Intern. J. Comput.-Support. Collab. Learn (2021) 16:367–401
https://doi.org/10.1007/s11412-021-09350-w

The impact of a gamified mobile question-asking app on museum visitor group interactions: an ICAP framing

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

Received: 25 October 2020 / Accepted: 12 July 2021 / Published online: 9 September 2021

Abstract

Mobile devices and apps have become a standard for the museum experience. Many studies have begun to explore the impact mobile apps may have on user experience and informal learning. However, there has been relatively little research on how visitor groups interact collaboratively while using these devices in computer-supported collaborative learning (CSCL) environments. In this paper, we explore the impact of a mobile question-asking app on museum visitor group interactions using the Interactive-Constructive-Active–Passive (ICAP) framework, a hierarchical taxonomy that differentiates modes of cognitive engagement. In a post-hoc analysis of survey findings from a study conducted at two large museums in the American southwest, we found that our app encouraged sharing of information among group members. In addition, users of a gamified version of the app were significantly more likely to report engaging in a group discussion during question-asking than groups using a non-game version of the app. We also found that group collaboration levels depended on the group-designated primary user of the app. Whenever a child or the group collaboratively asked the most questions, group discussion frequency was significantly higher. The study's findings support mobile question-asking apps' viability as a means to better understanding of museum visitor groups' interactions with exhibit content and provide evidence that game-based mobile apps, designed to foster question-asking by visitors, may bolster collaborative group interactions and informal learning.

Keywords Group interaction · Informal learning · Computer-supported collaborative learning · ICAP · Gamification · Mobile apps · Museum · Museum evaluation · Museum apps

✉ Jesse Ha
jesseha@asu.edu

[1] Mary Lou Fulton Teachers College, Arizona State University, Tempe, AZ, USA

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

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

Introduction

For decades, museums have attempted to bring more engagement and interest to their exhibits (Anderson, 2004). Museums have explored a range of design strategies, such as multiple options for visitor manipulation (Humphrey & Gutwill, 2005) and for simultaneous visitor manipulation (Gutwill, 2005). Recently, researchers and museums have attempted to increase opportunities for engagement with the design of digital apps that leverage the growing ubiquity of portable devices across the United States. The Pew Research Center (2019) found that 96% of Americans own a cellphone of some kind, of which 81% are smartphones. In addition, Bowman et al. (2019) found that 98% of museum visitors owned smartphones or tablets.

The push to modernize the museum experience by infusing it with some form of mobile technology is so prevalent that few museums today are without some form of a mobile-based app (Marty & Jones, 2012). These mobile-based apps often employ some combination of animation, music, and sound effects that are communicated through touch screens and delivered via interactive experiences such as games (Gilbert, 2016; Lohr, 2014). The use of mobile apps in museums reflects the current situation that "the issue is no longer whether to use media to enhance museum exhibitions, but how to use it." (Stogner, 2009, p. 385).

There are an increasing number of mobile apps developed for varied uses within museums such as the Brooklyn Museum's Ask, the Cleveland Museum's ArtLens, the Children's Museum of Houston's More, the Smithsonian's Infinity of Nations, the Museum of Natural History's Skin & Bones, and Roma Tre University's Inclusive Memory. Scholars and researchers have explored apps such as these (e.g., Bickmore et al., 2013; Katz et al., 2011; Poce et al., 2019; Proctor, 2015; Røtne & Kaptelinin, 2013; Valente-Marques, 2017). These cases represent a considered effort of museums and researchers to modernize the museum visitor's experience by leveraging widely available portable electronic devices.

Furthermore, museums have begun using digital mobile devices as learning tools (Parry & Ortiz-Williams, 2007; Parry & Sawyer, 2005). Several studies describe the use of mobile devices in such a capacity (e.g., Ahn et al., 2011; Chen et al., 2017; Jewitt, 2012; Manabe & Lydens, 2007; Roberts & Lyons, 2017). For example, Roberts and Lyons (2017) conducted an in situ experimental study that evaluated the differences in levels of productive learning talk in a 2×2 design, varying the method of user interaction with the exhibit (full-body digital interactive system vs. handheld Android tablet app) by the number of available input devices for interactions (single-user vs. multi-user) within a group of visitors. The study found that the handheld single-input condition produced the highest quality of collaborative learning talk, providing a basis for the current study's handheld single-input study design (Roberts & Lyons, 2017).

