Education Tech Research Dev (2023) 71:2483–2506
https://doi.org/10.1007/s11423-023-10265-6

CULTURAL AND REGIONAL PERSPECTIVES

Understanding museum visitors' question-asking through a mobile app

Man Su[1] · Jesse Ha[2] · Luis E. Pérez Cortés[3] · Jeremy Bernier[1] · Lin Yan[1] · Brian C. Nelson[1] · Judd D. Bowman[4] · Catherine D. D. Bowman[4]

Accepted: 22 June 2023 / Published online: 10 July 2023

Abstract

Question-asking is essential for reasoning, understanding, and investigating scientific problems within and beyond traditional classrooms. Nevertheless, questions generated in formal and informal learning environments can be infrequent and unsophisticated. This study explores museum visitors' question-asking quality by considering their interactions with two different modes of a question-asking mobile app (Ask or Game Mode) in two different museum environments (linear non-interactive or non-linear interactive exhibits). Results showed that visitors' question-asking quality was influenced by app modes and by museum environments. Specifically, we found that visitors' question-asking quality was significantly higher when using the gamified version of the app (Game Mode) compared to a non-gamified version (Ask Mode) in a linear non-interactive exhibit. Findings also revealed that question-asking performance could be significantly influenced by instrumental factors (such as app performance in answering questions) and socio-contextual factors (such as visitor group inquiry frequency). The study provides fundamental and comprehensive insights for designing active learning environments by considering the influential factors of question-asking.

Keywords Question-asking · Mobile app · Gamification · Museum education · Informal learning environments

✉ Man Su
mansu@asu.edu

[1] Mary Lou Fulton Teachers College, Arizona State University, Payne Hall, 122K, 1000 S Forest Mall, Tempe, AZ 85287-2111, USA

[2] UCLA School of Education & Information Studies, Los Angeles, USA

[3] Learning Research and Development Center, University of Pittsburgh, Pittsburgh, USA

[4] School of Earth and Space Exploration, Arizona State University, Tempe, USA

Introduction

Question-asking is one of the essential processes of scientific inquiry (Bybee, 2006) and, as an active learning strategy, is closely associated with academic achievement (Cano et al., 2014) and cognitive engagement (Chi, 2021). However, researchers have noted that question-asking is often infrequent and unsophisticated in formal and informal learning environments (Bradley et al., 2008; Graesser & Person, 1994; Jirout & Klahr, 2011). According to Graesser and Person (1994), unsophisticated questions are Shallow, short-answer questions that address the content and interpretation of explicit material; they are rarely high-level questions that involve inferences, multistep reasoning, the application of an idea to a new domain of knowledge, the synthesis of a new idea from multiple information sources, or the critical evaluation of a claim. (p. 105).

Recently, artificial intelligence (AI) such as ChatGPT has made its way into the general public's awareness. The potentials and dangers of such AI technologies have seemingly catalyzed heightened interest in the quality of questions that people ask. The focus on learners' questioning quality is no longer solely an issue in educational practice but has become an important and ongoing research topic, which aims to cultivate more individuals who can explore and solve complex problems and engage in critical thinking. Thus, understanding learners' questioning quality is essential for preparing them for the demands of the future workforce in the age of AI (Abdelghani et al., 2022; Humphries & Ness, 2015).

Recent works have also highlighted the importance of question-asking in various educational contexts. Museums have been identified as a rich source of inquiry-based learning opportunities that can foster higher-order thinking skills and engagement (Allen, 2004; Gutwill & Allen, 2012). Research has been conducted from various perspectives to investigate the role of questioning in museum visits. For example, studies have examined the conversational and questioning styles of parents (Ash, 2004), considered the use of guided questions (Hohenstein & Tran, 2007), and analyzed the questioning techniques used by museum educators (Braund & Lelliott, 2017; Tigert et al., 2021) in fostering participation, engagement, and higher inquiry skills during museum visits (Allen & Gutwill, 2009).

Despite the potential of museums to facilitate inquiry-based learning, there is a research gap in the analysis of visitors' question-asking quality in museum settings and the factors that can influence it. The present study aims to address this gap by making both theoretical and practical contributions on understanding visitors' question-asking quality. Specifically, the goal of the study is twofold. First, it aims to build on existing frameworks that examine influential factors of question-asking at museums. Second, this study aims to explore visitor groups' question-asking quality using a mobile app, Dr. Discovery (hereafter the Dr. D app).

The Dr. D app features two modes: Game Mode and Ask Mode. The Game Mode is designed to include simple casual game mechanics (e.g., earning digital badges by asking more questions). In comparison, the Ask Mode provides a baseline version which is essentially a simple question-asking app with no gamified elements. By using these modes, this study investigated the impact of different versions of a mobile app on visitors' question-asking quality in different museums located in the American Southwest: a linear non-interactive exhibit at Museum A and a non-linear interactive exhibit at Museum B. In doing so, we aimed to contribute to the development of practical strategies to enhance visitors' questioning skills and museum learning experiences.

In the following section, we discuss the theoretical frameworks underpinning this study and prior work that led to the formation of our research questions and hypotheses.

Theoretical framework and relevant literature

Frameworks for assessing question quality

Our prior work (Ha et al., 2021) critically reviewed existing frameworks and coding schema for assessing informal learning environments (Borun, 2002; Tissenbaum et al., 2017). These frameworks tend to evaluate the general aspects of learning behaviors or collaborative interactions occurring at museums rather than providing a hierarchical framework to clearly distinguish question-asking quality that would indicate widely different levels of cognitive engagement. Given that learning principles from formal environments could also be used to enhance learning in informal settings, we thereby considered frameworks and schema for analyzing the qualitative characteristics of questions in formal learning environments (Gallagher & Aschner, 1963; Graesser & Person, 1994; Scardamalia & Bereiter, 1992). One common feature of the question-asking frameworks used in classrooms is that they classify knowledge-based questions into multiple levels. For example, Gallagher and Ascher (1963) classified knowledge-based questions as low-level versus high-level. Low-level questions are memory-based or convergent-thinking questions that relate ideas by comparing, contrasting, or explaining. In contrast, high-level questions are more profound questions that involve divergent thinking to predict, infer, hypothesize, reconstruct information, incorporate new knowledge, and make subjective judgments. Graesser and Person (1994) later classified questions into 18 types. For instance, "low-level, convergent-thinking questions" include "concept completion" and "definition" (p. 110) types of questions that can be answered with a word or a phrase. In contrast, "high-level, divergent-thinking questions" contain "causal consequence", "goal orientation", and "judgmental" (p. 111) types of questions which often require long, elaborated answers with deeper reasoning. Similarly, Scardamalia and Bereiter (1992) categorized questions as either "basic information" questions that are asked in need of information to orient towards a topic or "wonderment" questions that "reflect curiosity, puzzlement, skepticism, or a knowledge-based speculation" (p. 187).

To our knowledge, there is no existing framework that guides the assessment of question quality in museums. Drawing upon previous works in both formal and informal learning environments, the present study proposes a framework that can be used to categorize visitor-generated questions into various types, thereby comprehensively understanding visitors' question-asking quality at museums.

Understanding influential factors for question-asking at museums

To understand visitors' question-asking performance (i.e., the question quality and quantity) at museums, it is critical to explore the factors that are associated with question-asking mechanisms (Graesser & Person, 1994; Ronfard et al., 2018). According to Ronfard et al. (2018), question-asking mechanisms can be explained from a process-oriented perspective and can be broken down into four stages: initiation, formulation, expression, and evaluation. "Initiation" involves the understanding that there is some information attainable by asking a question; "formulation" involves identifying the information needed and phrasing a question requesting that information; and "expression" is deciding whether to ask the question, considering both whether there is a reliable source of information and if the context is appropriate for asking a question. These three stages are followed by the fourth

stage, "evaluation" of the response received, which determines whether the sequence will end, a question will be restated, or a new question will be asked. These four stages constitute a model for elaborating the question-asking mechanism, which is essentially an iterative and dynamic process with different components influencing one another without a specific order. Therefore, one perplexing issue for museum researchers to consider is how to provide learning scaffolds at different stages of the question-generation process that encourage visitors to spontaneously engage in the iterative and dynamic process of question-asking. In theory, visitors may benefit from the scaffolds and be able to generate more high-level or deeper questions (Gutwill & Allen, 2012; Yoon et al., 2013).

