

Exploring the interplay between question-answering systems and communication with instructors in facilitating learning

Cheng Yi, Runge Zhu and Qi Wang

School of Economics and Management, Tsinghua University, Beijing, China

Abstract

Purpose – Question-answering (QA) systems are being increasingly applied in learning contexts. However, the authors' understanding of the relationship between such tools and traditional QA channels remains limited. Focusing on question-answering learning activities, the current research investigates the effect of QA systems on students' learning processes and outcomes, as well as the interplay between two QA channels, that is, QA systems and communication with instructors.

Design/methodology/approach – The authors designed and implemented a QA system for two university courses, and collected data from questionnaires and system logs that recorded the interaction between students and the system throughout a semester.

Findings – The results show that using a QA system alone does not improve students' learning processes or outcomes. However, the use of a QA system significantly improves the positive effect of instructor communication.

Originality/value – This study contributes to the literature on learning and education technology, and provides practical guidance on how to incorporate QA tools in learning.

Keywords Question-answering systems, Communication with instructors, Media richness, Learning process, Learning outcome

Paper type Research paper

1. Introduction

In recent years, question-answering (QA) systems have been incorporated into a wide range of learning activities to serve as tutors (Crockett *et al.*, 2017), pedagogical agents (Min *et al.*, 2019) or learning peers (Hayashi, 2014). Generally speaking, QA systems are information retrieval systems in which an answer is expected in response to a submitted query (Mervin, 2013). They are used to facilitate question-answering learning activities that help students clear up their doubts and improve their understanding of the subject they are learning, without time and location constraints (Hien *et al.*, 2018; Waltinger *et al.*, 2012). Specifically, QA systems can understand a query written by a user in a live chat window and then reply to it based on information from a knowledge base. Current QA systems are increasingly being designed to support natural language-based interaction (Hien *et al.*, 2018; Hussain and

© Cheng Yi, Runge Zhu and Qi Wang. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

The authors thank the EIC, the associate editor and the two anonymous reviewers for their valuable comments and suggestions.

Funding: Financial support was received from the National Natural Science Foundation of China [Projects 72022008 and 71490724], and Tsinghua University Initiative Scientific Research Program [Grant 20205080019].



Athula, 2018). In the learning domain, these QA systems have already been adopted by some educational institutions and online learning platforms such as Summit Learning [1], and are being increasingly accepted by students (Wu *et al.*, 2020). A recent study of educational QA systems built on the Facebook Messenger platform showed that around 30% of respondents were willing to use QA systems to obtain quick answers when instructors were not accessible (Schmulian and Coetzee, 2019), and the acceptance rate can be expected to reach 70% by 2022 (Grudin and Jacques, 2019).

The use of QA systems in learning has also attracted some attention from academia. Researchers have mainly focused on improving the performance of QA systems in understanding and answering students' natural language queries (Kokku *et al.*, 2018; Tang *et al.*, 2018). Some recent studies have also examined the use of such QA systems in various learning scenarios in terms of students' adoption and evaluation of the systems (Lee *et al.*, 2020). However, it is still not clear whether such QA systems are helpful or how they can help improve students' learning experience and outcomes (Winkler and Söllner, 2018). The first focus of this study is thus to investigate the effect of using QA systems on students' satisfaction with the learning process and their learning outcomes.

Moreover, although QA systems are becoming increasingly popular, students seldom use them as a primary tool for clarifying their doubts. Rather, QA systems are usually deployed as a supplementary or supportive channel for students' communication with instructors. For example, some massive open online course (MOOC) platforms provide QA tools to deliver learning materials and answer students' questions when instructors are not online (Aguirre *et al.*, 2018). Communication with instructors, while constrained by time and location, is still critical for clarifying students' understanding of a subject (Harper, 2018). Extant research, however, has mostly focused on students' interaction with QA systems *per se* but has ignored the interplay between this emerging technology tool and traditional communication channels (Gupta and Bostrom, 2009).

In fact, the interplay between educational tools and instructors has been an important aspect of prior e-learning research (Dermentzi and Papagiannidis, 2018; Gupta *et al.*, 2010). It has been suggested that with the use of educational tools, students' interaction with learning materials and instructors will change (Alavi and Leidner, 2001; Fryer *et al.*, 2019). For example, Afzal *et al.* (2019) designed a tutoring system to promote students' engagement and self-reflection in learning, which effectively facilitated one-on-one tutoring between instructors and students by allowing students to better express their areas of uncertainty and instructors to better understand each student's knowledge gaps. Bhattacharjee (2009) found that the use of a tutoring system alone did not improve students' learning performance but suggested that further research should explore how students could benefit from the cooperation between the system and instructors. A recent report from the McKinsey Global Institute predicts that with the increasing automation of instructors' tasks using educational tools, the nature of learning, especially the interaction between students and instructors, will also change (Zhang *et al.*, 2018). In the current context, using QA systems implies that students can direct some of their queries to systems rather than rely on instructor communication. However, it is still unclear how students' use of QA systems will change the dynamic of student-instructor communication and the role of instructors. Hence, another focus of this study is whether and how QA systems interact with instructor communication in affecting students' learning processes and outcomes (Gupta and Bostrom, 2009).

To this end, we draw upon Media Richness Theory (MRT), which states that users will use different communication channels for different information tasks in order to achieve a fit between the characteristics of the media and the type of task. This leads us to propose that the use of QA systems will positively influence learning experiences and outcomes, as it can help students resolve relatively basic, fact-based questions. Moreover, through the use of QA

systems, students' communication with instructors about more equivocal, personalized questions will become more effective because of increased instructor capacity and a higher quality of questions raised by students. We conducted a field study by implementing a QA system for two major courses in a university and collecting data from questionnaires and system logs throughout a semester. Our findings reveal that, on average, QA system use does not improve students' learning processes or outcomes. However, the use of a QA system does improve the effectiveness of student-instructor communication.

Overall, this study contributes to the literature by empirically examining the effect of QA systems on students' learning processes and outcomes and shedding light on the use of such systems in the learning context. Moreover, while the majority of prior studies focus on the effect of technology tools *per se*, this study explores whether and how technology tools (i.e. QA systems) interact with traditional learning channels (i.e. instructor communication) in influencing students' learning processes and outcomes through the theoretical lens of MRT.

2. Literature Review

2.1 Question-answering learning methods and QA channels

Question answering is one of the most important learning methods (Davey and McBride, 1986). It is widely recognized that the process of raising questions and seeking answers may help students deepen their understanding and improve their learning outcomes (Waltinger *et al.*, 2012). Specifically, composing questions requires students to focus on elaborating and internalizing received information, while obtaining answers helps them resolve problems and improve their comprehension. Pedagogical research suggests that student-generated questions can be classified as *fact-demanding* questions and *thought-provoking* questions according to the level of thought required to answer them (Barak and Rafaeli, 2004; Chin, 2002; Scardamalia and Bereiter, 1992). Fact-demanding questions reflect students' uncertainty about basic concepts and facts, which can usually be answered by clarifying definitions and giving examples. When students form a solid understanding of basic concepts by clarifying their fact-demanding questions, they will be capable of engaging in deep thinking and raising thought-provoking questions. The process of raising and resolving thought-provoking questions requires in-depth reasoning, information integration and critical thinking, which cultivate a high level of mastery of the subject (Chin, 2002; Chin and Brown, 2002). Studies have found that students who raise questions more frequently are more likely to achieve a deep understanding of learning materials and are also more capable of posing thought-provoking questions, which leads to a better learning outcome (Yu, 2009).

In traditional learning contexts, students often seek to communicate with instructors when they have questions that need answering. The critical role of students' communication with instructors has been demonstrated in many pedagogical studies (Pizzini and Shepardson, 1991). It has been suggested that a dialogic communication process between students and instructors can effectively clear up students' doubts and misconceptions, as well as stimulate their investigations into the subject domain, hence facilitating the construction of new knowledge (Chin, 2002; Chin and Brown, 2002). However, due to temporal and geographic constraints, an instructor may not be able to meet the QA needs of a large number of students. Various technology tools, such as QA systems, that allow students to generate questions and provide feedback, have therefore been developed and implemented in learning contexts.