Group interactions, such as collaborative learning talk, are widely recognized as crucial for informal science learning in museums (Allen, 2002; Dierking & Falk, 1994). Accordingly, informal collaborative learning in museum experiences has been an ongoing discussion in the computer-supported collaborative learning (CSCL) community (Davis et al., 2015; Roberts & Lyons, 2017; Yoon et al., 2012). However, few studies have focused on informal learning and visitor group interactions with mobile device apps (Melero et al., 2015; Nelson et al., 2017; Roberts & Lyons, 2017).

The current study seeks to add to the body of CSCL literature through a post-hoc analysis of museum survey data collected using a mobile question-asking app, Dr. Discovery, developed for a National Science Foundation study that sought to address the need for scalable museum evaluation (Nelson et al., 2017). Furthermore, this current study seeks to frame informal museum learning via the Interactive-Constructive-Active -Passive (ICAP) framework (Chi, 2009; Chi & Wylie, 2014), a hierarchical taxonomy of cognitive engagement that predicts student deeper learning. This paper explores participant groups' interactions as reflected in app-recorded group question-asking patterns and self-reported collaborative activities, overall and by app mode (gamified question-asking Game Mode versus baseline question-asking Ask Mode), and analyzes the data to understand better if and how mobile apps in museums can support a crucial prerequisite for informal collaborative learning: visitor groups interacting and collaborating within their group around museum exhibit content using a mobile app.

While the original project (Nelson et al., 2017) did not directly investigate informal visitor learning, data was collected on precursors of collaborative learning, specifically self-reported data on the extent to which members of a visitor group interacted as well as app-logged data in the form of user-generated questions. Though this study cannot make definitive claims about visitor learning, we explored whether the app may provide the opportunity for learning by analyzing the design of our app using the ICAP framework.

Implications of this study include building an understanding of the viability of gamified mobile question-asking apps for better interpreting museum visitor groups' interactions and exploring the possibility for informal learning that accompanies using the app collaboratively. Moreover, this study contributes to the growing body of research into how people learn and interact in museums and informs the design of CSCL technologies for use in museums.

Theoretical framework and prior works

Frameworks for informal learning

Informal learning is any activity that involves the pursuit of knowledge without the presence of externally imposed curricula (Livingstone, 1999), including museum visits. Informal learning is dynamic, non-directed, exploratory, social, and user-driven (Honey & Hilton, 2011; McManus, 1994; Ramey-Gassert et al., 1994). Constructivists view informal learning as the active process of learning and discovery through constructing meaning from observation and sensory input (Hein, 1996; Russell, 1995). Museums incorporate an array of exhibits and experiences to foster informal learning, including static displays, demonstrations, hands-on exhibits, interactive exhibits, and multimedia interactives (Griffin, 1998).

Considering most visitors go to museums in groups, informal learning in museums generally happens in the context of social interactions such as family visitors or school field trips via questions and discussions generated by looking at exhibits and reading labels (Falk & Dierking, 1992). Scholars have discussed the potential benefits and impact of social interaction and collaboration in museums on student learning (Laursen, 2012; Yoon et al., 2012). To systematically measure informal learning and group inquiry in museums, scholars have developed frameworks to guide the design and assessment of informal learning experiences and group interaction in museums. Here we describe and explore some of the affordances and limitations of several frameworks that have been used for measuring informal learning in museums.