With the advancement of technologies, museums are rapidly adopting diverse forms of mobile technologies to scaffold learning and offer diverse opportunities for visitors to interact and learn (Charitonos et al., 2012; Harley et al., 2016; Lee et al., 2020). Despite different modalities, promoting inquiry-based thinking and question-asking skills through interaction with exhibits is an essential goal (Efstathiou et al., 2018; Tissenbaum, 2020). For example, researchers closely examined different design elements to enhance inquiry skills afforded by mobile devices (Jong et al., 2018; Liu et al., 2021; Suarez et al., 2018). Gamification has been utilized as a design element in museum apps to enhance visitor engagement, learning, and enjoyment (Nelson et al., 2020; Squire & Jan, 2007). One popular gamification element that has been integrated into museum mobile apps is the use of digital badges (Cordova-Rangel & Caro, 2023; Negruşa et al., 2015). Research has shown that the use of digital badges can effectively motivate visitors to explore and engage with the exhibits more thoroughly, learn more about the collections, and provide a sense of accomplishment and recognition (Jovanovic & Devedzic, 2015). However, it is important to note that the use of digital badges needs to be combined with other educational strategies (i.e., promoting question-asking through other means such as prompts) to create a comprehensive and engaging museum visiting experience. Consequently, the design of a mobile device in terms of whether it contains gamified elements (e.g., digital badges), and the usability or actual performance of the device are instrumental factors that could potentially influence visitors' question-asking performance (Gutwill & Allen, 2012).

In addition to the instrumental factors, the literature also suggests considering socio-contextual factors: the physical learning environments and the social settings (Vartiainen & Enkenberg, 2013; Wang et al., 2017). The physical layout of a museum can have a significant impact on visitors' experience and question expression in museums. For example, exhibits that offer multiple hands-on and interactive investigations appeared to increase visitors' tendency to engage in inquiry behaviors (Allen & Gutwill, 2009) whereas overcrowding for a linear non-interactive exhibit may pressure visitors to quickly end their inquiry and move on to the next exhibit (Gutwill & Allen, 2012). A recent study (Pérez Cortés et al., 2023) compared the effects of linear versus non-linear interactive museum layouts on visitors' question-asking performance. In this study, a linear museum layout involves a sequential arrangement of exhibits in a single path or route. Visitors follow a predetermined path and view the exhibits in a fixed order. In contrast, a non-linear interactive layout provides visitors with more flexibility and autonomy to choose their own path and explore the exhibits in a non-sequential way. More research is needed to examine whether visitors in a non-linear interactive exhibit had higher levels of engagement, higher levels of retention and recall, spent more time in the museum, and asked more questions compared to those in the linear exhibit.

Moreover, the expression of questions also involves considerations of social settings, such as considering whether someone is available to provide reliable information; and if such a person is available, whether it is acceptable to request information from this person

given the social context (Graesser & Person, 1994). In museums, visitors may not articulate their questions because they are concerned with interrupting others, challenging the views of adults or more experienced group members, or cannot find a docent or museum professional (Allen & Gutwill, 2009). To create a social setting favorable for question expression, researchers found that parents who used a conversational style and open-ended questions tended to have children who were more engaged in inquiry behaviors than those whose parents took on a teacher-like role and used a structured question-asking pattern (Callanan et al., 2021; Hohenstein & Tran, 2007). Recent works on socio-contextual factors of informal learning primarily focused on examining conversations, dialogues, and other social collaborative interactions among learners and groups of learners (Ha et al., 2021; Tissenbaum, 2020; Zimmerman & Land, 2022). The present study shifted the focus to an equally critical examination of socio-contextual factors on visitor groups' question-asking quality and performance at museums.

Furthermore, to formulate a question, one must know what is missing and what is necessary for further understanding. Therefore, it is critical to understand cognitive factors associated with question formulation, such as the level of epistemic curiosity. Epistemic curiosity is a psychological need to acquire knowledge to fill a knowledge gap or correct a knowledge deficit, that motivates question formulation (Jirout & Klahr, 2011; Ram, 1991). In other words, how learners ask questions depends on how curious they are in terms of what they are learning or observing, or whether they can identify their knowledge gaps. In Alaimi et al. (2020), children who were more curious by trait asked more complex divergent-thinking questions. Similarly, Jirout and Klahr (2011) found a positive relationship between children's curiosity and their range of question-asking abilities, indicating that more curious children ask more questions, and are better at generating higher-level questions.

In summary, the literature suggests researchers explore three main factors that might be associated with visitors' question-asking performance at museums: instrumental, socio-contextual, and cognitive factors. Adopting the process-oriented perspective, these three factors might influence question-asking performance at the initiation, formulation, expression, and evaluation stage (Ronfard et al., 2018). Considering previous research, we investigated through a quantitative analytical approach and considered multiple factors that might influence visitors' question-asking performance. Specifically, we examined the influence of cognitive factors (i.e., the level of epistemic curiosity), instrumental factors (i.e., the app mode, the app performance in answering questions), and socio-contextual factors (i.e., exhibit layout, visitor group question-asking frequency) on museum visitors' question-asking quality and performance.

The present study

This study is the first to focus on visitors' question-asking quality at different museums using different versions of a mobile app. We explored three research questions:

RQ1: What is the overall question-asking quality and performance among visitor groups with distinct demographic characteristics using the app?
RQ2: Does the effect of different modes of the mobile app (Ask mode vs. Game mode) on visitors' question-asking quality vary by whether visitors engage in a linear non-interactive exhibit versus a non-linear interactive exhibit (Museum A vs. Museum B)?

RQ3: What are the factors that impact museum visitors' question-asking performance (considering both quality and quantity) when using a mobile app? To what degree do these factors impact visitors' question-asking performance?

For RQ1, we intended to present data on the overall question-asking quality captured by our app. For RQ2, we first hypothesized that visitors who used the Game Mode of the app would be more engaged in asking higher-level questions than those using the Ask Mode given that the literature documented the positive effect of gamification in fostering learning and engagement (Sailer & Homner, 2020; Zainuddin et al., 2020). However, after considering the effects of using the app at two museums with different exhibit layouts (linear non-interactive vs. non-linear interactive), we expected that the effect of the mobile app would vary depending on different exhibit layouts. Studies have shown that an exhibit with an open-room layout and interactive facilities is already engaging (Allen & Gutwill, 2009; Gutwill & Allen, 2012), so we assumed that the use of Game mode of the app would not substantially affect visitors' question-asking quality at the non-linear interactive exhibit. However, it is likely that using the Game mode of the app would have a significant influence on question-asking quality in a linear non-interactive exhibit. Specifically, we formed the following two hypotheses (H) for RQ2: Visitors who used the Game Mode of the Dr. D app will possess higher question-asking quality compared to those who used the Ask Mode at the linear non-interactive Museum A (H1); Visitors who used the Game Mode of the Dr. D app will demonstrate similar question-asking quality to those who used the Ask Mode at the non-linear interactive Museum B (H2).

For RQ3, the literature has posited cognitive factors (CF), instrumental factors (IF), and socio-contextual factors (SCF) as influential factors for question-asking performance in museums. However, the degree to which these factors could influence question-asking performance is yet to be known. Therefore, we hypothesized that (H3): Visitors' question-asking performance would be influenced by the level of epistemic curiosity (CF_1), the mode of the app (IF_1), the app performance in answering questions (IF_2), the museum layout (SCF_1), and visitor groups' question-asking frequency (SCF_2).