QA systems have been around for some years. Most early QA systems only supported keyword-based queries (Barak and Rafaeli, 2004). While such QA systems have been shown to improve students' learning engagement by allowing them to post course-related queries, they often suffer from issues of low accuracy (Song *et al.*, 2017). Since users have to frame their questions in the form of keywords, they may not be able to express their questions

precisely, and it might also be difficult for the system to understand the users' intentions and provide them with an accurate response. Hence, the application and benefits of keyword-based QA systems in learning contexts are limited (Tang *et al.*, 2018).

In recent years, natural language-based QA systems such as chatbots and conversational agents have been increasingly used to support the learning process. These systems allow students to raise questions using natural language, and can "understand" their questions, retrieve vital information from a knowledge base, and offer precise answers (Song *et al.*, 2017). Machine-learning-based classification of natural language questions significantly improves the response rate and the accuracy of generated answers (Sharma and Gupta, 2018; Waltinger *et al.*, 2012). The development and application of QA systems [2] have attracted increasing research attention in the fields of data science, pedagogy, and information systems (see Appendix 1 for a summary of studies). Mainstream studies have focused on the technical aspects of QA systems, such as optimizing algorithms to improve QA system performance in "understanding" and answering students' questions (Kokku *et al.*, 2018; Tang *et al.*, 2018). Some other studies have implemented QA systems in learning scenarios, and have examined the response rates and users' satisfaction with the systems (Fryer *et al.*, 2019; Lee *et al.*, 2020). However, a recent review of studies on pedagogical QA systems shows that how such systems affect students' learning processes and outcomes is still largely unknown, and that more empirical evidence still needs to be collected (Winkler and Söllner, 2018).

2.2 Richness of different QA channels

Students can raise questions and seek answers from technology-mediated channels such as QA systems, or from traditional channels such as communication with instructors. We draw upon media richness theory to better understand the characteristics of the two different channels, and how users may use different channels to support QA learning activities.

According to MRT, communication media can be characterized by their accessibility, availability of instant feedback, level of personalization, and communicative cues (Daft and Lengel, 1986). Human interaction is often regarded as a rich medium, whereas lean media include various self-serving electronic media such as frequently asked questions (FAQ) on company websites and self-service terminals. Communication in rich media may involve verbal expressions, facial expressions and body language, whereas lean media usually provide fewer information cues and less customized feedback. Nonetheless, lean media can offer users higher accessibility and availability as they can be accessed at any time, and thus increase the efficiency of information acquisition (Kock, 2009; Seeber *et al.*, 2020). Therefore, past studies suggest that equivocal questions with multiple interpretations and solutions are better supported by rich media (Scherer *et al.*, 2015), whereas standard questions with clear answers can be supported by lean media (Kahai and Cooper, 2003). In general, the central proposition of MRT is that users will choose communication media by aligning the equivocality of information to be conveyed to the richness of the media in order to achieve better task efficiency and effectiveness.

The question of whether users tend to treat QA systems as rich or lean media has attracted recent research attention. Studies find that most users have their own expectations as to what QA systems can support, and tend to use systems for fact-based and unequivocal tasks. This is because, compared with human interaction, QA systems allow limited communication cues and low variety of language, so that users find it difficult to convey and clarify personalized and thought-provoking questions. Similarly, when interacting with QA systems, users have also been found to be less open and agreeable than they are with humans, probably due to the difficulty in forming a mutual understanding during communication on equivocal topics (Luo *et al.*, 2019). Based on these findings, it is likely that users tend to regard a QA system as a relatively lean medium compared with person-to-person communication.

Generally speaking, in the current context, a QA system can be characterized as a lean medium, and users are more likely to use it to resolve relatively simple, fact-demanding questions, given that mainly text-based cues are provided. Users can gain immediate feedback from QA systems without time and space constraints. In contrast, communication with instructors can be considered a rich channel for dealing with more personalized, thought-provoking questions using rich information cues. However, such a channel may have limited accessibility due to limited instructor resources (see Table 1 for a summary). According to MRT, effective learning should occur when learning channels are compatible with users' learning activities. In other words, when fact-demanding questions are resolved efficiently via QA systems, limited instructor resources can be reserved for answering thought-provoking questions, and hence utilized more effectively.

3. Hypotheses development

This study focuses on the effect of students' use of QA systems on their satisfaction with the learning process and their learning outcomes, as well as the interplay between the use of QA systems and communication with instructors. In particular, we define students' satisfaction with the learning process as the extent to which students are content with their learning experience (Xu *et al.*, 2014). We define a learning outcome as the level of knowledge or skills students acquire after learning a course (Söllner *et al.*, 2018). Learning processes and outcomes have been the two focal evaluation dimensions of many prior studies in e-learning (e.g. Gupta and Bostrom, 2009; Gupta *et al.*, 2010). In this section, we first propose the effects of the traditional QA channel (i.e. communication with instructors) on learning processes and outcomes based on pedagogical literature, and then the effects of QA systems, as well as their interplay.

3.1 The effect of Communication with instructors

It has been well documented in pedagogical literature that QA is an important learning method (Pizzini and Shepardson, 1991). In particular, communicating with instructors is a primary and effective way of answering students' questions (Chin, 2002). As mentioned earlier, during a learning process, students may raise different kinds of questions, ranging from standard, repetitive questions about basic concepts (i.e. fact-demanding questions) to more equivocal and personalized questions (i.e. thought-provoking questions) (e.g. Chin, 2002; Barak and Rafaeli, 2004). According to MRT, since students and instructors can use rich cues to convey information to each other, such communication can best enable students to articulate their personalized questions and confusions, and also allow instructors to provide rich, clear explanations to answer them. This will greatly reduce students' frustration during

Table 1.
Comparison of
different QA channels

Channels Characteristics	QA systems	Communication with instructors
Accessibility	Students access QA systems through a mobile phone or computer without temporal or space constraints	Students need to schedule a QA session with instructors in a particular location
Feedback	Immediate feedback once a question is raised	Immediate feedback only in a scheduled QA session
Personalization	Preset and standard answers from database	Diverse and personalized answers from instructors
Communicative cues	Answers are delivered through text/picture without social cues	Answers are delivered by instructors with social and informational cues

the learning process (Pizzini and Shepardson, 1991). In addition, more communication with instructors also fosters a closer personal connection between students and instructors, which has been shown to increase students' attachment to the course and their satisfaction with the learning process (Rosenshine *et al.*, 1996). Therefore, we propose,

H1a. Communication with instructors has a positive effect on students' satisfaction with the learning process.

By resolving students' uncertainties about the course content, communication with instructors can help students form a solid understanding of the course content and become well-prepared to learn the upcoming materials (Chin, 2002). Moreover, being able to directly communicate with instructors encourages students to raise more thought-provoking questions and stimulates their critical thinking (Chin and Brown, 2002). This will likely facilitate their knowledge construction process and lead to an improved learning outcome. Therefore, we propose,

H1b. Communication with instructors has a positive effect on students' learning outcomes.

3.2 The effects of using QA systems

Fact-demanding questions are common in students' learning processes and can usually be answered with clear definitions and examples (Chin and Brown, 2002). According to MRT, users are likely to use QA systems to resolve their fact-demanding questions as these systems have been developed to offer precise answers to relatively standard questions based on a knowledge database (Scherer *et al.*, 2015). The ability to raise questions and obtain answers from a QA system implies that students can immediately resolve their uncertainties about basic concepts without temporal and geographic constraints. This will make the learning process more efficient and thus increase users' satisfaction with the process. Students will also have more control over their learning pace than when QA systems are not used. Prior studies have suggested that a high level of control over the learning process has a positive effect on student satisfaction (Xu *et al.*, 2014; Zimmerman and Schunk, 2001). Therefore, we propose,

H2a. Using QA systems has a positive effect on students' satisfaction with the learning process.

Using QA systems may also be associated with improved learning outcomes, for several reasons. First, as students can ask QA systems whenever they have uncertainties about basic concepts and have these uncertainties cleared up quickly, they are likely to form a solid knowledge base. Indeed, the process of raising questions in such a channel also requires students to reflect on their learning materials and pay attention to their confusions and misconceptions (Yu, 2009). Second, when questions about basic concepts can be resolved promptly, students are more likely to explore further and think more deeply. The positive relationship between deep thinking and learning outcomes is well documented (Chin, 2002). Hence, we propose,

H2b. Using QA systems has a positive effect on students' learning outcomes.