Barriault (1999) developed the Visitor Engagement Framework (VEF) to assess and code visitor informal learning experiences in connection with museum exhibits. The framework consists of seven discrete learning behaviors that are observable, including "initiation behaviors" such as 1) doing the activity or 2) watching others engaging in the activity; "transition behaviors" such as 3) repeating an activity or 4) exhibiting emotional responses; and "breakthrough behaviors" such as 5) referring to past experiences while engaging in the activity; 6) seeking and sharing information with others; and 7) engaging in inquisitive behavior such as testing variables, reading signage, remaining on task, making comparisons, and, applying knowledge (Barriault & Pearson, 2010, pp. 94–95). The major affordance of the VEF is the frequency measurement of the impact that exhibits have on visitor behavior. Analysis of this frequency data can reveal a general idea of how effective a particular exhibit is in engaging visitors. An ideal frequency profile would have visitors exhibiting more breakthrough behaviors. Using these frequency profiles, museums could iteratively redesign exhibits to improve VEF frequency profiles. Moreover, lessons from applying this framework can be applied to the future design of exhibits. A limitation of the VEF is its inability to rank the different types of breakthrough behaviors. For example, reading signage, remaining on task, and making comparisons are different cognitive activities that would have widely varying effects on learning.

Borun (2002) developed a hierarchical learning-levels schema for measuring family learning via coded visitor behaviors and conversations at three levels: "identifying," "describing," and "interpreting and applying" (Borun, 2002, p. 254). The first level, identifying, includes one-word statement/answers that involve few associations to exhibit content. Moreover, items coded at this level involve visitors making connections to content that "miss the point" of the exhibit (Borun, p. 254). The second level, describing, includes multi-word answers or statements that involve correct connections to visible exhibit content. In addition, items coded at this level involve visitors making connections to personal experiences based on visible exhibit characteristics but not the underlying concepts of the exhibit. The third level, interpreting/applying, includes multi-word answers or statements that involve correct connections to concepts underlying the exhibits. At this level, visitors make connections between the concepts of the exhibit and their life experiences. The learning-level schema can classify visitor behaviors that are relevant to learning as described above. An affordance of this schema is that the hierarchy of learning can be used to measure the quality of visitor learning behavior: identifying, describing, or interpreting and applying. For example, reading signage would classify as level two, describing, whereas making comparisons would be classified as level three, interpreting and applying. At the same time, a limitation of this schema lies in its inability to distinguish the quality of connections that visitors make between concepts and life experiences. For example, the third level would not distinguish between a visitor who simply states a concept that connects to their prior knowledge and a visitor who applies the concept in self-explaining prior knowledge.

Allen (2002) developed a coding framework that distinguishes five categories of museum visitor learning-talk: "perceptual," "conceptual," "connecting," "strategic," and "affective" (Allen, 2002, pp. 274–277). Perceptual talk refers to visitors drawing attention to a particular stimulus within an exhibit. Perceptual talk involves identification of something, stating the name of an object, pointing out concrete aspects of exhibits, or reading aloud a label in the exhibit. Conceptual talk refers to visitors making cognitive interpretations regarding elements of an exhibit, including simple interpretive statements, complex inferences such as hypothesizing, generalizing information, comparing and contrasting, and predictions or expectations of what will happen. Connecting talk refers to the connections between items in exhibits and visitor prior knowledge, including prior life experiences, prior declarative knowledge, and connections between exhibit elements. Strategic talk refers to discussion about how to use or approach exhibits as well as expressions of evaluation about visitor performance, actions, or abilities. Affective talk refers to expressions of emotions such as pleasure, displeasure, or intrigue. An affordance of Allen's framework is that it distinguishes four different conceptual subcodes: simple, complex, prediction, and metacognition. A limitation of the framework is that it does not directly address collaborative-learning talk within the coding scheme.