Through answering these questions and hypotheses, we hope to provide fundamental and comprehensive insights for designing and implementing mobile apps that can foster high-level questioning among visitors and improve their museum learning experience.

Description of the app

The Dr. Discovery (Dr. D) app, created as part of a National Science Foundation-funded project examining the potential benefits of mobile technologies for supporting better understanding of visitor interactions with museum exhibits, features an interface in the form of an animated astronaut floating against a starfield. Visitors interact with the app through text in either English or Spanish. Spell correction is utilized on typed input. The Dr. D app is composed of two parts: (1) a front-end "virtual scientist" who responds to visitors' questions by providing vetted answers and (2) a back-end analytics portal to be mined by museum staff, evaluators, and researchers (Nelson et al., 2017).

Our app was designed with two modes: Ask Mode and Game Mode (Fig. 1). Ask Mode provided a baseline version which was essentially a simple question-asking app with no gamification elements. In comparison, Game Mode was designed with a gamification component (e.g., providing badges to reward visitors' question-asking behaviors based on the number of questions asked). The purpose of Game Mode was to

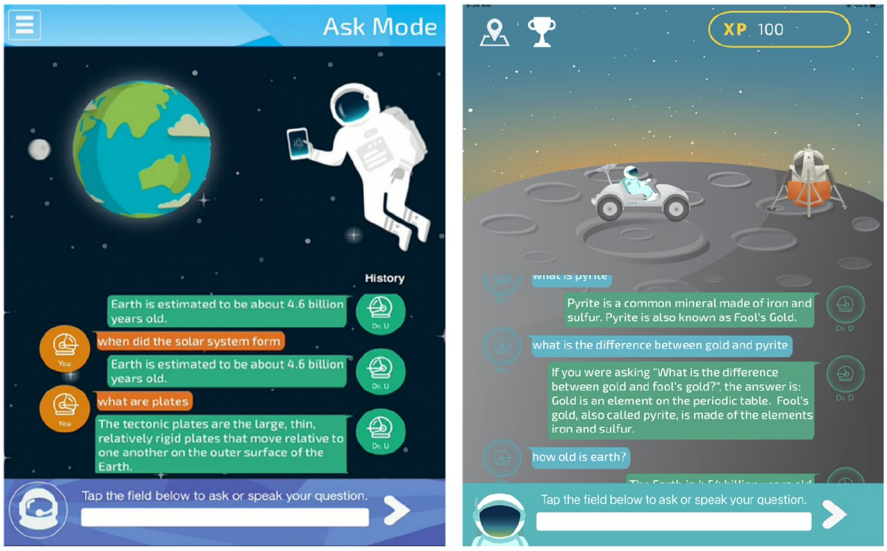

Fig. 1 The App Interface: Ask Mode (left), Game Mode (right)

encourage visitors to ask more questions about the exhibit content, providing valuable insights for museum personnel. However, there was a challenge in finding a balance between engaging visitors in the gamified app and ensuring they still fully engage with the physical museum experience. To address this, the Game Mode underwent several design iterations, aiming to make it engaging while keeping the focus on the exhibit content. In the Game Mode, players help Dr. D return to a space lander by asking questions, and with each question, Dr. D's vehicle moves forward. Unlike existing interventions that used carefully designed questioning prompts or scaffolds (Gutwill & Allen, 2012; Yoon et al., 2013), our app introduced Dr. D as a "virtual scientist" in the form of a gender-neutral astronaut who could answer visitors' questions anytime during their visits. This design could have potentially reduced the efforts put into expressing a question because the perceived social cost of asking a question to a mobile app was considerably lower than asking questions to a docent or others (Huang et al., 2017). This design also led to the inclusion of badges and trophies as rewards for specific question-asking achievements. Therefore, the study compared whether the gamified badging reward system could improve the quality of questions asked in the app.

In addition, both versions of the app used a tiered response system to interact with visitors by answering questions and could report the level of certainty in answering questions. For instance, when questions were perfectly matched to the database, the front-end of the app would reply with associated answers and the back-end portal would report tier "1" as a perfect match. If the question match was close, but not perfect, our app would respond "If you meant [question from the database], the answer is [answer from the database]." Then, the portal would document tier "2" for a close match. Samples that describe the tiered response system can be found in Appendix A.

Methods

Participants and settings

This study used a two (Game Mode vs. Ask Mode) by two (linear non-interactive exhibit vs. non-linear interactive exhibit) design. The research team introduced the Dr. D app to visitor groups and invited them to use it while exploring a designated exhibit hall at either of the museum sites. Participant groups decided when and where to use the project app while exploring the exhibits at their own pace. Different versions of the app were implemented on an alternating schedule such that on a given day all visitors would be offered one version of the app. When considering the contexts of the two museums (A and B), the average group size at both consisted of approximately four individuals, with a higher representation of women than men. The most common age range among group members fell within the "child" category, encompassing ages 5 to 12. In terms of survey respondents, more than two-thirds of the visitor groups are small-sized groups containing two to four people. Nearly two-thirds identified as female across both museum sites. Similarly, roughly two-thirds of the main participants reported their age to be between 26 and 45. Regarding racial and ethnic background, nearly two-thirds of primary participants identified as White, making it the largest demographic group across two sites. The Hispanic/Latinx community represented the second largest group, while the remaining racial and ethnic groups accounted for approximately five percent or less. More demographic information can be found in our recent paper (Pérez Cortés et al., 2023). In addition, Museum A features a linear non-interactive exhibit called "Origins" that occupies a long hallway from the entrance to the exit. In contrast, Museum B features a non-linear interactive exhibit about "Forces of Nature" that takes up an open room and semi-circular space (See Fig. 2). A total of 1, 331 visitor groups used our app during their visit to two museums by inputting at least one question. Visitors self-selected into their groups. Among these groups, 463 (35%) groups visited Museum A while 868 (65%) groups visited Museum B. A total of 510 groups (38%) used Ask Mode and 821 (62%) used Game Mode (See Table 1). There were four conditions involved in the study:

- Ask Mode of the app used in a linear, non-interactive exhibit (ASK_MUSEUM A)
- Ask Mode of the app used in a non-linear, interactive exhibit (ASK_MUSEUM B)
- Game Mode of the app used in a linear, non-interactive exhibit (GAME_MUSEUM A)

Fig. 2 The museum layout: museum A (left), museum B (right)

Table 1 Visitor group numbers by venue and app mode

	Ask mode	Game mode	Total
Museum A	214	249	463
Museum B	296	572	868
Total	510	821	1331

Note that despite allocating an equal number of implementation days for each mode at each museum, the number of participants in each category from our in-situ study was substantially unbalanced due to external influences. Specifically, there was a higher number of participant groups in the Game Mode at Museum B. This variation can be primarily attributed to two significant factors: the presence of vacation days for children and teenagers, and the proximity of certain days to major holidays

- Game Mode of the app used in a non-linear, interactive exhibit (GAME_MUSEUM B)

Procedure

At each museum, we implemented Ask and Game Modes in a rotational manner, ensuring that visitors were offered participation in only one condition on any given day. The implementation periods typically lasted for 3–4 h each day. We had a total of 13 iPad minis available for lending to participants at each site. To facilitate data collection, our team set up two tables at each museum. One table was positioned at the entrance of the exhibit, where visitors were invited to take part in the study and complete a pre-questionnaire. The other table was located at the exit of the exhibit, where visitors filled out a post-questionnaire. At the entrance table, our researchers approached every visitor group that included at least one adult over the age of 18 and invited them to participate, following a standardized script. If a group declined to participate, we documented the general composition of the group (number of adults, youth, and babies/toddlers) and made note of any relevant comments they expressed. For groups that accepted, one adult from each group completed the pre-questionnaire. Once the necessary paperwork was completed, participants were given an iPad with a unique number, which was recorded on their questionnaire along with the entry time to the exhibit. Participants were informed that our research team had developed an iPad app designed to address their questions as they explored the exhibit. The adult participant who filled out the pre-survey was designated as the "question enterer" for the group, while other members were encouraged to contribute their own questions to the designated person. Participants were informed that they were free to spend as much time as they desired in the exhibit. After completing the exhibit, participants returned the iPad to the exit table. Our researchers recorded the exit time on the post-questionnaire before handing it to the adult participant for completion. As a token of appreciation for their participation, participant groups were given stickers and temporary tattoos.

