Effects of Question Type Presentation on Raised Questions in a Video Learning Framework

Hinako Izumi[1(✉)], Masaki Matsubara[2], Chiemi Watanabe[3], and Atsuyuki Morishima[2]

[1] Master's Programs in Informatics, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-2 Kasuga, Tsukuba-shi, Ibaraki 305-8550, Japan
hinako.izumi.2019b@mlab.info

[2] Faculty of Library, Information and Media Science, University of Tsukuba, 1-2 Kasuga, Tsukuba-shi, Ibaraki 305-8550, Japan
{masaki,mori}@slis.tsukuba.ac.jp

[3] Faculty of Industrial Technology, Tsukuba Univesity of Technology, Amakubo 4-3-15, Tsukuba, Ibaraki 305-8520, Japan
chiemi@a.tsukuba-tech.ac.jp

Abstract. This paper examines the effect of explicitly presenting types of questions to students in an online learning environment. In our setting, people are shown online videos for lectures and then asked to raise questions. We compare the collected questions with those collected in the same situation but without presenting types of questions. We find that presenting types of questions improves the quality of questions, but does not increase the number of questions raised.

Keywords: Learning framework · Question prompt · Facilitation

1 Introduction

Understanding what we do not understand is an important step for learning, and one of the most difficult things for beginners [10,14]. One technique for helping students understand what they do not understand is presenting types of questions to prompt the students to raise questions. In offline classes, presenting types of questions was shown to be effective for learners to raise questions [9]. In their settings, facilitators present types of questions when they prompt questions from students.

This paper examines the effect of explicitly presenting types of questions to students in an online video learning environment.

Our challenge is to develop a *fully automated workflow* to prompt questions with presenting types of questions for online classes. The focus of this paper is whether presenting types of questions to students is still effective in such a setting. In our setting, students watch online videos for lectures (Fig. 1). The

K. Toeppe et al. (Eds.): iConference 2021, LNCS 12646, pp. 114–126, 2021.
https://doi.org/10.1007/978-3-030-71305-8_9

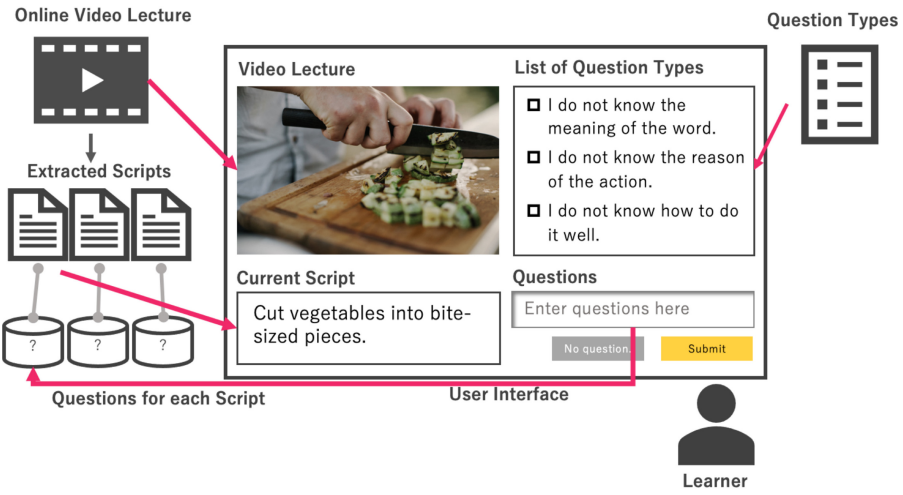

Fig. 1. An automatic framework for prompting questions with presenting types of questions for online classes

process of prompting questions is fully automated as follows. First, the script of the video lecture is extracted and divided into chunks of sentences in advance[1]. Second, the video lecture is played in the screen, and scripts are shown in synchronization with the video. A list of question types is shown with checkbox to help the learner come up with their question. An example of question type is "I do not know the meaning of the word." There is a text field for typing a question with two buttons "submit" and "no question." The video does not move to the next chunk if none of the two buttons are pressed before the current chunk ends.