Davis et al. (2015) developed a coding framework for analyzing interaction and meaning-making, which revealed four distinct patterns of collaborative interaction among dyads in museums:"reactive," "autonomous," "planning," and "contemplative" (Davis et al., 2015, pp. 63–68). These patterns represent progressive levels of interactional complexity and vary both in the dimensions of the level of coordination between dyad members and the target of the action. The reactive pattern is characterized by dyad discussions with low coordination and a focus mainly on exploring an exhibit's mechanical goals. The autonomous pattern is characterized by dyad interactions with low coordination and with one dyad member focusing on a conceptual goal and the other dyad member focusing on mechanical goals or detaching from the dyad interaction entirely. The planning pattern is characterized by dyad interactions that include an active articulation of planned mechanical goals without much focus on conceptual goals. The contemplative pattern is characterized by dyads vocalizing planned complex and longer-term goals that focus on conceptual goals. The framework provides a taxonomy of four distinct collaborative discussion patterns that can be analyzed for how museum technology may support learning across these patterns of interactions. A major affordance of this framework is the inclusion of collaboration among visitors and the coding of the level of coordination between dyad members. A limitation of this framework is that it does not distinguish the quality of conceptual goals.

Tissenbaum et al. (2017) proposed the Divergent Collaborative Learning Mechanisms (DCLM) framework for coding the types of collaborative actions and talk in museum settings. The DCLM extended the Collaborative Learning Mechanisms (CLM) framework (Fleck et al., 2009) to encompass a broader spectrum of productive collaborative behaviors within open-ended, exploratory activities where participants can pursue divergent goals. In addition to the four learning mechanisms (i.e., "making and accepting suggestions"; "negotiating"; "joint attention and awareness"; "narrations") covered in the CLM framework, the DCLM framework expanded the mechanisms of the collaborative discussion by adding "clarification" and "seeking help" (Tissenbaum et al., 2017, p. 44). Making and accepting suggestions refers to groups of visitors introducing knowledge and ideas to one another; Negotiating refers to talk in which group members critically and constructively engage in negotiating ideas; Joint attention and awareness refers to the awareness of group members regarding jointly attending to the exhibit and ideas; Narrations refer to individual group members who narrate and say out loud what they are doing in order to enable monitoring within the group; Clarification refers to the discussion within the group that disambiguates actions of individual members that serve to coordinate group exploration; Seeking help refers to group members asking and learning from the expertise of more knowledgeable members within the group. Furthermore, the DCLM framework also revised the mechanisms for coordinating collaboration in CLM to embrace the mechanisms of enacting divergent collaboration. Four new mechanisms were included, namely "goal adaptation," "boundary spanning actions," "boundary spanning perception," and "modeling" (Tissenbaum et al., pp. 45–46). Goal adaptation refers to the moments in which group members change and adapt their goals and shift towards actively exploring exhibits;

Boundary spanning actions refers to group members actively tinkering and manipulating objects and spaces within the exhibit; Boundary spanning perception refers to multiple group members working synchronously on similar challenges in a manner that allows for simultaneous learning through observation of other group members; Modeling refers to more knowledgeable group members explaining what they are thinking and doing while manipulating exhibits in order to facilitate less knowledgeable group members in replicating actions. An affordance of this framework is the extensive categorization of collaborative learning actions and the variety of investigations into group visitations that allow researchers and museums to examine visitor interactions at a finer-grained level. A limitation of this framework is that there is no clear hierarchy for ranking the various codes.

Roberts and Lyons (2017) developed the scoring qualitative informal learning dialogue (SQuILD) method that addresses interactions within exhibits via five coding categories: "manage," "instantiate," "evaluate," "integrate," and "generate" (Roberts & Lyons, 2017, pp. 354–355). Manage refers to talk that is related to the establishment of joint attention, negotiation of action, or scaffolding exhibit use similar to that of Allen's (2002) strategic category. Instantiate refers to users reporting information aloud that may provide opportunity for other visitors to perceive that information from the exhibit. Evaluate refers to assigning value through statements that make a judgment or assessment regarding information from exhibits. Integrate refers to the connections among multiple pieces of information presented in an exhibit. Generate refers to statements that go beyond what is exhibited that combine information from the exhibit with visitors' prior knowledge. Though there is no discrete hierarchy among these five codes, Roberts and Lyons did rank subcodes into a hierarchy of low, mid, and high relevance. Generally, evaluate, generate, and integrate subcodes are classified as higher relevance than instantiate and manage subcodes. An affordance of this method is the inclusion of generation as a code considering generative learning promotes deeper learning (Chi, 2009). Other frameworks discussed above have categories and codes that could include generative learning strategies but do not explicitly categorize generative learning. In addition, a hierarchy of subcodes affords researchers a tool for interpreting differences in the quality of visitor dialogue. A limitation of this framework is that generative visitor actions are not necessarily exclusive to the generate coding category since certain evaluative and integrative actions can also be considered generative.