Data and measures

In this study, the use of the app doubled as an unobtrusive data-gatherer for researchers to not only collect vast amounts of authentic questions asked by visitors but also gather information about the app's performance in responding to visitors' questions. Specifically, we examined three strands of data.

The first strand was the app-collected question-asking data. The second strand was survey data from pre-post questionnaires, which were reliable measures reported in our previous paper (Nelson et al., 2020). The third strand was researcher-coded data. For the third strand, we scored each question collected from the app based on our coding scheme. To address our research questions, we identified two dependent variables (DVs) and five independent variables (IVs) drawn from the literature and the available data. The definition and explanations of all the variables can be found below.

Question-asking quality and performance (DVs)

To assess visitors' question-asking quality and performance, we calculated both the average and the total question score per visitor group. Since the app collected all the questions asked by each visitor group, we carefully assigned them to the corresponding question type and scored each question based on our coding scheme (See Table 2). Because some groups had more members than others which might contribute to a higher number of questions asked within the group, we decided to use the average question score per visitor group instead of the total question score as an indicator of question-asking quality (DV1). However, we later used the total question score per visitor group as the indicator for the overall question-asking performance (DV2), which took both the question quality and quantity into consideration.

Cognitive factor

Epistemic curiosity is a composite cognitive factor created from a series of pre-questionnaire questions asking visitors about their frequency of participation in free-time science activities. Participants were asked "In your free time, how often do you (1) read news about science? (2) read books about science? (3) visit websites about science? (4) watch TV about science? (5) talk to others about science? (6) visit science-related museums? (7) do science activities/projects?" and the choices were "Never/almost never", "Once a year", "A few times a year", "Monthly", "Weekly", "Daily". We scored the choices from 1 ("Never/almost never") to 6 ("Daily") and calculated an average composite score of these questions as a proxy measure for epistemic curiosity per group. This cognitive factor (CF_1) was a continuous variable in our analysis.

Instrumental factors

The instrumental factors were sourced directly from the app. App mode was a categorical variable (IF_IV2): (0) Ask Mode, (1) Game Mode. Since the app used a tiered response system to report how certain it was in responding to each question. For the analysis, we estimated the average tier score for each visitor group as the measure of app performance. This app performance variable was a continuous one (IF_IV3).

Socio-contextual factors

Participants were from two different museums with different layouts, so museum layout was a categorical variable (SCF_IV4): (0) Museum A (linear exhibit), (1) Museum B (non-linear exhibit). In addition, the socio-contextual factors considered survey data such as self-reported group question-asking frequency collected from the questionnaire. Participants were asked to

Table 2 Question coding scheme

Question type		Explanation	Examples
Question category	Question subcategory		
Basic Information Questions Scored as "1"	Factual	Question seeks one answer, a very limited range of correct answers, or a negative or affirmative answer	How heavy is the Tucson meteorite?
			Where is cuprite found?
			Who found earth?
	Declarative		What is a stalactite?
			What is earth made of?
			Tell/Teach/Show me about… (an actual question)
	Binary question		Is there wind in space?
			Is there fossil on other planets?
			Does the earth rotate in different directions?
Regular Wonderment Questions Scored as "2"	Process	Open question needs more elaboration, or generates a range of answers	How did the earth begin?
			How does selenite form?
			What makes fog?
	Explanatory		What happens when a rock is put under pressure? Why is Quartz purple?
			What was Earth like 30 million years ago?
	Causal		What causes/makes crystal caves?

Table 2 (continued)

Question type		Explanation	Examples
Question category	Question subcategory		
High-level Wonderment Questions Scored as "3"	Comparison and contrast	Question involves divergent and critical thinking to predict, infer, hypothesize, and generate new ideas or information in addition to the exhibit contents	What is the difference between the stalactite and stalagmite?
			What elements are found in both humans and rocks of the earth?
	Critical/Hypothetical		What is the purpose in finding marked locations on the opposite wall? Are you trying to promote non-linear thinking and problem solving?
			What would the seasons be like if the earth was not tilted on its axis?
	Interpretive		What is the purpose of a meteorite?
	Communal		What did people think the human being looked like?
			In life on earth 30 million years ago how do we know what the animals really looked like?
			Why do we not feel the earth move?
Nonsensical or Off-topic Questions Scored as "0"	Personal info	Question that is not about exhibit contents	What is your name?
	Irrelevant information		Are there any pokemon around here?
	Repetitive questioning		Where do fossi? (code this as "0") Where do fossil fuels come from? (code this as "1")
	Random strings		uu?
	Incomplete		earth?

rate "how often did your group discuss what questions to ask Dr. D?" on a 1 to 5 Likert scale (1 = not at all; 5 = every time). This variable was included in our analysis as a continuous variable (SCF_IV5).

Coding procedure and scheme

The coding procedure involved two main steps. First, one of the researchers closely reviewed 3000 (25%) questions and proposed an initial categorization of the different types of questions based mainly on Scardamalia and Bereiter's coding rubric (Scardamalia & Bereiter, 1992). Given the diversity of questions captured in our app, we formulated four main types and fifteen subtypes of questions after multiple rounds of team discussion and review. As a result, we distinguished wonderment questions into two types: (a) regular wonderment questions that asked for explanations of science processes or causality for a scientific phenomenon; (b) high-level wonderment questions that asked for comparison and contrast, involving inquiries that manifest critical thinking, scientific hypothetical reasoning, or questions that may generate interpretive or communal discussions. The finalized coding scheme is shown in Table 2. Upon reaching a consensus on the coding, the second step was to code the questions. The first and the third author independently coded 2400 (20%) questions that were evenly and randomly selected from all conditions. During the coding process, app modes and museum sites were blinded. Disagreements were discussed and addressed during team meetings. The overall interrater reliability was very high (Cohen's $k = .90$, $p < .001$). The rest of the questions were then coded by the first author.

Data analysis

For answering RQ1, descriptive statistics were used to present the results of visitors' overall question-asking quality. For addressing RQ2, a two (Ask Mode vs. Game Mode) by two (Linear vs. Non-linear) between subjects ANOVA was conducted to study question-asking quality (DV1) differences between visitors who used the Ask Mode of the app and the Game Mode of the app at two museums featuring different exhibit layouts. A series of pairwise comparisons were also conducted to examine the significance of differences between each pair of conditions. For RQ3, to identify which factors could influence museum visitors' question-asking performance and to what degree the influences are, we regressed visitors' total question score per group (DV2) on different App Mode x Museum Layout (IF_IV2 x SCF_IV4) conditions, epistemic curiosity (CF_IV1), app performance (IF_IV3), and group inquiry frequency (SCF_IV5), while controlling for group size. Considering the 2×2 interaction effect between the categorical variables, we converted the four conditional categories into three dummy variables: Game_Linear, Ask_Non-linear, and Game_Non-linear with Ask_Linear as the reference group. Epistemic curiosity is indicated by the frequency level of visitors' free-time science activities. App performance is the average tier score for the app answering visitor groups' questions. It is noted that the lower the average tier score, the better the app performance. Group inquiry frequency is the frequency level of visitors discussing what questions to ask the app within their group.

Results

Overall question-asking quality and performance (RQ1)

The app used in this study collected a total of 11,909 questions from museum visitors across two sites. Among these questions, 8126 (68%) were basic information questions and were scored as "1" for each question, 2144 (18%) were regular wonderment questions and were scored as "2", and only 82 (1%) were high-level wonderment questions and were scored as "3". There were 1557 (13%) questions that were nonsensical or off-topic and were scored as "0" according to our coding scheme. Table 3 shows the descriptive statistics for the dependent variable, question-asking quality (DV1), measured by the average question score per visitor group. Table 3 also provides information on another dependent variable, question-asking performance (DV2), measured by the total question score per visitor group without averaging it by the number of questions asked.