This paper focuses on the following two research questions (Fig. 2). The first question is whether showing types of questions increases the number of raised questions or not, and the second one is whether it affects the quality of questions or not.

To obtain the answers to the two questions, we conduct an experiment in a crowdsourcing setting. We hire crowd workers and ask them to raise questions on one of two online lectures. We compare the collected questions with those collected in the same situation but without presenting types of questions. We find that presenting types of questions improves the quality of questions.

The contributions of this paper are as follows.

(1) It shows that presenting types of questions is effective in terms of the quality of raised questions, in an automatic framework that does not require human facilitators' intervention, for prompting questions on online lectures from learners.
(2) It gives the result of a large scale experiment on the framework with hundreds of people recruited from a crowdsourcing service.

[1] In the simplest way, each chunk is composed of one sentence.

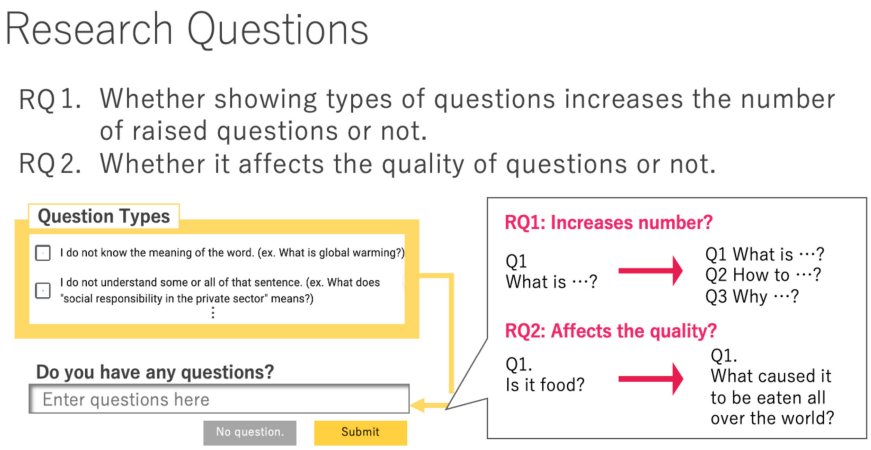

Fig. 2. Research questions

The remainder of this paper is as follows. Section 2 introduces related work. Section 3 explains the experiment and its result. Section 4 gives a discussion. Section 5 concludes.

2 Related Work

Our challenge is to develop an automatic and effective way to prompt questions in an online learning environment. However, effective ways to prompting questions is not straightforward. Miyake et al. [13] showed one of the reasons students have difficulty eliciting questions is that they do not understand what they do not understand. Therefore, we have been exploring an approach to promote metacognition of students. In our previous work [8], we found that when eliciting questions, having them select parts of the script that they do not understand elicits more specific questions. We adopted the task design used in [8] that we found effective.

Existing approaches to facilitating learning include peer review of writing [11], having students solve problems created by each other [6,7,12], and supporting the process of recognizing and solving complex problems [1,4]. In particular, studies that focus specifically on eliciting questions from students include King [9] and Endo [3]. Both studies showed an improvement in performance as a result of prompting question types and other students' existing questions. In addition, [9] showed an increase in the number of some types of questions, too. Since it has been shown that promoting questioning is an effective approach to aid learning, we try to incorporate it into an automatic online workflow.

CoNet-C [2] is similar to our approach in that it is an online scaffolding tool for discussion that gives students several types of questions in each of three patterns of clarification, rebuttal, and viewpoint change so that they can ask questions about each student's opinions. CoNet-C allows students to individually

ask their classmates. Since its focus was on a discussion situation, it was assumed that the questions would be about the opinions of each student. Our study differs from the CoNet-C in that ours aims to promote questions on the content of the class in a general classroom situation where the teacher gives a lecture to students.

3 Experiment

We conducted an experiment in which students were asked questions about a video lecture's content to examine the following research questions (Fig. 2).

RQ1: Whether showing types of questions increases the number of raised questions or not.
RQ2: Whether it affects the quality of questions or not.

In the experiment, we provided the online video and its script to the participants, and asked them to list up pairs of their question and the place (chunk) to ask their question.