The ICAP framework for informal learning

We propose a framework that addresses some of the limitations of these aforementioned frameworks while at the same time embodies many of the affordances of those frameworks. In this section, we briefly summarize the ICAP framework and discuss the unique affordances of the framework for measuring informal learning as well as the specific reasons for its use in our study.

The ICAP hierarchical framework predicts that deeper learning is commensurate to the mode of engagement: interactive > constructive > active > passive, see Fig. 1 (Chi, 2009; Chi & Wylie, 2014). The passive mode is defined as simply paying attention and receiving information from instructors or instructional material without doing anything else. The active mode is defined as the passive mode in addition to manipulating anything related to learning but without the construction or generation of new ideas or information. The constructive mode is defined as the active mode in addition to the construction or generation of new ideas or mental models. The interactive mode is defined as the constructive

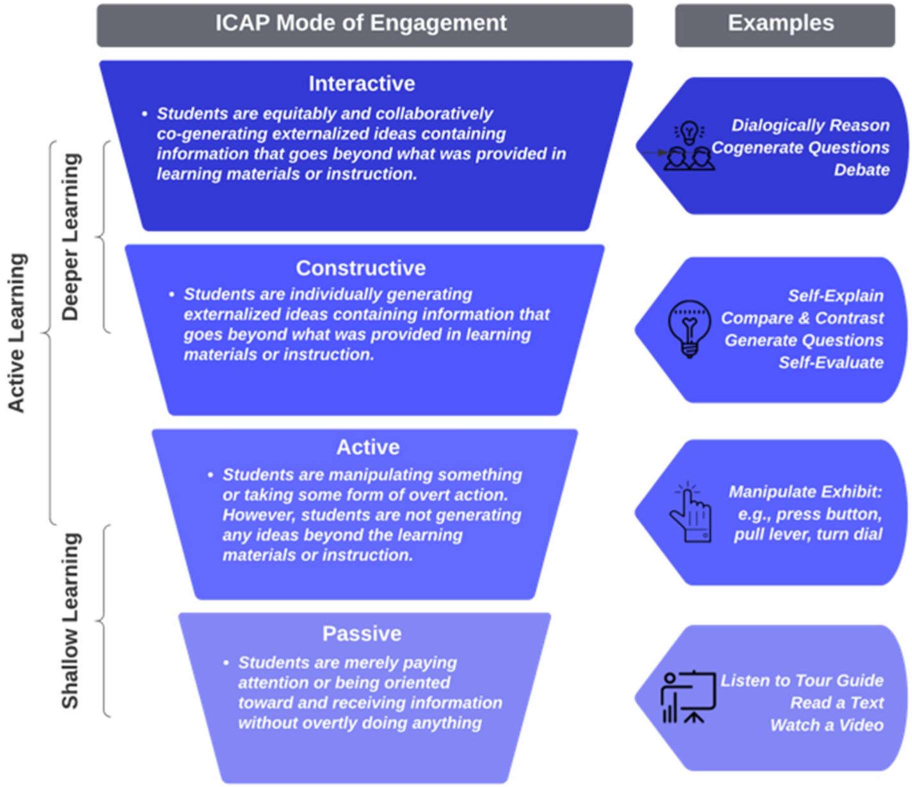


Fig. 1 The ICAP Framework

mode in addition to co-constructive or co-generative interactions among two or more peers. The interactive mode is often conflated with people interacting; the simple act of students interacting does not reach the interactive mode within the ICAP framework. The interactive mode requires people to co-generate ideas within groups in which all group members contribute equally and build on each other's ideas. Collaboration within the interactive category also requires full synchronous coordination of activities; groups that divide and conquer the work would be considered the constructive mode of engagement since group members are, in essence, learning in parallel.