The results of the analysis on overall question-asking quality (DV1) and question-asking performance (DV2) revealed interesting patterns across different group sizes, age groups, gender, and race. When examining DV1, it did not show much difference across different group sizes. However, it was found that the most commonly seen small-sized groups with 2 to 4 individuals exhibit a higher average of DV2 (M = 9.25, SD = 7.16) than larger groups containing 9 or 10 people (M = 8.17, SD = 7.35). Similarly, when analyzing age groups, different age groups did not demonstrate many differences in DV1 while age groups (36–55 years old) demonstrated higher average of DV2 (M = 10.24, SD = 7.72) compared to younger age groups (18–35 years old, M = 9.26, SD = 7.34) and older age groups (elder than 56 yrs old, M = 7.44, SD = 6.55). Moreover, the analysis indicated that females (M = 1.08, SD = 0.37) tended to exhibit greater question-asking quality (DV1) than males (M = 1.03, SD = 0.37). This advantage was mitigated when examining DV2. Additionally, when considering race and ethnicity, participants from diverse racial backgrounds displayed varying levels of question-asking quality (DV1). Native Hawaiian (M = 1.15, SD = 0.14, N = 7) was the highest, followed by White (M = 1.08, SD = 0.36, N = 942), Hispanic (M = 1.06, SD = 0.38, N = 220), Black (M = 0.98, SD = 0.39, N = 46, Native American (M = 0.88, SD = 0.30, N = 21), and Asian (M = 0.88, SD = 0.40, N = 60). In terms of question-asking performance (DV2), the findings indicated not much difference across ethnic groups. Overall, these results shed light on the nuanced dynamics of the overall question-asking qualities and

Table 3 Descriptive statistics for the dependent variables

	N	Min	Max	Mean	SD	Percentiles		
						25%	50%	75%
Average Question Score Per Group (DV1)	1331	0	3	1.06	0.37	0.94	1	1.25
Total Question Score Per Group (DV2)	1331	0	50	9.51	7.38	4	8	14

performances within distinct group sizes, ages, gender, and racial groups, providing valuable insights for further research and development in customizing question-asking apps based on different demographic characteristics.

Effects of the mobile app on question-asking quality (RQ2)

In addition to analyzing the overall average results of all participants, we were interested in further examining question-asking quality under different conditions. There was a significant main effect of App Mode, F (1, 1327) = 6.37, p = .012. Visitors who used the Ask Mode (M = 1.03, SD = 0.43) had a significantly lower question quality than those who used the Game Mode (M = 1.08, SD = 0.33). The main effect for museum layout was also significant, F (1, 1327) = 32.41, p < .001, such that visitors in the linear non-interactive exhibit (M = 0.99, SD = 0.38) had significantly lower question quality than those in the non-linear interactive exhibit (M = 1.10, SD = 0.36). However, the interaction effect was also significant, F (1, 1327) = 6.84, p = .009, indicating the effect of different versions of the mobile app (Ask Mode vs. Game Mode) on visitors' question-asking quality is dependent on museum layout (Linear non-interactive vs. Non-linear interactive).

The observed mean, standard deviation, median, and range of average question scores per visitor group across the four conditions are displayed in Table 4. In particular, using the Game Mode of the app at a linear non-interactive exhibit (GAME_MUSEUM A) led to a significantly higher question score than using the Ask Mode of the app at the same museum (ASK_MUSEUM A), p < 0.001. However, there was no significant difference found between the other two conditions using different versions of the mobile app at a non-linear, interactive exhibit (ASK_MUSEUM B = GAME_MUSEUM B).

As shown in Fig. 3, although the question-asking quality at Museum B is generally higher than that at Museum A, visitor groups who used the Game Mode of the app at Museum A had significantly higher question-asking quality (measured by the higher average question score per group) than those who used the Ask Mode at Museum A. This pattern did not appear at Museum B given the relatively flat line shown in Fig. 3. This difference suggests that the effect of different modes of the mobile app on visitors' question-asking quality varies by the different museum layouts in which visitors are engaged.

Overall, this set of analyses successfully provided evidence for the two hypotheses generated from RQ2: (1) H1: Visitors who used the Game Mode of the Dr. D app possess significantly higher question-asking quality than those who used the Ask Mode of the app at the linear non-interactive Museum A. In other words, the effect of different modes of the app on visitors' question-asking quality varied substantially when visitors engaged in a linear, non-interactive exhibit. (2) H2: Visitors who used the Game Mode of the Dr. D app demonstrated similar question-asking quality to those who used the Ask Mode of the

Table 4 Question asking quality by conditions

	ASK_MUSEUM A	GAME_MUSEUM A	ASK_MUSEUM B	GAME_MUSEUM B
N	214	249	296	572
M (SD)	0.93 (0.40)	1.04 (0.36)	1.11 (0.43)	1.10 (0.32)
Median [min, max]	1 [0, 2]	1 [0, 2]	1 [0, 3]	1.1 [0, 3]

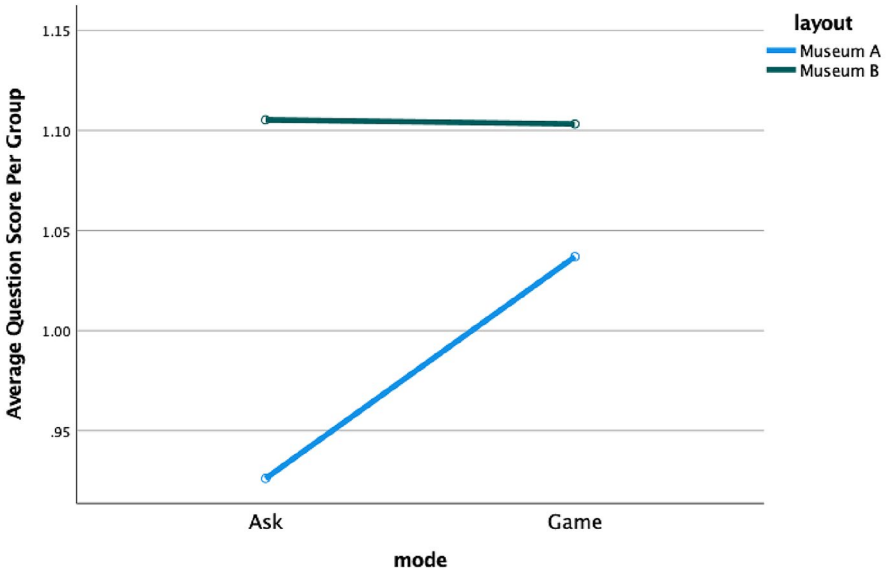

Fig. 3 Average question score by condition

app at the non-linear interactive Museum B. In other words, the effect of different modes of the app on visitors' question-asking quality did not vary when visitors engaged in the non-linear interactive exhibit. We will elaborate on this result in the Discussion section.

Influential factors of museum visitors' question-asking performance (RQ3)

Overall, all the variables combined explained 26.4% of the variance in visitors' question-asking performance, and the overall regression equation was statistically significant (F [7, 1242] = 63.791, $p < 0.001$). However, Table 5 shows that not all the variables are important in this regression. The cognitive factor, epistemic curiosity, was not statistically significant (ß = 0.041, $p = .096$) in explaining visitors' question-asking performance. In contrast, the instrumental factor, app performance, had a substantial and statistically significant effect

Table 5 Regression analysis for question-asking performance

Variable	Standardized coefficients		
	ß	t	Sig
Constant		2.79	.005
Group Size	0.08	3.40	< .001
Game_Linear	0.18	5.69	< .001
Ask_Non-linear	0.20	5.95	< .001
Game_Non-linear	0.55	15.45	< .001
Curiosity	0.04	1.66	.096
App performance	− 0.13	− 5.29	< .001
Group inquiry frequency	0.18	7.22	< .001

on question-asking performance (ß = − 0.131, $p < .001$), as did the socio-contextual factor, group inquiry frequency (ß = 0.180, $p < .001$). Moreover, the three dummy variables also showed statistically significant effects on question-asking performance at varying degrees.