For investigating the research questions, we applied the following two types of question collecting processes to compare the number and the qualities of questions.

- When collecting the questions, we show the types of general questions.
- When collecting the questions, we do not show the question types.

The detailed procedure is described in Sect. 3.3. The IRB approved the experiment at the Faculty of Library, Information, and Media Science at the University of Tsukuba.

Summary of Results: RQ1 did not support, and RQ2 did support.

3.1 Teaching Materials

The topic of teaching materials were a cooking recipe video[2] (COOK) and TED Talks[3] (social issue, SI). Table 1 shows title, video length, and number of chunks of each video. We extracted a script from the subtitles in the video, and divided it into chunks according to the timing of the screen transitions in the video. We used that video because the script of the video content was available and that it was the right time length for the crowd workers to engage.

[2] "How to make boiled and soaked aubergines" https://www.youtube.com/watch?v=27sVEYPlomw.
[3] "The evolution of the coffee cup lid" https://www.ted.com/talks/a_j_jacobs_the_evolution_of_the_coffee_cup_lid.

Table 1. Details of teaching materials

Domain	Title	Video length	Number of chunks
COOK	How to make boiled and soaked aubergines	00:03:42	17
SI	The evolution of the coffee cup lid	00:03:02	8

3.2 Participants

A total of 400 participants (100 in each condition and domain) were recruited via Yahoo! Crowdsourcing. We didn't set any requirements for the age or gender, but the workers can work on a PC for the environment.

To investigate effectiveness of presenting question type, the participants were divided into two groups: an experimental group and a control group. In the experiment group, the question type was shown when the participants asked the questions, whereas it was not in the control group.)

Before watching the online video, participants were shown a table (Table 2) to explain the skill levels for each domain and asked to assess and declare their skills based on the table.

Table 2. Criteria for self assessment of skill level in each domain

Level	COOK	SI[a]
1	I've barely ever cooked	I watch or read the news at least once a month for a total of 5 to 30 min
2	I may cook a simple meal once a week to once a month	I watch or read the news at least once a month for a total of 30 min or more
3	I may cook an elaborate meal once a week to once a month	I watch or read the news at least once a week for a total of 5 to 30 min
4	Cook a simple meal almost every day	I watch or read the news at least once a week for a total of 30 min or more
5	I cook an elaborate meal almost every day	I watch or read the news for a total of 5 to 30 min almost every day
6	I have worked in a cooking job for more than a month	I watch and read the news for a total of 30 min or more almost every day

[a] The content of TED talks is one of social issues. Since many news and newspapers also deal with social issues, people who read news frequently are likely to be familiar with understanding social issues. For this reason, we used this index.

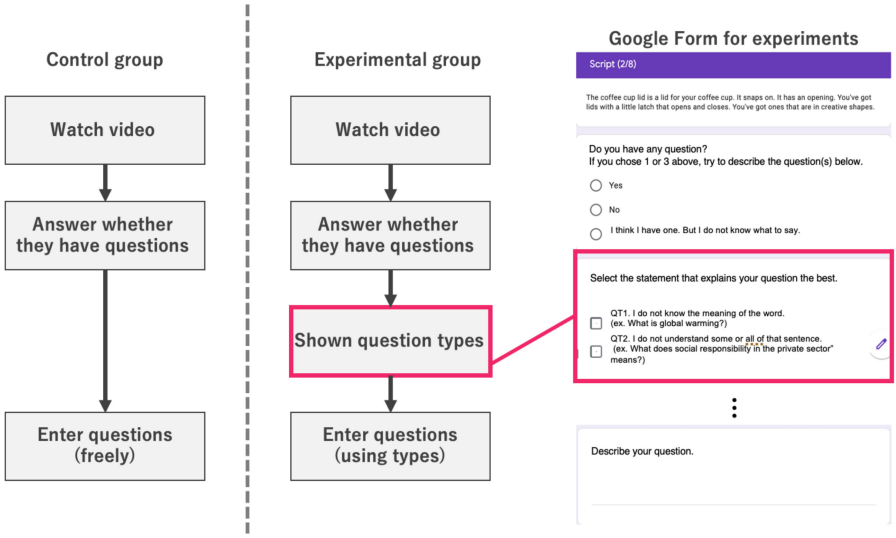

Fig. 3. Experimental procedure: First, participants watch a video. Second, they are asked whether they have questions. If they have a question, they enter it. In the experimental group, with type presentation only in the experimental group.