3.3 The interplay between using QA systems and Communication with instructors

As MRT suggests, users will choose their communication channels based on the richness of the media and the information to be conveyed. When a fit between communication channels and tasks is achieved, efficiency and task performance will be improved (Scherer *et al.*, 2015). In the current context, communication with instructors is a rich, personal communication

channel that may help resolve personalized and equivocal questions but is constrained by instructor resources (Lee *et al.*, 2020). QA systems are relatively lean media for basic, repetitive questions but are highly accessible. We therefore argue that students' use of QA systems to resolve basic questions will help alleviate instructors' resource constraints and achieve a fit between communication channels and QA activities, hence reinforcing the benefits of communication with instructors.

Specifically, when some students' questions are resolved through QA systems, instructors can spend less time on repetitive fact-demanding questions and hence devote more time to resolving personalized and thought-provoking questions. This will help make the best use of such a channel, given that it is a rich medium appropriate for resolving such questions. In other words, students are more likely to have access to instructors when they want to discuss their personalized questions in depth, leading to more efficient and effective interactions between students and instructors. Hence, the effect of communication with instructors on students' learning processes and outcomes is improved.

In addition, studies have suggested that question-generating skills can be developed through repeated practice (Yu, 2009). By raising and resolving basic questions via QA systems, students will build a solid knowledge base and be more capable of raising insightful and thought-provoking questions. They will thus be more confident in raising high-quality questions to instructors and understanding instructors' explanations, leading to more fluent and fruitful communication. Overall, using QA systems enables students to benefit more from interacting with instructors, due to increased instructor capacity and the improved quality of questions raised by students. Hence, we propose,

- H3a. Using QA systems enhances the effect of communication with instructors on improving students' satisfaction with the learning process.
- H3b. Using QA systems enhances the effect of communication with instructors on improving students' learning outcomes.

4. Research method

We conducted a field study in a major public university in China. A QA system was built for two major courses offered by the university, in which a total of over 300 students were enrolled.

4.1 System design

Two core modules in the School of Economics and Management, *Principles of Economics* and *Principles of Accounting*, were selected as the target courses for which QA systems were built. We chose these courses for two reasons. First, both courses had a class size of over 100 students, which was considered a large number for one instructor to handle. This class size provided us with a reasonable context in which to build and study the effect of QA systems. Second, students generally tended to be active in asking questions and seeking to clarify their uncertainties because their performance in these core modules was very important.

We collaborated with a well-reputed third-party developer to develop the QA system. The system consisted of three key elements. First, we worked with the course instructors and teaching assistants to establish a knowledge base for each course. Information in the knowledge base was mainly sourced from textbooks, lecture slides, relevant cases and essays, answers to textbook exercises, and records of students' questions and answers in past semesters. This information was processed and restructured to form question-answer pairs.

We also included QA pairs related to administrative matters such as lecture time and assignment deadline.

Second, we used text-similarity matching methods to match questions raised by students with the questions in our knowledge base. In addition, to improve the system's understanding of users' natural language queries, we tried to identify users' intentions by building an intent classifier for the queries. In particular, we defined four common intents based on all the questions in the knowledge base: *definition*, *example*, *case study* and *news*. Each query was then classified into one of the four intent categories.

Third, since the courses were taught bilingually, a translator engine was used to allow both Chinese and English questions. The system also supported simple social interactions such as greetings by calling up an existing chatbot application programming interface.

After designing and implementing the system, we trained the system for two months by inviting a group of students who had taken these courses in previous semesters to raise questions and rate the answers provided. In particular, when a student raised a question, the system would first confirm with the student the meaning of the question by providing several question options. After receiving the student's feedback, the system would return the corresponding answers. Next, a student was asked to label each answer as relevant or irrelevant, and the system would automatically update the accuracy of possible answers based on users' interactions and feedback. By this means, the QA ability of the system continuously increased as more questions were raised and more answers were rated or updated.

The QA system was deployed on the WeChat platform, which is the most widely used mobile messaging application in China. We created a WeChat public account for each course, and every student could exchange messages with the account after subscribing to it, as though they were chatting with a real user. At the beginning of the semester, all students were asked to subscribe to the public account of the course for free. They could then start a conversation with the account at any time. For each question, the system would provide a best-matched answer on top, followed by four related readings. Users could click the answer to check its detailed content. Figure 1 shows the system interface with examples of questions and answers.

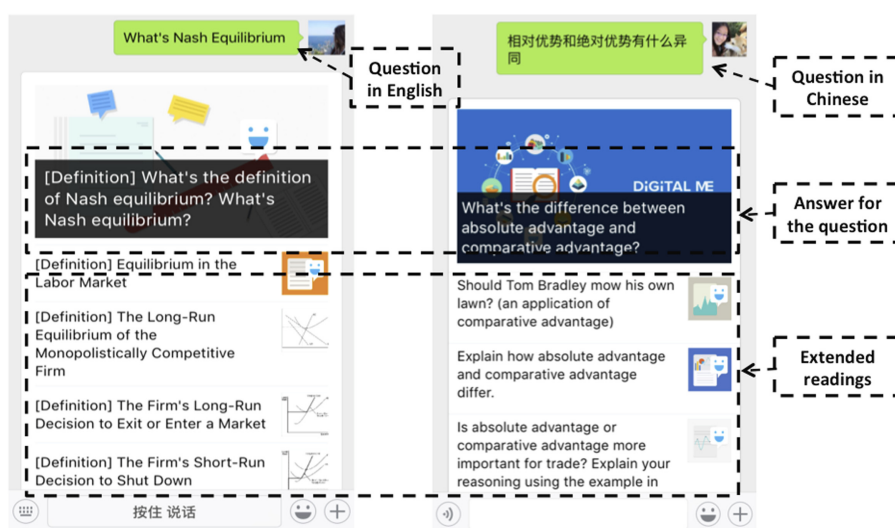


Figure 1.
QA system interface

When a question could not be answered by the system (i.e. no matching answer was found in the knowledge base), the system would apologize and inform the student that it would try to provide an answer within 48 hours. The research assistants, who were constantly monitoring the system at the back end, would contact the instructor for an answer quickly and reply via the system.

4.2 Data collection

The system was used in the two courses throughout the fall semester in 2017. All students were undergraduates (mostly freshmen and sophomores). They all used smartphones and were generally familiar with WeChat. Data were collected in three phases and from different sources, as elaborated below.

Phase 1: Pre-questionnaire. A pre-questionnaire was issued at the start of the semester to collect students' demographic information and learning-related traits (see the next section for elaboration and [Appendix 2](#) for measurement items). A total of 232 responses were collected at this phase.

Phase 2: System log. During the semester, the system log recorded all the dialogues between the students and the system, including question content, course, student ID, time to raise the question, and the system's reply. The main independent variable, that is, the use of the QA systems, was measured based on this system log. During the semester, 1,263 interaction records were collected at this phase, involving five types of interaction messages (see [Table 2](#) for details).

Phase 3: Post-questionnaire. A post-questionnaire was issued before the final exam to collect responses to measures of the dependent variables (i.e. satisfaction with the learning process and learning outcome), the independent variable (i.e. communication with instructors), as well as other variables related to the learning process (see the next section for elaboration and [Appendix 2](#) for measurement items). A total of 176 responses were collected at this phase.

After matching responses were collected from the three phases, we had 152 students with complete data from all phases, and our data analysis was based on this sample. Of these students, 86 (56.6%) were female, and 66 (43.4 %) were male.

5. Model and analysis

5.1 Empirical model

This study aims to investigate the impact of using the QA system on students' learning processes and outcomes. Our independent variable, *use of QA systems (U_Systems)*, was a

Table 2.
Descriptive statistics of
system log

Message type	Explanation	Number	Percentage
System answered questions	Queries answered by the system	990	78.4
Manually answered questions	Queries which were not resolved by the system but answered by research assistants through the system within 48 hours	37	2.9
Unanswered questions	Queries which were not resolved by the system and were not answered by research assistants within 48 hours	111	8.8
Administrative questions	Queries about course schedule, location, tests and other administrative information	70	5.5
Chatting messages	Messages such as "hello," "thank you," "who are you"	55	4.4
Total		1,263	100.0

binary indicator: it was coded “1” if a student used the system to raise at least one question related to learning content, and zero otherwise. The other independent variable, *Communication with instructors* ($C_Instructors$), was obtained from the self-reported frequency of interacting with instructors in the post-study questionnaire. The two dependent variables, *satisfaction with the learning process* ($L_Process$) and *the learning outcome* ($L_Outcome$), were both obtained from the post-study questionnaire, based on previously-validated measures (Xu *et al.*, 2014). Exploratory factor analysis was performed on the learning process and learning outcome. Results showed that measurement items were loaded heavily on their intended factor and lightly on the other factors, indicating adequate convergent and discriminant validity [3] (see Table 3 for factor loadings). The Cronbach’s alpha of the measure of the learning process and learning outcome were 0.93 and 0.91, respectively, indicating adequate reliability of the measurement scales.