To be specific, being in the Game_Linear condition led to a 0.18 SD increase in question-asking performance; being in the Ask_Non-linear condition led to a 0.20 SD increase; being in the Game_Non-linear condition led to a 0.55 SD increase in question-asking performance, with group size controlled. The regression results inform that although the literature suggests consideration of cognitive factors (i.e., epistemic curiosity) as an influential factor for question-asking performance, our model shows that the effect of epistemic curiosity on question-asking performance was not statistically significant for museum visitors. This suggests that our app might be an effective intervention tool that promotes visitors' question-asking performance at museums regardless of the frequency level of visitors' free-time science activities (a proxy measure for epistemic curiosity). Therefore, the regression analysis successfully provided evidence for H3 derived from RQ3: visitors' question-asking performance is largely influenced by the interaction between the app mode (IF_1) and the museum layout (SCF_1), the app performance in answering questions (IF_2), and visitor groups' question-asking frequency (SCF_2).

Discussion

The ultimate goal of the study was to foster high-level questioning among visitors and improve their museum learning experience by understanding visitors' question-asking quality and the factors associated with question-asking performance. To achieve this goal, we addressed three research questions related to (a) the overall question-asking quality and performance among visitor groups with distinct demographic characteristics using the Dr. D app; (2) the effect of different modes of the Dr. D app (Ask Mode vs. Game Mode) on visitors' question-asking quality at two different museums (Linear Museum A vs. Non-linear Museum B); (3) how are museum visitors' question-asking performance explained by cognitive, instrumental, and socio-contextual factors?

For RQ1, we found that the overall question-asking quality was low for the full sample. Based on our question coding scheme, most questions asked using the Dr. D app were basic information questions (68%). Wonderment questions took up 19% with only 1% for high-level wonderment questions which involved divergent and critical thinking to predict, infer, hypothesize, and generate new ideas or information in addition to the exhibit contents. The finding of an overall low question quality highlights the significance of the study in examining how different versions of a mobile app might affect visitors' question-asking quality. The low question-asking quality could be attributed to several reasons. For instance, the prevalence of basic information questions suggests that participants may have lacked deeper engagement or curiosity about the exhibit content. They may have been more focused on seeking factual details rather than exploring the concepts or ideas presented in the exhibits. The scarcity of high-level wonderment questions indicates a limited use of critical and divergent thinking skills among the participants. This could be due to various factors, such as a lack of prior knowledge or experience, insufficient motivation or interest in engaging with the exhibits, or a lack of guidance or prompts within the app to encourage higher-level questioning. The results of low overall question-asking quality underscore the importance of developing customized mobile apps or scaffolding tools to facilitate higher

question-asking quality and performance among museum visitors with distinct demographic backgrounds.

For RQ2, we investigated visitors' question-asking quality using the two modes of the Dr. D app across two different museum sites. We found that visitors' question-asking quality did not have a substantial difference depending on the app mode they used in a non-linear interactive exhibit. However, at a linear, non-interactive exhibit, visitors who used Game Mode scored significantly higher in their question-asking quality than those who used the Ask Mode. This finding suggests that using the gamified badging reward system is not a one-size-fits-all strategy to promote high question-asking quality and it often requires careful consideration of the learning contexts (Dichev & Dicheva, 2017; Sailer & Homner, 2020). To be specific, when learners are exposed to environments that are less interactive in terms of stimulating curiosity and interests, a gamified mobile app might help learners to increase their engagement by fostering extrinsic motivation. However, when the environment itself already provides learners with an engaging and interactive experience and expects high intrinsic motivation from the learners, the role of a gamified mobile app might be limited. Given the fact that changing the "hardscape" (physical layout) of museum exhibits is often costly and unrealistic, understanding how to design flexible elements (i.e., adding or not adding a gamified element to a mobile app) around the physical settings of the museum becomes critical.

For RQ3, as informed by the literature, we investigated whether the cognitive factor, instrumental factors, and socio-contextual factors could impact museum visitors' question-asking performance. The regression results suggest that these factors indeed have different weights in affecting visitors' question-asking performance. The cognitive factor represented by visitors' epistemic curiosity had no statistically significant effect on visitors' question-asking performance. In contrast, instrumental factors such as the app's performance in answering questions, and socio-contextual factors such as group inquiry frequency had substantial influences on question-asking performance. Visitor groups that received more assured responses to their questions, as well as those that had a higher frequency level to discuss what questions to ask the app, would normally have better question-asking performance. More importantly, the interaction between the instrumental factor of *App mode* and the socio-contextual factor of *Museum layout* contributed a large proportion of the variance in explaining visitor groups' question-asking performance. Holding other variables constant, visitors in the Game_Non-linear condition had the highest question-asking quality, followed by Ask_Non-linear, Game_Linear, and Ask_Linear. This finding advances the existing research in two directions.

First, recent efforts on designing mobile apps (Paliokas et al., 2020; Zimmerman & Land, 2022) for museums and other informal learning environments have not yet comprehensively considered both the socio-contextual factors and instrumental factors together. Existing mobile apps for museums might have focused solely on instrumental factors, such as providing information about the exhibits or offering interactive quizzes (Paliokas et al., 2020). While these features are valuable, they may not have considered the socio-contextual factors (e.g., visitor's age group, educational background, prior knowledge, and social dynamics within the museum) that influence visitors' engagement and questioning behaviors (Zimmerman & Land, 2022). Given that this study proposed a question coding framework as well as considering multiple factors influencing question-asking quality, future researchers can draw on the framework as well as the multiple-factor perspective to evaluate whether mobile apps were designed effectively to support high-level questioning across different learning environments. To be more specific, future mobile apps should involve features like prompts for promoting social interactions and collaborative questioning,

access to additional resources or multimedia content, or mechanisms to provide adaptive feedback on the quality of the questions asked.

Second, our study indicates that the performance of an app in answering visitors' questions could largely explain visitors' question-asking performance. For instance, when the Dr. D. app was more successful in answering questions posed, subsequent questions asked tended to be more high-level. This suggests that as visitors gained confidence in the app's ability to answer questions, they felt more comfortable asking higher-level questions. Conversely, if the app could not answer questions well, visitors were less likely to follow up with more complex questions. Therefore, we suggest future efforts to further improve the question-answering or any other designed systems of mobile technologies since an app's performance and its feedback could profoundly influence learners' question-asking performance. Future research should consider providing more personalized feedback and matched responses to visitors' questions or requests, designing a more intelligent system that could support deeper inquiry and reasoning, and scaffolding visitors' self-directed learning and explorations at museums or other informal learning environments.

Limitations

Although our study has yielded meaningful insights, it is important to acknowledge its limitations. Specifically, one of the primary limitations stems from our reliance on self-reported survey data and the log data obtained through our question-asking mobile app without considering other data sources. In light of the advancements in wearable technologies, it would be beneficial for museum researchers to collect multi-channel data, such as eye movement, facial expressions, posture, blood pressure, heart rate, and body temperature. Additionally, incorporating observational data could offer valuable insights. While we aimed to gather data in a non-intrusive manner and refrained from using wearable devices or employing observers to track visitor groups, future research could consider transforming the question-asking coding scheme proposed in this study into an observational protocol. This approach could potentially guide museum researchers in conducting field observations or analyzing the question-asking behaviors of visitor groups as they engage with exhibits. Future studies could be improved by including more data sources to triangulate findings, especially to enhance the validity and reliability in measuring museum visitors' question-asking quality and performance. Second, we were not able to collect question-asking data in the control conditions where visitors explored the exhibits without using our mobile app. Caution should be taken when emphasizing the effect of a question-asking app in fostering visitor engagement in museums compared with regular museum visits with no app in hand. Future efforts may need to further investigate the potential effects of using a question-asking app versus not using it. Third, the research presented here is an early attempt at incorporating science and engineering practices (in the form of asking questions) into museum education. While focusing merely on question-asking may initially be considered a limitation for the generalizability of this work, we feel that our study offers up a new opportunity and opens new avenues for educational designers and researchers. We encourage future works to think deeply about the role of question-asking in one's learning by comprehensively examining the factors that could mediate question-asking quality and performance, and to collaborate on developing and implementing new technologies to engage learners of diverse backgrounds in more science and engineering practices (e.g.,

analyzing and interpreting data; constructing explanations; engaging in argument from evidence; obtaining, evaluating, and communicating information) that builds on their abilities to ask questions.