3.3 Procedure

Figure 3 shows the experimental procedure. First, the participants were asked to watch the video, then the chunk of segmented scripts was presented to them in turn. They were instructed to ask any questions if they have at each chunk. In asking questions, we used Google Form. Each section has one chunk and a text area to write questions down if they have.

While the control group was asked questions without presenting types, the experimental group was presented with types in each section of the form. It was asked to choose the one that matches the content of the question they wanted to ask and report it to the participants. The purpose of presenting the types is to have the participants use the presented types as hints to raise their own questions.

Question Types. The question types were designed to present questions with roughly the same meanings in COOK and SI, based on the questions on the mathematics domain collected in [8]. Table 3 shows question types and example questions in two domains. In the table, the "type" column explain what are asked in question types, and "Examples in COOK" and "Examples in SI" columns show examples of the question types in the cooking and social issues domain we used for our two videos COOK and SI, respectively.

Table 3. The types of questions presented to the participants in the experimental group

No	Type	Examples in COOK	Examples in SI
QT1	I do not know the meaning of the word	What is flambé?	What is global warming?
QT2	I do not know the specific numbers about it	How much soy sauce should I add?	How many years is there a sufficient embargo on fishing?
QT3	I want to know if I can substitute it for something else	Can I use something instead of sugar?	Does gardening at home work as an alternative to planting trees?
QT4	I want to make sure my understanding	Does that mean A?	
QT5	I do not understand some or all of that sentence	What does "mixing like cutting" means?	What does "social responsibility in the private sector" means?
QT6	I do not know how to do it well	I get bubbles in my pudding, how can I remedy this?	How does the dam regulate the volume of water?
QT7	I want to propose something	How about doing A?	
QT8	I do not know why you do it	Why do you let that food sit in the fridge?	Why do we need to separate our waste?
QT9	I do not know why you say "A, so B."	"Radishes can fall apart in boiling, so cut the corners." Why do you say so?	"Sea surface temperatures will rise, so dwellings could be flooded." Why do you say so?
QT10	A question that is none of the above	–	

Table 4. Number of subjects and questions in each conditions. In both domains, the number of questions did not increase or decrease despite the type presentation.

	COOK		SI	
	Control	Experimental	Control	Experimental
Valid answers	98	98	100	100
Questions	289	289	222	218

3.4 Result

Table 4 shows the numbers of valid answers (participants) and the numbers of questions raised by participants in each domain. In the COOK domains, the number of raised questions was the same (289). In SI, 222 questions were raised by participants in the control group and 218 in the experimental group.

Table 5 shows the distributions of skill levels of participants. In COOK, the majority of participants were either beginners (Levels 1, 2) or moderate (4), while in SI, they were mostly skilled workers (5, 6).

Table 5. Self-reported skill of subjects. Many participants reported skill level 1 or 2 or 4 in COOK domains and skill level 5 or 6 in SI domains.

Skill level	COOK		SI	
	Control	Experimental	Control	Experimental
1	19	11	6	3
2	32	37	1	0
3	6	5	3	3
4	32	35	4	2
5	6	7	42	51
6	3	3	44	41

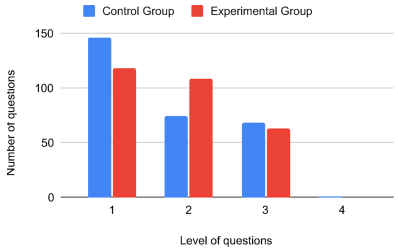

Fig. 4. Level of questions in COOK domain. The presentation of the type reduced the number of skill level 1 questions and increased the number of skill level 2 questions ($p = 0.01468$).