The control variables included the use of other potential QA channels, such as searching online materials ($F_Materials$, *frequency of using online learning materials*), discussing with peer students (F_Peers , *frequency of discussing with peers*), and reviewing textbooks (F_Books , *frequency of reviewing textbooks*). These learning behaviors might also have direct impacts on the learning process and performance (Rosenshine *et al.*, 1996). We measured the use of these alternative methods in the post-study questionnaire. We also included control variables related to one’s past experience and potential acceptance of new technology applications, such as one’s *experience of using online learning tools* (E_Tools) (Arbaugh and Duray, 2002), *personal innovativeness in information technology* (I_IT) (Agarwal and Prasad, 1998), and *playfulness* (PL) which represented one’s willingness to play with and explore a new system (Dewey, 1913). Further, since students might differ in terms of their capability for learning their chosen subjects, we also measured students’ *academic self-efficacy* ($A_Efficacy$), which represented their belief that they could successfully accomplish the target course (Bandura *et al.*, 1987). These variables were measured in the pre-study questionnaire, and all the measurement items were adapted from prior studies (see Appendix 2).

To address our research questions, we first modeled the effect of instructor communication and QA system use on users’ satisfaction with the learning process, as well as their interplay, using the following equation:

$$\begin{aligned} L_Process = & \beta_1 U_Systems + \beta_2 C_Instructors + \beta_3 U_Systems \times C_Instructors \\ & + \beta_4 F_Materials + \beta_5 F_Peers + \beta_6 F_Books + \beta_7 I_IT + \beta_8 PL \\ & + \beta_9 A_Efficacy + \beta_{10} E_Tools \end{aligned}$$

We modeled the effect of instructor communication and QA system use on learning outcome as well as their interplay as follows:

Items	Factors	
	Satisfaction on learning process	Learning outcome
Process 1	0.923	0.432
Process 2	0.918	0.231
Outcome 1	0.332	0.901
Outcome 2	0.301	0.849

Note(s): Extraction method: principal component analysis; rotation method: varimax with Kaiser normalization

Table 3.
Rotated factor loadings

$$\begin{aligned} L_Outcome = & \beta_1 U_Systems + \beta_2 C_Instructors + \beta_3 U_Systems \times C_Instructors \\ & + \beta_4 F_Materials + \beta_5 F_Peers + \beta_6 F_Books + \beta_7 I_IT + \beta_8 PL \\ & + \beta_9 A_Efficacy + \beta_{10} E_Tools \end{aligned}$$

Table 4 presents the descriptive statistics and the correlation matrix of all variables based on 152 observations.

5.2 Results

We first estimated the ordinary least square (OLS) model on all control variables. As reported in columns (1) and (4) in Table 5, academic self-efficacy positively affected the learning process and learning outcome, while the frequency of reviewing textbooks positively affected the learning outcome. We then estimated the effect of using the QA system and communicating with instructors by further including *U_Systems* and *C_Instructors*. The results are presented in columns (2) and (5). As indicated, communicating with instructors positively affected the learning process ($\beta = 0.209, p < 0.001$) and learning outcome ($\beta = 0.167, p < 0.05$), providing support to H1a and H1b. However, using the QA system alone did not have a significant effect on the learning process ($\beta = 0.140, p > 0.1$) or learning outcome ($\beta = -0.158, p > 0.1$). Hence, H2a and H2b were not supported.

To test the interplay between using the QA system and communicating with instructors, we further included the interaction term *U_Systems* \times *C_Instructors* in the full model. As column (3) and column (6) showed, there was an interaction effect between using the QA system and communicating with instructors on both the learning process ($\beta = 0.051, p < 0.001$) and the learning outcome ($\beta = 0.042, p < 0.001$). To further clarify the nature of this interaction effect, we performed separate analyses for users and non-users of the QA system and tested the simple main effects of communication with instructors. Specifically, for students who used the QA system, more instructor communication led to higher satisfaction with the learning process ($\beta = 0.789, p < 0.001$) and a better learning outcome ($\beta = 0.746, p < 0.001$). For students who did not use the QA system, the positive effect of communication with instructors on the learning process still held, but it was less prominent ($\beta = 0.189, p < 0.05$, see Figure 2a), and there was no significant effect of instructor communication on the learning outcome ($\beta = 0.114, p > 0.1$, see Figure 2b). Thus, H3a and H3b were supported.

5.3 Identification

The above result shows positive interactions between QA systems and instructor communication in terms of improving students' learning processes and outcomes. However, there might be potential endogeneity issues if students who did not use the QA system (i.e., *U_Systems* = 0) differed from those who used the system (i.e., *U_Systems* > 0) in other observable dimensions. In particular, we identified three such dimensions in which the two student groups differed, including (1) gender, (2) general willingness to ask questions (*W_Questions*), and (3) computer self-efficacy (*C_Efficacy*). Specifically, these individual characteristics might potentially affect students' use of the QA system. For example, prior research has shown that males and users with higher computer self-efficacy are more willing to try new technologies (Dong and Zhang, 2011). Also, if a student is generally more willing to ask questions, he/she is more likely to use the QA system to address his/her questions when it is available. In other words, the previous results on the moderating effect of the QA system may not be attributed to system use but to these individual characteristics.

We thus used the three factors as matching variables and adopted the Propensity Score Matching (PSM) technique to construct a control group (i.e., *U_Systems* = 0) that was similar

Variable	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) <i>L_Process</i>	4.092	1.528	–										
(2) <i>L_Outcome</i>	4.278	1.534	0.712	–									
(3) <i>U_Systems</i>	0.276	0.449	0.088	0.008	–								
(4) <i>C_Instructors</i>	1.585	2.033	0.353	0.312	0.058	–							
(5) <i>F_Materials</i>	4.658	1.504	0.052	–0.021	–0.021	0.249	–						
(6) <i>F_Peers</i>	4.820	1.524	0.122	–0.179	–0.175	0.335	0.121	–					
(7) <i>F_Books</i>	5.010	1.632	0.199	0.331	–0.111	0.344	0.654	0.314	–				
(8) <i>I_IT</i>	4.500	1.208	0.116	0.072	0.031	0.084	0.029	0.102	0.066	–			
(9) <i>PL</i>	5.468	1.129	0.119	0.113	–0.045	–0.030	0.094	0.026	0.124	0.623	–		
(10) <i>A_Efficacy</i>	4.802	0.968	0.332	0.107	0.061	0.171	–0.102	0.055	0.054	0.234	0.300	–	
(11) <i>E_Tools</i>	5.001	1.323	0.197	0.263	0.074	0.149	0.162	0.069	0.114	0.324	0.259	0.400	–