Conclusions

In conclusion, the findings of our study have both theoretical and practical implications. In terms of the theoretical implications, our study conducted a systematic review of question-asking mechanisms and examined a wide range of factors that could influence question-asking performance. This may point future researchers in a direction that comprehensively considers cognitive, instrumental, and socio-contextual factors when conducting research or making design decisions in informal settings. For the practical implications, we believe that the question coding scheme derived from our data analysis process can make a significant contribution to the practice of museum education and informal learning. We foresee that future researchers can directly apply our coding schema to analyze question-asking quality and associate it with different levels of cognitive engagement in informal learning spaces. We also anticipate future works to easily adapt our coding scheme into guiding protocols or principles that would potentially support field observations or the design of new mobile apps which focus on analyzing questions or developing an intelligent, automated question analyzing and responding system. Given how previous work has noted the difficulty of obtaining a large number of knowledge-based questions from students under ordinary formal school circumstances, we also believe that our investigation of question-asking quality and the relevant factors for museum visitors' question-asking performance may help shed light on question-asking research in different learning environments, formal or informal.

Appendix A

How the Tier is Defined in Our App's Response System.

Examples	App reaction	Tier (how certain the app responds to the Qs)
Q: What is stardust? Our app: Stardust was designed to catch comet dust and pre-solar interstellar materials and return them safely to Earth	Our app directly answers the question	1 (perfect match)
Q: How many stars in the sky? Our app: If you were asking "How many stars in the galaxy?" The answer is: Current estimates suggest that there are 200 billion stars in the Milky Way galaxy	Our app reinterprets the question, with an answer provided	2 (close match)

Examples	App reaction	Tier (how certain the app responds to the Qs)
Q: What makes a wave? Our app: Sorry, I don't understand. Did you mean to ask: "What is a wave?" "What makes a tornado?" Neither of these	Our app raises two seemingly relevant questions with hyperlinks to their answers	3 (partial match)
Q: Did the dinosaurs die from a meteor? Our app: Oops! I haven't learned that yet. But did you know: Volcanoes are built by the accumulation of their own eruptive products – lava, bombs (crusted over ash flows), and tephra (airborne ash and dust)	Our app replies that it has not yet learned that information and provides the user with a random science fact from its database	4 (no match)

Note that users could indicate that none of the alternative questions suggested by the app represent their originally posted questions by tapping "seems wrong" when viewing the answer screen, which allowed the research team to keep improving the Q&A database.

Acknowledgements This work was supported by the National Science Foundation [Grant Number 1438825].

Data availability The data that supports the findings of this study is available at Su, M. (2022).

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Research involving human participants and/or animals The study was approved by the Human Subjects Institutional Review Board (IRB) at Arizona State University.

Informed consent Informed consent was obtained from all participants in this study.

References

Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Sauzéon, H., & Oudeyer, P. Y. (2022). GPT-3-driven pedagogical agents for training children's curious question-asking skills. *arXiv preprint* arXiv:2211.14228. https://doi.org/10.48550/arXiv.2211.14228

Alaimi, M., Law, E., Pantasdo, K. D., Oudeyer, P. Y., & Sauzeon, H. (2020). Pedagogical agents for fostering question-asking skills in children. *Conference on Human Factors in Computing Systems Proceedings*. https://doi.org/10.1145/3313831.3376776

Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education, 88*(S1), S17–S33. https://doi.org/10.1002/sce.20016

Allen, S., & Gutwill, J. P. (2009). Creating a program to deepen family inquiry at interactive science exhibits. *Curator: The Museum Journal, 52*(3), 289–306. https://doi.org/10.1111/j.2151-6952.2009.tb00352.x

Ash, D. (2004). How families use questions at dioramas: Ideas for exhibit design. *Curator: The Museum Journal, 47*(1), 84–100. https://doi.org/10.1111/j.2151-6952.2004.tb00367.x

Borun. (2002). Object-based learning and family groups. *Perspectives on Object-Centered Learning in Museums*, 245–260.

Bradley, M. E., Thom, L. R., Hayes, J., & Hay, C. (2008). Ask and you will receive: How question type influences quantity and quality of online discussions. *British Journal of Educational Technology, 39*(5), 888–900. https://doi.org/10.1111/j.1467-8535.2007.00804.x

Braund, M., & Lelliott, A. (2017). Opening up the dialogic space. Using questions to facilitate deeper informal learning. *Preparing Informal Science Educators, 14*, 15. https://doi.org/10.1007/978-3-319-50398-1_28

Bybee, R. W. (2006). Scientific inquiry and science teaching. In L. B. Flick & N. G. Lederman (Eds.), *Scientific Inquiry and Nature of Science. Science & Technology Education Library.* (Vol. 25). Springer.

Callanan, M. A., Castañeda, C. L., Solis, G., Luce, M. R., Diep, M., McHugh, S. R., Martin, J. L., Scotchmoor, J., & DeAngelis, S. (2021). "He fell in and that's how he became a fossil!": Engagement with a storytelling exhibit predicts families' explanatory science talk during a museum visit. *Frontiers in Psychology, 12*, 689649. https://doi.org/10.3389/fpsyg.2021.689649

Cano, F., García, Á., Berbén, A. B. G., & Justicia, F. (2014). Science Learning: A path analysis of its links with reading comprehension, question-asking in class and science achievement. *International Journal of Science Education, 36*(10), 1710–1732. https://doi.org/10.1080/09500693.2013.876678

Charitonos, K., Blake, C., Scanlon, E., & Jones, A. (2012). Museum learning via social and mobile technologies: (How) can online interactions enhance the visitor experience? *British Journal of Educational Technology, 43*(5), 802–819. https://doi.org/10.1111/j.1467-8535.2012.01360.x

Chi, M. T. H. (2021). Translating a theory of active learning: An attempt to close the research-practice gap in education. *Topics in Cognitive Science, 13*(3), 441–463. https://doi.org/10.1111/tops.12539

Cordova-Rangel, J., & Caro, K. (2023). Designing and evaluating aventura marina: A serious game to promote visitors' engagement in a science museum exhibition. *Interacting with Computers*. https://doi.org/10.1093/iwc/iwad017

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education, 14*(1), 1–36. https://doi.org/10.1186/s41239-017-0042-5

Efstathiou, I., Kyza, E. A., & Georgiou, Y. (2018). An inquiry-based augmented reality mobile learning approach to fostering primary school students' historical reasoning in non-formal settings. *Interactive Learning Environments, 26*(1), 22–41. https://doi.org/10.1080/10494820.2016.1276076

Gallagher, J. J., & Aschner, M. J. (1963). A preliminary report on analyses of classroom interaction. *Merrill-Palmer Quarterly of Behavior and Development, 9*(3), 183–194.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137. https://doi.org/10.3102/000283120310011

Gutwill, J. P., & Allen, S. (2012). Deepening students' scientific inquiry skills during a science museum field trip. *Journal of the Learning Sciences, 21*(1), 130–181. https://doi.org/10.1080/10508406.2011.555938

Ha, J., Pérez Cortés, L. E., Su, M. Nelson, B. C., Bowman, C., & Bowman, J. D. (2021). The impact of a gamified mobile question-asking app on museum visitor group interactions: An ICAP framing. *The International Journal of Computer-Supported Collaborative Learning*, *16*, 367–401.