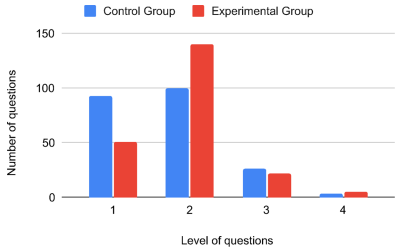

Fig. 5. Level of questions in SI domain. The presentation of the type reduced the number of level 1 questions and increased the number of level 2 questions ($p = 0.00019$).

Figures 4 and 5 show the distributions of *questions levels* raised by participants in the COOK and SI domains, respectively. Here, we adopted the question levels defined in [5]:

Level 1 Questions that can be answered with Yes/No or a few words.
Level 2 Questions that can be answered with a simple explanation at the general theory level.
Level 3 Those that require a reasoned and complex explanation.
Level 4 Those that require a reasoned explanation of the causal relationships between multiple concepts.

Levels of the raised questions were determined by a different set of five evaluators, one of which is an author of this paper. They give the levels independently (Cronbach's alpha = 0.99), and we computed the level by majority voting of the results.

For both COOK and SI, the number of Level 1 questions decreased and that of Level 2 questions increased significantly in the experimental group. The -square test results showed a significant difference in the distribution of question levels in both domains (COOK: $p = 0.01468$, SI: $p = 0.00019$).

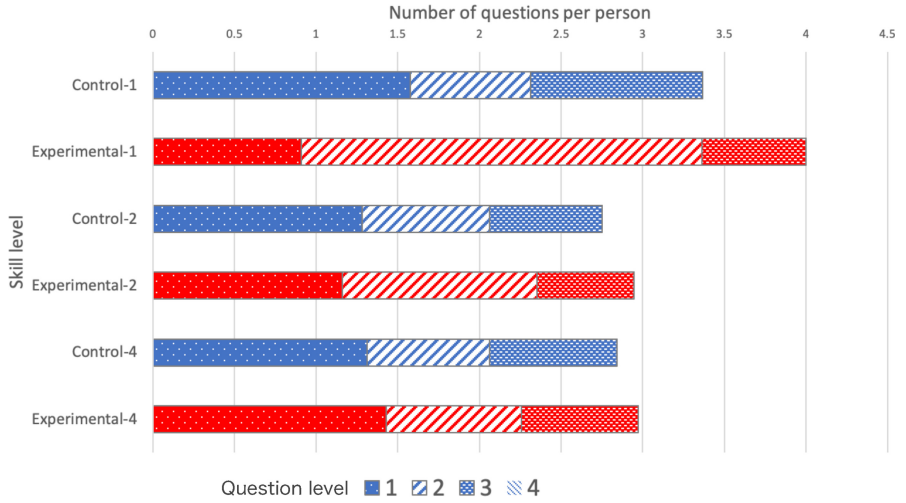

Fig. 6. Number of questions per person by skill level in COOK domain (As Table 5 shows, at some levels the numbers are too small to make comparisons, so these graphs show only skill levels with a population of 10 or more.). There is a decrease in the number of level 1 questions at skill level 1 and an increase in the number of level 2 questions at skill levels 1 and 2.