Table 4.
Descriptive statistics
and correlation matrix

Table 5.
OLS results on learning
process and learning
outcome

Variable	Dependent variable: Learning process			Dependent variable: Learning outcome		
	Control (1)	Main effect (2)	Full model (3)	Control (4)	Main effect (5)	Full model (6)
<i>U_Systems</i>		0.140 (0.268)	0.191 (0.257)		-0.158 (0.286)	0.213 (0.280)
<i>C_Instructors</i>		0.209*** (0.066)	0.224*** (0.063)		0.167* (0.094)	0.120** (0.067)
<i>U_Systems</i> × <i>C_Instructors</i>			0.051*** (0.014)			0.042*** (0.015)
<i>F_Materials</i>	-0.063 (0.105)	-0.103 (0.102)	-0.124 (0.101)	0.039 (0.106)	-0.004 (0.111)	-0.015 (0.108)
<i>F_Peers</i>	-0.023 (0.066)	-0.032 (0.068)	-0.029 (0.067)	0.102 (0.066)	0.056 (0.069)	0.066 (0.071)
<i>F_Books</i>	0.181 (0.095)	0.104 (0.090)	0.151 (0.090)	0.226* (0.097)	0.205* (0.098)	0.212* (0.096)
<i>I_IT</i>	-0.044 (0.127)	-0.102 (0.123)	-0.099 (0.122)	-0.115 (0.127)	-0.154 (0.127)	-0.238 (0.124)
<i>PL</i>	0.043 (0.135)	0.145 (0.134)	0.346* (0.138)	0.097 (0.136)	0.167 (0.138)	0.349* (0.141)
<i>A_Efficacy</i>	0.450** (0.140)	0.138** (0.010)	0.320* (0.133)	0.320* (0.098)	0.212 (0.140)	0.163 (0.135)
<i>E_Tools</i>	0.101 (0.101)	0.091 (0.097)	0.026 (0.096)	0.194 (0.101)	0.188 (0.099)	0.154 (0.096)
Course dummy	-0.809** (0.270)	-0.890** (0.260)	-0.960*** (0.250)	-0.676* (0.271)	-0.716** (0.275)	-0.606* (0.103)
Constant	1.782* (0.893)	1.276 (0.850)	1.069 (0.817)	0.764 (0.896)	1.248 (0.905)	1.289 (0.932)
Observations	152	152	152	152	152	152
R^2	0.151	0.217	0.285	0.203	0.221	0.262
Adjusted R^2	0.110	0.167	0.234	0.158	0.166	0.204

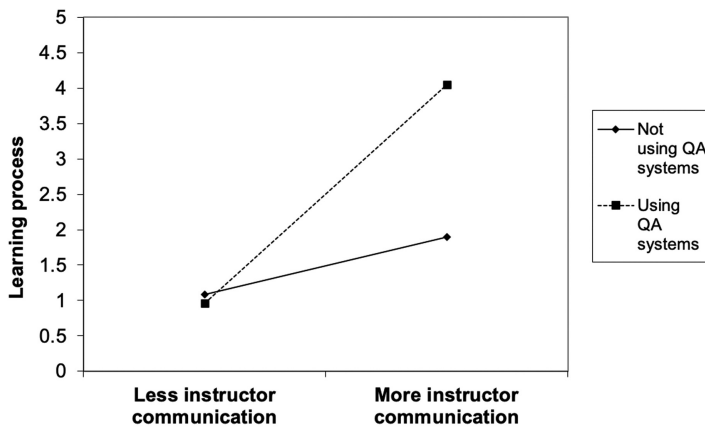
Note(s): * $p < 0.05$; ** $0 < 0.01$; *** $p < 0.001$; standard errors in parentheses

to the treatment group who used the system (i.e., $U_Systems > 0$) in terms of these variables. Specifically, we performed PSM with the one-to-one nearest-neighbor matching (without replacement) algorithm (see [Appendix 3](#) for PSM results). After matching, the two groups had no significant differences across all variables. This implies that the PSM was successful, and that the difference between the two groups came mainly from using the QA system. As the PSM-constructed sample was considered small, we performed the bootstrapping method to estimate the regression. The results are summarized in [Table 6](#), column (1) and column (8) [\[4\]](#), and show findings consistent with the OLS analysis.

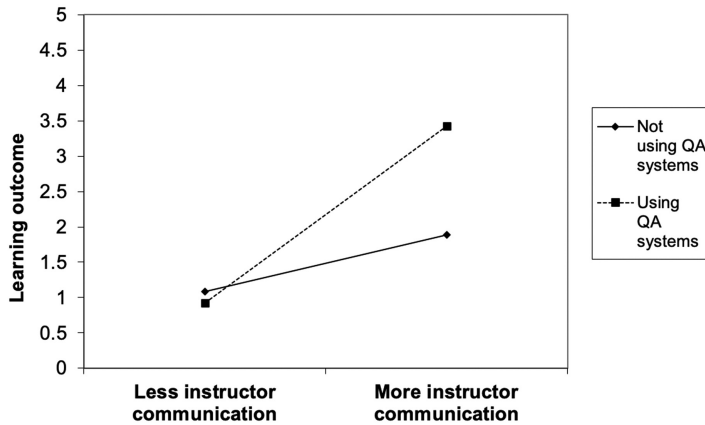
5.4 Robustness check

We further corroborated our findings by checking the robustness and consistency of the findings in multiple ways.

First, in the original OLS model, $U_Systems$ was modeled as a binary variable indicating whether students used the system to raise at least one question. To account for possible information loss in the binary measure, we adjusted and re-estimated the model by using the frequency of QA system use ($F_Systems$, mean = 7.48, SD = 14.58), measured by the total number of questions raised by each student. The estimates of coefficients were reported in column (2) and column (9) of [Table 6](#). In addition, since this variable was obtained from system logs and its standard deviation was considered large, we also log-transformed



(a)



(b)

Figure 2.
Plots of interaction
effect on learning
process and learning
outcome

$F_Systems$ to estimate a log-level model. The results are presented in column (3) and column (10) of Table 6. Overall, the results of these models were similar to the OLS model.

Second, to address the possible concern on the multicollinearity issue, we re-estimated the model based on the centralized variables and summarized the results in columns (4) and (11) of Table 6. The variance inflation factor (VIF) values of the variables were all below 2. The estimates of coefficients of the interaction term were similar to those in the OLS model.

Third, to establish the robustness of our results across different PSM algorithms, we constructed the control group based on three other matching algorithms, that is, the one-to-one, one-to-two, and one-to-three nearest-neighbor matching (with replacement) algorithm (see Appendix 3 for the PSM results). We estimated the full model on the learning process and learning outcome, respectively. The results, based on three different matching algorithms, are presented in columns (5) to (7) and columns (12) to (14) of Table 6. As indicated, all the coefficients of the interaction term remained positive and significant.

Table 6.
Regression results for
identification and
robustness checks (1)

Variable	Dependent variable: Learning process				Dependent variable: Learning outcome									
	I.1 without replace (1)	Frequency (2)	Log- frequency (3)	Centralization (4)	I.1 replacement (5)	I.2 replacement (6)	I.3 replacement (7)	I.1 without replace (8)	Frequency (9)	Log- frequency (10)	Centralization (11)	I.1 replacement (12)	I.2 replacement (13)	I.3 replacement (14)
<i>U. Systems/</i>	0.483	0.036	0.016	0.071	0.007	0.003	0.013	0.462	0.003	-0.003	0.065	0.002	0.001	0.014
<i>F. Systems</i>	(0.285)	(0.030)	(0.031)	(0.039)	(0.039)	(0.038)	(0.038)	(0.328)	(0.034)	(0.034)	(0.157)	(0.043)	(0.043)	(0.042)
<i>C. Instructors</i>	0.145*	0.225***	0.241***	0.241***	0.184*	0.119	0.142	0.134*	0.139*	0.139*	0.126*	0.084	0.048	0.073
	(0.087)	(0.060)	(0.060)	(0.060)	(0.085)	(0.087)	(0.082)	(0.074)	(0.058)	(0.055)	(0.070)	(0.116)	(0.099)	(0.097)
<i>U. Systems/</i>	0.414*	0.035***	0.2441***	0.243***	0.051***	0.032***	0.040**	0.392***	0.041**	0.257***	0.254***	0.045*	0.046**	0.043*
<i>F. Systems ×</i>	(0.191)	(0.013)	(0.066)	(0.065)	(0.019)	(0.018)	(0.018)	(0.186)	(0.015)	(0.072)	(0.071)	(0.002)	(0.020)	(0.021)
<i>C. Instructors</i>														
<i>Control</i>	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
<i>variables</i>														
<i>Observations</i>	84	152	152	152	80	91	100	84	152	152	152	80	91	100
<i>R-squared</i>	0.265	0.358	0.340	0.340	0.426	0.427	0.276	0.342	0.259	0.286	0.288	0.432	0.443	0.378
<i>Adjusted</i>	0.245	0.308	0.288	0.288	0.336	0.347	0.298	0.217	0.200	0.230	0.232	0.343	0.366	0.301
<i>R-squared</i>														

Note(s): * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; standard errors in parentheses

Fourth, to understand whether the effect of QA system use and instructor communication can hold across different demographic subgroups, we broke down the overall subject population into different subgroups according to gender, students' willingness to ask questions, and computer self-efficacy, which might affect their system use behavior, and tested the effects of system use respectively. Specifically, in terms of gender, the sample was directly divided into male and female subsets. In terms of willingness to ask questions and computer self-efficacy, we performed a median split and divided subjects into high and low levels. As the reconstructed subsets were considered small, we performed the bootstrapping method to estimate the regressions for each subset. The results are presented in Table 7. As indicated, the coefficients of the interaction effect between QA system use and instructor communication remained positive and significant in all the subgroups.

