30, no. 2, pp. 193–213, 2019. [in Korean] https://doi.org/10.35736/JCS.30.2.9
[38] Y. Min, J. Ahn, and S. Kim, "Structural relationship analysis between the intention to use educational chatbots and influential factors based on technology acceptance model," *The Journal of the Education Information and Media*, vol. 26, no. 3, pp. 799–825, 2020. [in Korean] https://doi.org/10.15833/KAFEIAM.26.4.799
[39] D. S. Zwakman, D. Pal, and C. Arpnikanondt, "Usability evaluation of artificial intelligence-based voice assistants: The case of Amazon Alexa," *SN Computer Science*, vol. 2, no. 1, pp. 1–16, 2021. https://doi.org/10.1007/s42979-020-00424-4
[40] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale," *Journal of Usability Studies*, vol. 13, no. 3, pp. 158–167, 2018.
[41] J. Kim, E. Shin, K. Han, S. Park, J. H. Youn, G. Jin, and J. Y. Lee, "Efficacy of smart speaker–based metamemory training in older adults: Case-control cohort study," *Journal of Medical Internet Research*, vol. 23, no. 2, e20177, 2021. https://doi.org/10.2196/20177

[42] N.D. Fleming, and C. Milss, "Not another inventory, rather a catalyst for reflection," *To Improve the Academy*, vol. 11, pp. 137–155. https://doi.org/10.1002/j.2334-4822.1992.tb00213.x
[43] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human–Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008. https://doi.org/10.1080/10447310802205776

9 Authors

Gyeong-Geon Lee is a Ph.D. candidate in the Department of Chemistry Education, College of Education, Seoul National University. He holds a B.S. (Chemistry Education) and a B.E. (Computer Science & Engineering) from Seoul National University. His research interests pertain to various fields of educational studies, from history and philosophy of science education, curriculum studies concerning competencies, science teaching and learning focusing on inquiry and laboratory, educational technology focusing on instructional design, and artificial intelligence in education for automated assessment and classroom environment. He seeks to incorporate various approaches to science education into a comprehensive knowledge base (crusaderlee@snu.ac.kr).

Minji Choi is a master's student in the Department of Chemistry Education, College of Education, Seoul National University. She holds a B.S. (Chemical Engineering) from University of Seoul. Her research interests pertain to non-face-to-face experimental classes, game-based learning, and school science. She studies how students can have positive experiences with science classes and improve science academic emotion (minji0102@snu.ac.kr).

Taesoo An is a master's student in the Department of Chemistry Education, College of Education, Seoul National University. He has educational experience in developing countries, including Korea. He is interested in the education that all students are interested in, and the core competencies that must be learned in the future. His field of research is interested in the universal application and expansion of scientific inquiry methods beyond the scientific field (obrigado@snu.ac.kr).

Seonyeong Mun is a master's student in the Department of Chemistry Education, College of Education, Seoul National University. She holds a B.S. (Chemistry Education) from Seoul National University. Her research interests pertain mainly to history and philosophy of science and science education, science teaching and learning focusing on inquiry and laboratory. She seeks to how students can participate in science classes more meaningfully, construct and acquire science knowledge in the context of science (suyo2478@snu.ac.kr).

Hun-Gi Hong is a Professor in the Department of Chemistry Education at the College of Education, Seoul National University, Seoul, Republic of Korea. He finished his PhD in Analytical Chemistry from the University of Texas at Austin and postdoctoral fellowship at the Beckman Institute of Science and Technology, University of Illinois at Urbana-Champaign. He has been a Chemistry Professor at Sejong University and a visiting professor at Iowa State University. His research interests include various fields of analytical electrochemistry, chemistry education, gifted student education, and educational technology (hghong@snu.ac.kr).

Article submitted 2022-08-22. Resubmitted 2022-10-29. Final acceptance 2022-10-29. Final version published as submitted by the authors.