Harley, J. M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016). Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes. *Educational Technology Research and Development, 64*, 359–388. https://doi.org/10.1007/s11423-015-9420-7

Hohenstein, J., & Tran, L. U. (2007). Use of questions in exhibit labels to generate explanatory conversation among science museum visitors. *International Journal of Science Education, 29*(12), 1557–1580. https://doi.org/10.1080/09500690701494068

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology, 113*(3), 430–452. https://doi.org/10.1037/pspi0000097

Humphries, J., & Ness, M. (2015). Beyond who, what, where, when, why, and how: Preparing students to generate questions in the age of common core standards. *Journal of Research in Childhood Education, 29*(4), 551–564. https://doi.org/10.1080/02568543.2015.1073199

Jirout, J. & Klahr, D. (2011). Children's Question Asking and Curiosity: A Training Study. *Society for Research on Educational Effectiveness*. https://eric.ed.gov/?id=ED528504

Jong, M.S.-Y., Chan, T., Hue, M.-T., & Tam, V. W. L. (2018). Gamifying and mobilising social enquiry-based learning in authentic outdoor environments. *Journal of Educational Technology & Society, 21*(4), 277–292.

Jovanovic, J., & Devedzic, V. (2015). Open badges: Novel means to motivate, scaffold and recognize learning. *Technology, Knowledge and Learning, 20*, 115–122. https://doi.org/10.1007/s10758-014-9232-6

Lee, H., Jung, T. H., tom Dieck, M. C., & Chung, N. (2020). Experiencing immersive virtual reality in museums. *Information & Management, 57*(5), 103229. https://doi.org/10.1016/j.im.2019.103229

Liu, C., Zowghi, D., Kearney, M., & Bano, M. (2021). Inquiry-based mobile learning in secondary school science education: A systematic review. *Journal of Computer Assisted Learning, 37*(1), 1–23. https://doi.org/10.1111/jcal.12505

Negruşa, A. L., Toader, V., Sofică, A., Tutunea, M. F., & Rus, R. V. (2015). Exploring gamification techniques and applications for sustainable tourism. *Sustainability, 7*(8), 11160–11189. https://doi.org/10.3390/su70811160

Nelson, B. C., Bowman, C. & Bowman, J. (2017). Designing for data with ask Dr. Discovery: Design approaches for facilitating museum evaluation with real-time data mining. *Technology, Knowledge and Learning, 22*, 427–442. https://doi.org/10.1007/s10758-017-9313-4

Nelson, B. C., Bowman, C. D. D., Bowman, J. D., Pérez Cortés, L. E., Adkins, A., Escalante, E., Owen, B. L., Ha, J., & Su, M. (2020). Ask Dr. Discovery: The impact of a casual mobile game on visitor engagement with science museum content. *Educational Technology Research and Development*, *68*, 345–362. https://doi.org/10.1007/s11423-019-09696-x

Paliokas, I., Patenidis, A. T., Mitsopoulou, E. E., Tsita, C., Pehlivanides, G., Karyati, E., Tsafaras, S., et al. (2020). A gamified augmented reality application for digital heritage and tourism. *Applied Sciences, 10*(21), 7868. https://doi.org/10.3390/app10217868

Pérez Cortés, L. E., Ha, J., Su, M., Nelson, B., & Bowman, C. (2023). Gleaning museum visitors' behaviors by analyzing questions asked in a mobile app. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-023-10208-1

Ram, A. (1991). A theory of questions and question asking. *Journal of the Learning Sciences, 1*(3–4), 273–318. https://doi.org/10.1080/10508406.1991.9671973

Ronfard, S., Zambrana, I. M., Hermansen, T. K., & Kelemen, D. (2018). Question-asking in childhood: A review of the literature and a framework for understanding its development. *Developmental Review, 49*(March), 101–120. https://doi.org/10.1016/j.dr.2018.05.002

Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction., 9*(3), 177–199. https://doi.org/10.1207/s1532690xci0903_1

Squire, K. D., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1), 5–29. https://doi.org/10.1007/s10956-006-9037-z

Su, M. (2022). Museum question-asking data file belonging to the research: Dr. Discovery question-asking mobile app in museums. 4TU.ResearchData. https://doi.org/10.4121/21085036.v1

Suarez, A., Specht, M., Prinsen, F., Kalz, M., & Ternier, S. (2018). A review of the types of mobile activities in mobile inquiry-based learning. *Computers & Education, 118*, 38–55. https://doi.org/10.1016/j.compedu.2017.11.004

Tigert, J. M., Fotouhi, G., & Kirschbaum, S. (2021). An investigation of museum educators' questioning during field trips. *Learning, Culture and Social Interaction, 31*, 100571. https://doi.org/10.1016/j.lcsi.2021.100571

Tissenbaum, M. (2020). I see what you did there! Divergent collaboration and learner transitions from unproductive to productive states in open-ended inquiry. *Computers & Education, 145*, 103739. https://doi.org/10.1016/j.compedu.2019.103739

Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: Understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning, 12*(1), 35–64. https://doi.org/10.1007/s11412-017-9249-7

Vartiainen, H., & Enkenberg, J. (2013). Learning from and with museum objects: Design perspectives, environment, and emerging learning systems. *Educational Technology Research and Development, 61*, 841–862. https://doi.org/10.1007/s11423-013-9311-8

Wang, H. Y., Liu, G. Z., & Hwang, G. J. (2017). Integrating socio-cultural contexts and location-based systems for ubiquitous language learning in museums: A state of the art review of 2009–2014. *British Journal of Educational Technology, 48*(2), 653–671. https://doi.org/10.1111/bjet.12424

Yoon, S. A., Elinich, K., Wang, J., Van Schooneveld, J. B., & Anderson, E. (2013). Scaffolding informal learning in science museums: How much is too much? *Science Education, 97*(6), 848–877. https://doi.org/10.1002/sce.21079

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review, 30*, 100326. https://doi.org/10.1016/j.edurev.2020.100326

Zimmerman, H. T., & Land, S. M. (2022). Supporting children's place-based observations and explanations using collaboration scripts while learning-on-the-move outdoors. *International Journal of Computer-Supported Collaborative Learning, 17*(1), 107–134. https://doi.org/10.1007/s11412-022-09366-w

Man Su earned her Ph.D. from the Mary Lou Fulton Teachers College, Arizona State University. Her research focuses on designing, implementing, and evaluating computer-supported learning technologies in both formal and informal environments. She is a postdoc research scientist at Saarland University Informatics Campus in Saarbrucken, Germany.

Jesse Ha is an assistant professor of Computer Science Education at Montclair State University. He is currently advising research for UCLA's STEM+C3 grant that focuses on preparing pre-service STEM teachers to integrate computational thinking and equity-centered practices into their pedagogy. His research interests include education technology, STEM/CS education, cognitive sciences, and asset-based pedagogies.

Luis E. Pérez Cortés earned his Ph.D. from the Mary Lou Fulton Teachers College, Arizona State University. He is an assistant professor at the University of Texas at Arlington.

Jeremy Bernier is a Ph.D. candidate at Mary Lou Fulton Teachers College, Arizona State University.

Lin Yan is a Ph.D. student at Mary Lou Fulton Teachers College, Arizona State University.

Brian C. Nelson is a professor of educational technology at Mary Lou Fulton Teachers College, Arizona State University. His research focuses on the theory, design, and implementation of immersive learning and assessment environments.

Judd D. Bowman is a professor in the School of Earth and Space Exploration at Arizona State University. He is an experimental cosmologist interested in the formation of structures in the early universe, including the first stars, galaxies, and black holes.

Catherine D. D. Bowman is an associate research professor in the School of Earth and Space Exploration at Arizona State University. Her research focuses on informal and higher education and evaluation.