In sum, the above checks demonstrated the robustness and consistency of our findings regarding the interplay between the use of QA systems and instructor communication.

6. Conclusion

6.1 Discussion of results

This study focuses on the individual and interaction effects of two important learning channels, QA systems and communication with instructors, on learning processes and outcomes. Our findings show that communication with instructors positively influences students' learning processes and outcomes. However, using a QA system alone does not have a significant impact on students' learning processes or outcomes. To further understand how students interacted with the system, we carefully examined the system log. We found that among all the course-related questions raised in the system, 92% could be categorized as fact-demanding questions, and the system response rate for these questions reached 99%. However, for the rest of the questions, which could be categorized as thought-provoking questions, the response rate was only 65%. Although we took some remedial approaches for questions that the system failed to answer, such delay or inability to respond would affect students' learning experience and their willingness to use the system. In general, the application of QA systems in the learning domain is still in an early phase. While helping students to answer standard conceptual questions is certainly important and meaningful, it is possible that failure to provide further learning support may also prevent students from benefiting fully from QA systems.

However, our findings also show that using the QA system significantly boosts the positive effect of communication with instructors on both the learning process and the learning outcome. It is possible that when basic questions can be solved effectively by the QA system, instructors will have enough resources to focus on those thought-provoking questions. In other words, students can build a solid knowledge base by having their simpler questions answered by the QA system and think more deeply to further develop their knowledge structure by communicating with instructors. Hence, QA systems and communication with instructors complement each other as learning channels, leading to a better learning process and outcome.

6.2 Limitations and future work

Our work is not without limitations. First, research has suggested that over time even very lean media can be perceived as rich, once users learn how to use them correctly and efficiently (Scherer *et al.*, 2015). As the capability of QA systems continuously improves and users become more familiar with the interaction style, users' perception of the systems may change. Hence, an important future research direction is to examine how users' interaction with the system and the impact of system use on their learning may change over time. Second, the participants in

Table 7.
Regression results for
identification and
robustness checks (2)

Variable	Dependent variable: Learning process				Dependent variable: Learning outcome				Low <i>C_Efficacy</i>	
	Male	Female	High <i>W_Questions</i>	Low <i>W_Questions</i>	High <i>W_Questions</i>	Female	Male	High <i>W_Questions</i>		Low <i>W_Questions</i>
<i>F_Systems</i>	-0.038 (0.039)	0.037 (0.067)	-0.065 (0.114)	0.075 (0.053)	-0.021 (0.044)	0.128 * (0.053)	-0.030 (0.044)	-0.103 (0.118)	0.066 (0.055)	-0.128
<i>C_Instructors</i>	0.312*** (0.083)	0.257** (0.074)	0.373* (0.176)	0.365** (0.138)	0.412* (0.160)	0.206* (0.088)	0.238* (0.083)	0.195* (0.078)	0.245* (0.070)	0.016 -
<i>F_Systems</i> × <i>C_Instructors</i>	0.040* (0.035)	0.066* (0.031)	0.077* (0.035)	0.088** (0.031)	0.052** (0.019)	0.065** (0.024)	0.047* (0.021)	0.071* (0.026)	0.088*** (0.034)	0.084
Control variables	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Observations	66	86	76	76	76	76	66	76	76	76
<i>R</i> -squared	0.423	0.313	0.284	0.305	0.321	0.401	0.293	0.264	0.340	0.475
Adjusted <i>R</i> -squared	0.342	0.241	0.186	0.211	0.229	0.320	0.254	0.164	0.250	0.404

Note(s): $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; standard errors in parentheses

this study were college students in a renowned university in a large city, and it is common for students to use digital tools to support their learning. Future research may want to extend the current study by examining the effect of multi-learning channels on other types of students.

6.3 Implications of the findings

This study has several theoretical and practical implications. Research on learning technology has accrued over the past few decades, and recent research on how education tools affect students' experience and performance is increasing (Luo *et al.*, 2019). In particular, most research thus far has examined the effect of technological tools *per se*, but has ignored how technology may change or interact with traditional learning channels (Gupta and Bostrom, 2009). Indeed, given the merits of traditional learning channels and the supporting role of technologies, it is important to understand how new technology may reinforce traditional learning channels or how different learning channels may be integrated to strengthen each other's effects. This study reveals a positive interplay between a QA system and communication with instructors. Our results thus provide an important extension to previous research by suggesting that incorporating new technologies and taking advantage of traditional approaches are both essential for achieving a good learning experience. More broadly, this study is also highly relevant to the lively discussion on how humans and machines evolve in working together (Amershi *et al.*, 2019; Rahwan *et al.*, 2019). Our findings suggest that integrating both human and machine communication channels is important, as student-instructor communication will be more effective with the presence of QA systems.

Furthermore, this study theorizes the mechanisms that underlie the interplay between two different learning channels based on MRT. MRT has been widely used in research on marketing, services, and information communication technology. This study represents an important application of MRT to the learning context by showing that different QA channels can be used to solve different types of questions and complement each other in improving learning processes and outcomes. In particular, QA systems serve as lean media that can efficiently solve fact-based questions, while instructors serve as rich media that are good for solving thought-provoking questions. The positive interplay between the two channels holds because students' use of QA systems to resolve fact-based questions will free instructor resources for resolving more difficult, personalized questions, hence helping to make the best use of instructor communication.

In terms of its practical implications, this study suggests that using QA systems may not always lead to better learning performance. Specifically, the insights from our research indicate that educators need to consider the capabilities of educational tools as well as their influence on students' specific learning activities, and hence to understand their roles in teaching and learning. It is also possible that as the capability of QA systems improves and students gain more experience in using them, the role of QA systems will change accordingly. Nonetheless, our study shows a positive interaction between using QA systems and communication with instructors. This means that with the presence of QA systems, instructors can play an even more important and effective role in improving students' learning performance. Instructors may want to design and adapt teaching methods accordingly, for example, by encouraging students to resolve basic questions via online channels while allocating more time for in-depth discussion and probing in order to improve students' learning processes and outcomes.

Notes

1. <https://www.summitlearning.org>
2. For the rest of this paper, we use "QA systems" to refer to QA systems with natural language interfaces.

3. An item was dropped for each construct due to a cross-loading issue.
4. For brevity, from this point onward, we only report the major variables of interest.