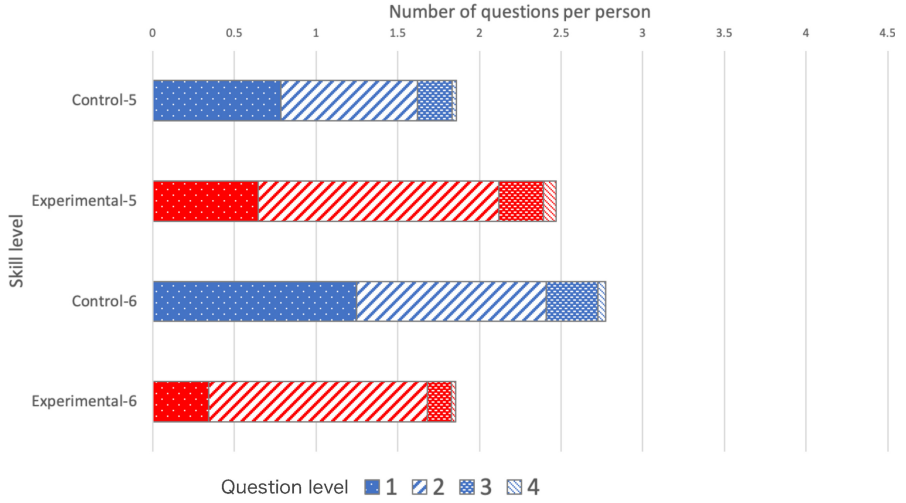

Fig. 7. Number of questions per person by skill level in SI domain (As Table 5 shows, at some levels the numbers are too small to make comparisons, so these graphs show only skill levels with a population of 10 or more.). There is a decrease in the number of level 1 questions and an increase in the number of level 2 questions at skill levels 5 and 6.

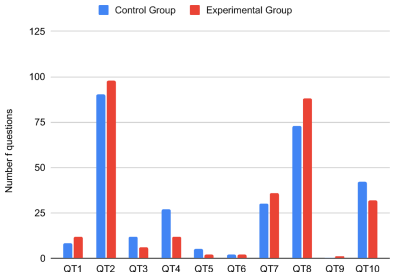

Fig. 8. Type of questions in COOK domain. In the experimental group, there are some noticeable different such as a decrease in the number of QT4 questions with a mean question level of around 1 (in Table 6), and an increase in the number of QT8 questions with a mean question level of around 3.

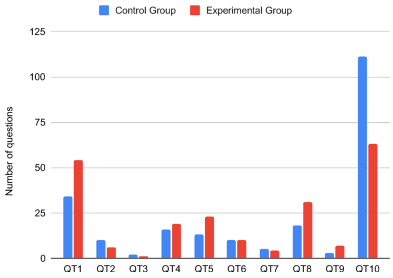

Fig. 9. Type of questions in SI domain. In the experimental group, there are some noticeable different such as an increase in the number of QT1 and QT8 questions with a mean question level of around 2 (in Table 6).

Figures 6 and 7 show the results of the analysis for each skill level. The horizontal axis is the number of questions at each question level per person on average. There is a significant increase in level 2 questions at skill levels 1 and 2 in COOK and at skill levels 5 and 6 in SI within the skill levels where there are enough people to make comparisons.

4 Discussion

As the result shows, presenting a list of question types did not affect the number of questions at all, although we did not force participants to raise questions. Note that the experiment setting was different from our framework's actual deployment, in that we asked crowd workers to raise questions. However, we did not find any cause that enforces the number of questions to be almost the same. Therefore, the result gave a negative answer to our first research question. Identifying factors to raise the number of questions is an interesting future work.

On the other hand, the result clearly showed that presenting a list of question types increased the number of questions at higher levels out of those raised by participants, and we can conclude that our second hypothesis, whether it affects the quality of questions, is true.

We assumed that presenting a list of question types reminded participants of questions they had not noticed so that they consider a more variety of questions. In order to see what happened in this aspect, we analyzed the types of questions raised by the participants.

We re-classified the collected questions because we found that some of the questions were labeled with wrong question types by participants. The five evaluators classified the questions raised by participants in both groups of both

domains (Cronbach's alpha = 0.98). If we compare this "gold standard" classification with the participants' classification results in controlled groups of COOK, the accuracy of the latter in COOK is 0.57, the recall is 0.51, and the precision is 0.57. Those values in SI are 0.49, 0.41, 0.40, respectively. Although the values are not high, as we show next, our result suggests that presenting types of questions is effective in increasing the number of questions in types whose average question level is higher.

Figures 8 and 9 compares the distributions of types of questions in the experimental and control groups. The types of questions that increased their numbers in the experimental groups of both domains are QT1, QT8, QT9. Those that decreased their numbers are QT3, QT10. Although some showed different behaviors (QT2, QT4, QT5, QT6, QT7), the result shows that presenting question types may affect the distribution of types of questions raised by participants. Table 6 shows the relationship between question types and question levels. The result shows that presenting question types increased the average of question levels because it increased the number of question types that result in higher question levels.

It has not been verified yet whether the types used in this study are effective for this system, or which types are more effective. These issues will be included in the future work.

In this experiment, the distribution of skills was skewed, especially in SI. In order to improve this, it is necessary to obtain the distribution of skills beforehand and measure the ability to understand the course content appropriately.