References

- Afzal, S., Dempsey, B., Helon, C., Mukhi, N., Pribic, M., Sickler, A., Strong, P., Vanchiswar, M. and Wilde, L. (2019), "The personality of AI systems in education: experiences with the Watson tutor, a one-on-one virtual tutoring system", *Childhood Education*, Vol. 95 No. 1, pp. 44-52.
- Agarwal, R. and Prasad, J. (1998), "A conceptual and operational definition of personal innovativeness in the domain of information technology", *Information Systems Research*, Vol. 9 No. 2, pp. 204-215.
- Aguirre, C.C., Kloos, C.D., Alario-Hoyos, C. and Muñoz-Merino, P.J. (2018), "Supporting a MOOC through a conversational agent. Design of a first prototype", *2018 International Symposium on Computers in Education*, IEEE Computer Society, Washington, DC, pp. 1-6.
- Alavi, M. and Leidner, D.E. (2001), "Research commentary: technology-mediated learning-a call for greater depth and breadth of research", *Information Systems Research*, Vol. 12 No. 1, pp. 1-10.
- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N. and Inkpen, K. (2019), "Guidelines for human-AI interaction", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, pp. 1-13.
- Arbaugh, J.B. and Duray, R. (2002), "Technological and structural characteristics, student learning and satisfaction with web-based courses: an exploratory study of two on-line MBA programs", *Management Learning*, Vol. 33 No. 3, pp. 331-347.
- Bandura, A., O'Leary, A., Taylor, C.B., Gauthier, J. and Gossard, D. (1987), "Perceived self-efficacy and pain control: opioid and nonopioid mechanisms", *Journal of Personality and Social Psychology*, Vol. 53 No. 3, pp. 563-571.
- Barak, M. and Rafaeli, S. (2004), "On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning", *International Journal of Human-Computer Studies*, Vol. 61 No. 1, pp. 84-103.
- Bhattacharjee, Y. (2009), "A personal tutor for algebra", *Science*, Vol. 323 No. 5910, pp. 64-65.
- Chin, C. (2002), "Student-generated questions: encouraging inquisitive minds in learning science", *Teaching and Learning*, Vol. 23 No. 1, pp. 59-67.
- Chin, C. and Brown, D.E. (2002), "Student-generated questions: a meaningful aspect of learning in science", *International Journal of Science Education*, Vol. 24 No. 5, pp. 521-549.
- Crockett, K., Latham, A. and Whitton, N. (2017), "On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees", *International Journal of Human-Computer Studies*, Vol. 97 No. 1, pp. 98-115.
- Daft, R.L. and Lengel, R.H. (1986), "Organizational information requirements, media richness and structural design", *Management Science*, Vol. 32 No. 5, pp. 554-571.
- Davey, B. and McBride, S. (1986), "Effects of question-generation training on reading comprehension", *Journal of Educational Psychology*, Vol. 78 No. 4, pp. 256-262.
- Dermentzi, E. and Papagiannidis, S. (2018), "Academics' intention to adopt online technologies for public engagement", *Internet Research*, Vol. 28 No. 1, pp. 191-212.
- Dewey, J. (1913), *Interest and Effort in Education*, Houghton Mifflin, Boston, MA.
- Dong, J.Q. and Zhang, X. (2011), "Gender differences in adoption of information systems: new findings from China", *Computers in Human Behavior*, Vol. 27 No. 1, pp. 384-390.

- Fryer, L.K., Nakao, K. and Thompson, A. (2019), "Chatbot learning partners: connecting learning experiences, interest and competence", *Computers in Human Behavior*, Vol. 93 No. 1, pp. 279-289.
- Grudin, J. and Jacques, R. (2019), "Chatbots, Humbots, and the quest for artificial general intelligence", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland, pp. 1-11.
- Gupta, S. and Bostrom, R.P. (2009), "Technology-mediated learning: a comprehensive theoretical model", *Journal of the Association for Information Systems*, Vol. 10 No. 9, pp. 686-714.
- Gupta, S., Bostrom, R.P. and Huber, M. (2010), "End-user training methods: what we know, need to know", *ACM SIGMIS Database: The Database for Advances in Information Systems*, Vol. 41 No. 4, pp. 9-39.
- Harper, B. (2018), "Technology and teacher-student interactions: a review of empirical research", *Journal of Research on Technology in Education*, Vol. 50 No. 3, pp. 214-225.
- Hayashi, Y. (2014), "Togetherness: multiple pedagogical conversational agents as companions in collaborative learning", *International Conference on Intelligent Tutoring Systems*, Springer, pp. 114-123.
- Hien, H.T., Cuong, P.-N., Nam, L.N.H., Nhung, H.L.T.K. and Thang, L.D. (2018), "Intelligent assistants in higher-education environments: the fit-eBot, a chatbot for administrative and learning support", *Proceedings of the 9th International Symposium on Information and Communication Technology*, Danang City, pp. 69-76.
- Hussain, S. and Athula, G. (2018), "Extending a conventional chatbot knowledge base to external knowledge source and introducing user based sessions for diabetes education", *The 32nd International Conference on Advanced Information Networking and Applications Workshops*, IEEE Computer Society, Washington, DC, pp. 698-703.
- Kahai, S.S. and Cooper, R.B. (2003), "Exploring the core concepts of media richness theory: the impact of cue multiplicity and feedback immediacy on decision quality", *Journal of Management Information Systems*, Vol. 20 No. 1, pp. 263-299.
- Kock, N. (2009), "Information systems theorizing based on evolutionary psychology: an interdisciplinary review and theory integration framework", *MIS Quarterly*, Vol. 33 No. 2, pp. 395-418.
- Kokku, R., Sundararajan, S., Dey, P., Sindhgatta, R., Nitta, S. and Sengupta, B. (2018), "Augmenting classrooms with AI for personalized education", *International Conference on Acoustics, Speech and Signal Processing*, IEEE Computer Society, Washington, DC, pp. 6976-6980.
- Lee, L.-K., Fung, Y.-C., Pun, Y.-W., Wong, K.-K., Yu, M.T.-Y. and Wu, N.-I. (2020), "Using a multiplatform chatbot as an online tutor in a university course", *International Symposium on Educational Technology*, IEEE Computer Society, Washington, DC, pp. 53-56.
- Luo, X., Tong, S., Fang, Z. and Qu, Z. (2019), "Frontiers: machines vs Humans: the impact of artificial intelligence chatbot disclosure on customer purchases", *Marketing Science*, Vol. 38 No. 6, pp. 937-947.
- Mervin, R. (2013), "An overview of question answering system", *International Journal of Research in Advanced Technology*, Vol. 1 No. 1, pp. 75-80.
- Min, W., Park, K., Wiggins, J., Mott, B., Wiebe, E., Boyer, K.E. and Lester, J. (2019), "Predicting dialogue breakdown in conversational pedagogical agents with multimodal lstms", *International Conference on Artificial Intelligence in Education*, Springer, pp. 195-200.
- Pizzini, E.L. and Shepardson, D.P. (1991), "Student questioning in the presence of the teacher during problem solving in science", *School Science and Mathematics*, Vol. 91 No. 8, pp. 348-352.
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F. and Breazeal, C. (2019), "Machine behaviour", *Nature*, Vol. 568 No. 7753, pp. 477-486.

- Rosenshine, B., Meister, C. and Chapman, S. (1996), "Teaching students to generate questions: a review of the intervention studies", *Review of Educational Research*, Vol. 66 No. 2, pp. 181-221.
- Scardamalia, M. and Bereiter, C. (1992), "Text-based and knowledge based questioning by children", *Cognition and Instruction*, Vol. 9 No. 3, pp. 177-199.
- Scherer, A., Wunderlich, N.V. and Wangenheim, F.V. (2015), "The value of self-service: long-term effects of technology-based self-service usage on customer retention", *MIS Quarterly*, Vol. 39 No. 1, pp. 177-200.
- Schmullian, A. and Coetzee, S.A. (2019), "The development of messenger bots for teaching and learning and accounting students' experience of the use thereof", *British Journal of Educational Technology*, Vol. 50 No. 5, pp. 2751-2777.
- Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I. and Lowry, P.B. (2020), "Collaborating with technology-based autonomous agents: issues and research opportunities", *Internet Research*, Vol. 30 No. 1, pp. 1-18.
- Sharma, Y. and Gupta, S. (2018), "Deep learning approaches for question answering system", *Procedia Computer Science*, Vol. 132 No. 1, pp. 785-794.
- Söllner, M., Bitzer, P., Janson, A. and Leimeister, J.M. (2018), "Process is king: evaluating the performance of technology-mediated learning in vocational software training", *Journal of Information Technology*, Vol. 33 No. 3, pp. 233-253.
- Song, D., Oh, E.Y. and Rice, M. (2017), "Interacting with a conversational agent system for educational purposes in online courses", *The 10th International Conference on Human System Interactions*, IEEE Computer Society, Washington, DC, pp. 78-82.
- Tang, D., Duan, N., Yan, Z., Zhang, Z., Sun, Y., Liu, S., Lv, Y. and Zhou, M. (2018), "Learning to collaborate for question answering and asking", *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, LA, pp. 1564-1574.
- Waltinger, U., Breuing, A. and Wachsmuth, I. (2012), "Connecting question answering and conversational agents", *KI-künstliche Intelligenz*, Vol. 26 No. 4, pp. 381-390.
- Winkler, R. and Söllner, M. (2018), "Unleashing the potential of chatbots in education: a state-of-the-art analysis", *Paper Presented at Meeting of the Academy of Management*, 9 July, Chicago, IL, available at: https://www.alexandria.unisg.ch/254848/1/JML_699.pdf (accessed 10 June 2020).
- Wu, H.K., Lin, C.H., Ou, Y.Y., Liu, C.Z. and Chao, C.Y. (2020), "Advantages and constraints of a Hybrid model K-12 E-learning assistant chatbot", *IEEE Access*, Vol. 8 No. 1, pp. 77788-77801.
- Xu, D., Huang, W.W., Wang, H. and Heales, J. (2014), "Enhancing E-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: an empirical investigation", *Information and Management*, Vol. 51 No. 4, pp. 430-440.
- Yu, F.-Y. (2009), "Scaffolding student-generated questions: design and development of a customizable online learning system", *Computers in Human Behavior*, Vol. 25 No. 5, pp. 1129-1138.
- Zhang, C.A., Dai, J. and Vasarhelyi, M.A. (2018), "The impact of disruptive technologies on accounting and auditing education: how should the profession adapt?", *The CPA Journal*, Vol. 88 No. 9, pp. 20-26.
- Zimmerman, B.J. and Schunk, D.H. (2001), *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, Routledge, New York, NY.