Table 6. Average level of questions in each types (In Table 6, the part that cannot be calculated because there is no question is left blank.). This table shows that there are several types of questions that tend to be at a higher level.

	COOK		SI	
	Control	Experimental	Control	Experimental
QT1	2.00	1.92	1.85	1.96
QT2	1.52	1.49	1.60	1.67
QT3	1.00	1.33	1.50	1.00
QT4	1.00	1.23	1.25	1.32
QT5	1.80	2.00	1.92	1.91
QT6	2.00	2.50	2.20	1.90
QT7	1.27	1.11	2.00	1.50
QT8	2.81	2.61	2.33	2.39
QT9	–	3.00	3.00	3.00
QT10	1.29	1.53	1.56	1.76

5 Conclusion

We examined the effect of explicitly presenting types of questions to students in an online learning environment. In our setting, people were shown online videos for lectures and then asked to raise questions. We compared the collected questions with those collected in the same situation but without presenting types of questions. We found that presenting types of questions improves the quality of questions, but did not increase the number of questions raised. Future work includes a more detailed exploration of the relationship between posed question types and collected questions, investigation of factors to increase the number of raised questions, and in-the-wild studies with deployed systems.

Acknowledgments. This work was partially supported by JST CREST Grant Number JPMJCR16E3, JSPS KAKENHI Grant Number JP19K11978, and JST-Mirai Program Grant Number JPMJMI19G8, Japan.

References

1. Chen, C.H., Bradshaw, A.C.: The effect of web-based question prompts on scaffolding knowledge integration and ill-structured problem solving. J. Res. Technol. Educ. **39**(4), 359–375 (2007)
2. Choi, I., Land, S.M., Turgeon, A.J.: Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. Instr. Sci. **33**(5–6), 483–511 (2005)
3. Endo, K.: A pilot study on the facilitating effect of the questioning behavior of students through the internet. J. Educ. Appl. Inf. Commun. Technol. **14**(1), 11–15 (2011)
4. Ge, X., Chen, C.H., Davis, K.A.: Scaffolding novice instructional designers' problem-solving processes using question prompts in a web-based learning environment. J. Educ. Comput. Res. **33**(2), 219–248 (2005)
5. Guthrie, J.T., et al.: Increasing reading comprehension and engagement through concept-oriented reading instruction. J. Educ. Psychol. **96**(3), 403 (2004)
6. Hirai, Y., Hazeyama, A.: A learning support system based on question-posing and its evaluation. In: Fifth International Conference on Creating, Connecting and Collaborating through Computing (C5 2007), pp. 178–184. IEEE (2007)
7. Hirai, Y., Hazeyama, A., Inoue, T.: Assessment of learning in concerto III: a collaborative learning support system based on question-posing. In: Proceedings of Computers and Advanced Technology in Education, pp. 36–43. ACTA Press, Calgary (2009)
8. Izumi, H., Matsubara, M., Watanabe, C., Morishima, A.: A microtask approach to identifying incomprehension for facilitating peer learning. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 4624–4627. IEEE (2019)
9. King, A., Rosenshine, B.: Effects of guided cooperative questioning on children's knowledge construction. J. Exp. Educ. **61**(2), 127–148 (1993)
10. Kruger, J., Dunning, D.: Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J. Pers. Soc. Psychol. **77**(6), 1121 (1999)

11. Lin, S.S.P., Samuel, M.: Scaffolding during peer response sessions. Proc. Soc. Behav. Sci. **90**, 737–744 (2013)
12. Luxton-Reilly, A., Bertinshaw, D., Denny, P., Plimmer, B., Sheehan, R.: The impact of question generation activities on performance. In: Proceedings of the 43rd ACM Technical Symposium on Computer Science Education, pp. 391–396 (2012)
13. Miyake, N., Norman, D.A.: To ask a question, one must know enough to know what is not known. J. Verbal Learn. Verbal Behav. **18**(3), 357–364 (1979)
14. White, R.T., Gunstone, R.F.: Metalearning and conceptual change. Int. J. Sci. Educ. **11**(5), 577–586 (1989)