Corresponding author

Cheng Yi can be contacted at: yich@sem.tsinghua.edu.cn

Author and year	QA systems	Research foci
Aguirre et al. (2018)	Voice QA assistant for online JAVA courses	Implemented a voice-based QA system on a MOOC platform and optimized its accuracy
Crockett et al. (2017)	Giving personalized tutoring materials	Proposed a new method to model users' learning styles and hence to provide better-suited tutoring materials
Fryer et al. (2019)	Language partner	Conducted a field experiment and found that practicing language with a QA system increased students' interest in the language course
Hayashi (2014)	Facilitating group learning	Conducted a field experiment and found that using QA systems increased students' social awareness and engagement in collaborative learning
Hien et al. (2018)	Providing administrative support	Adopted various artificial intelligence techniques to enhance QA system performance
Hussain and Athula (2018)	QA tasks for diabetes education	Extended QA system's knowledge base to an external knowledge source, Wikipedia, thereby increasing response rate
Kokku et al. (2018)	General QA tasks on a e-learning platform	Implemented a QA system to provide personalized tutoring, leading to increased student engagement and knowledge absorption
Lee et al. (2020)	General QA tasks on a e-learning platform	Developed a QA system to answer questions on course materials and administrative matters, which improved students' satisfaction with the learning process
Min et al. (2019)	Chatting with students in game-based learning activities	Incorporated a deep-learning method to choose appropriate responses, and hence enhanced users' QA experiences
Sharma and Gupta (2018)	No specific application	Analyzed, implemented and improved various popular methodologies in the field of question answering
Song et al. (2017)	Chatting with students on a MOOC platform	Developed a QA system to promote students' meaningful interaction to engage them in online learning
Tang et al. (2018)	No specific application	Proposed a new training algorithm, which effectively improved the accuracy of QA systems
Waltinger et al. (2012)	No specific application	Enhanced QA systems' response time and accuracy by considering contextual information

Table A1.
Literature review on
QA systems in
education

Appendix 2

Variable	Measures	Source
<i>L_Process</i> (learning process)	I Feel that the learning process of this course is very pleasant I Am very satisfied with my learning process of this course	Post-questionnaire (Xu <i>et al.</i> , 2014)
<i>L_Outcome</i> (learning outcome)	I Have understood all the learning materials related to this course I Have achieved good learning results related to this course	Post-questionnaire (Söllner <i>et al.</i> , 2018)
<i>U_Systems</i> (use of QA system)	1 – If the student used the system to raise at least one question related to learning content 0 – Otherwise	System log
<i>F_Systems</i> (frequency of using QA systems)	The total number of questions that related to learning content raised by each student	System log
<i>C_Instructors</i> (communication with instructors)	Self-evaluation of the frequency of communication with instructors 1 – Hardly ever, 2 – rarely, 3 – occasionally, 4 – sometimes, 5 – modestly frequently, 6 – frequently, 7 – very frequently	Post-questionnaire
<i>F_Materials</i> (frequency of checking online learning materials)	Self-evaluation of the frequency of checking online learning materials 1 – Hardly ever, 2 – rarely, 3 – occasionally, 4 – sometimes, 5 – modestly frequently, 6 – frequently, 7 – very frequently	Post-questionnaire
<i>F_Peers</i> (frequency of asking peers)	Self-evaluation of the frequency of discussion with classmates 1 – Hardly ever, 2 – rarely, 3 – occasionally, 4 – sometimes, 5 – modestly frequently, 6 – frequently, 7 – very frequently	Post-questionnaire
<i>F_Books</i> (frequency of studying textbooks)	Self-evaluation of the frequency of studying textbooks 1 – Hardly ever, 2 – rarely, 3 – occasionally, 4 – sometimes, 5 – modestly frequently, 6 – frequently, 7 – very frequently	Post-questionnaire
<i>E_Tools</i> (experience of using learning tools)	I have rich experience of using technology tools to support learning	Pre-questionnaire (Arbaugh and Duray, 2002)
<i>I_IT</i> (innovativeness in IT)	Compared to my classmates, I am the one who would like to try new IT products first I like experiencing novel IT products I'm not willing to try new IT products	Pre-questionnaire (Agarwal and Prasad, 1998)
<i>PL</i> (playfulness)	I think experiencing IT products is fun I'm willing to explore the functions of IT products I'm willing to try new functions of IT products with my imagination I interact with IT products in a flexible way	Pre-questionnaire (Dewey, 1913)
<i>W_Questions</i> (willingness to ask questions)	I often ask questions when learning I think asking questions is a useful way to learn	Pre-questionnaire (Davey and McBride, 1986)

Table A2.
Measurement items

(continued)

Note(s): Seven-point Likert Scale

Table A2.

Matching Method	Variable	Unmatched		Mean		% Bias	% Reduced bias	<i>t</i>	<i>p</i> -value
		Matched	Treated	Control					
One-to-one nearest (without replacement)	Gender	<i>U</i>	0.540	0.370	29.6		61.1	1.86	0.065
		<i>M</i>	0.514	0.571	-11.5			-0.47	0.637
	<i>W_Questions</i>	<i>U</i>	3.686	4.24	-33.8			-1.86	0.064
		<i>M</i>	3.686	3.714	-1.7		94.8	-0.07	0.942
	<i>C_Efficacy</i>	<i>U</i>	2.678	3.336	-40.5			-2.18	0.031
One-to-one nearest (with replacement)	Gender	<i>M</i>	2.678	2.886	-12.8		68.5	-0.53	0.600
		<i>U</i>	0.540	0.370	29.6			1.86	0.065
	<i>W_Questions</i>	<i>M</i>	0.469	0.563	-11.5		36.1	-0.47	0.461
		<i>U</i>	3.686	4.241	-33.8			-1.86	0.064
	<i>C_Efficacy</i>	<i>M</i>	3.938	3.938	0.0		100.0	-0.07	1.000
One-to-two nearest (with replacement)	Gender	<i>U</i>	2.678	3.336	-40.5			-2.18	0.031
		<i>M</i>	2.931	2.886	-17.3		57.2	-0.53	0.454
	<i>W_Questions</i>	<i>U</i>	0.540	0.370	29.6			1.86	0.065
		<i>M</i>	0.469	0.485	-3.2		89.4	-0.12	0.902
	<i>C_Efficacy</i>	<i>U</i>	3.686	4.241	-33.8			-1.86	0.064
One-to-three nearest (with replacement)	Gender	<i>M</i>	3.938	3.828	6.7		80.2	0.28	0.778
		<i>U</i>	2.678	3.336	-40.5			-2.18	0.031
	<i>W_Questions</i>	<i>M</i>	2.93	3.078	-9.1		77.4	-0.41	0.684
		<i>U</i>	0.540	0.370	29.6			1.86	0.065
	<i>C_Efficacy</i>	<i>M</i>	0.469	0.474	-1.1		96.5	-0.04	0.967
One-to-four nearest (with replacement)	Gender	<i>U</i>	3.686	4.241	-33.8			-1.86	0.064
		<i>M</i>	3.938	3.847	5.6		83.5	0.24	0.814
	<i>W_Questions</i>	<i>U</i>	2.678	3.336	-40.5			-2.18	0.031
		<i>M</i>	2.931	3.028	-6.0		85.1	-0.27	0.787

Table A3.
PSM Variables and
t-test Results