Like many of the other frameworks already used in museum research, the ICAP framework provides similar affordances. One such element is a hierarchy for evaluating the quality of visitor behavior. Borun (2002) includes a hierarchy of learning at three levels (identifying, describing, and interpreting/applying) that are relevant to the ICAP framework. Generally, identifying and describing behavior would be classified within the active mode of engagement, and interpreting/applying behavior would be classified as either the active or constructive mode of engagement. The ICAP framework further refines Borun's categories by using generative learning strategies as a distinguishing feature for the constructive and interactive modes of engagement. Generative learning as a requirement for higher levels of the ICAP framework affords the evaluation of the quality of interpretation and application and their likelihood for fostering deeper learning. This focus on generative learning echoes the SQuILD method. In a similar fashion, the ICAP framework also provides a hierarchy that could disaggregate the VEF code of breakthrough behaviors by examining the quality of behaviors with the lens of generative learning. In addition, the ICAP framework relates to the conceptual portions of the frameworks by Allen (2002) and Davis et al. (2015). Conceptual talk, as described by these frameworks, would reach either the active or constructive levels of engagement. The ICAP framework would further evaluate the concepts discussed by visitors by evaluating whether visitors are generating new ideas as opposed to simply discussing the concepts in the same manner they are presented within exhibits. Another crucial element that the ICAP framework embodies is the element of collaboration via the interactive mode of engagement. The VEF framework categorizes visitors seeking and sharing information. The ICAP framework further refines this type of interaction by examining whether the information being shared is generative and not the repetition of text displayed in exhibits. This also applies to the collaborative interaction framework by Davis et al. and the DCLM framework. In addition, the ICAP framework echoes Davis et al. in terms of collaborative coordination since the interactive mode assumes that learners are fully coordinated in their collaboration. Overall, the ICAP framework unifies many of the affordances of these coding frameworks.

Another major affordance for applying the ICAP framework into the museum experience is that the framework goes beyond coding. All the other frameworks mentioned were developed for the purpose of coding museum visitor behaviors and interactions, whereas the ICAP framework was developed for the purpose of classifying the quality of learner cognitive engagement. As compared to that of the other coding frameworks, the ICAP framework is a more generalized framework that has four categories that predict learning outcomes. The ICAP framework is firmly grounded on cognitivist and constructivist theories of learning and based on empirical research that supports its claims on the hierarchy of student learning. While the ICAP framework can certainly also be used for coding, a unique affordance that the ICAP framework also provides is that it can also be used as a designing framework for how to engage visitors in deeper learning. The other coding frameworks can certainly inform an iterative design of exhibits based on observations of visitors interacting within these exhibits. At the same time, the ICAP framework affords museums theoretical tools grounded in cognitive and constructivist learning theories that may enhance the initial design and implementation of exhibits.

There are also a number of limitations to the ICAP framework for evaluating informal learning in museums. First, the ICAP framework aggregates many codes into its four categories and thus has a reduced capacity for examining the nuances that can be examined by other frameworks. For example, the DCLM framework codes for ten distinct collaborative behaviors for which the ICAP has one equivalent category, the interactive mode of engagement. In a sense, the ICAP framework examines how learners are engaging cognitively at a coarser-grained level, whereas these other frameworks examine visitor behavior at a finer level. Second, the ICAP framework focuses solely on the cognitive engagement of learning, whereas other frameworks also include motivational aspects of visitor experiences in exhibits. For example, the VEF codes for emotional responses of visitors. Third, the ICAP framework is derived from classroom learning and not explicitly for informal learning. At the same time, this limitation presents an opportunity for unifying informal and formal (classroom) learning should museum researchers begin utilizing and adapting the ICAP framework for informal learning.

The current study does not focus on coding the behavior of visitors but rather seeks to understand if a mobile question-asking app may support informal collaborative learning. Thus, the use of a theoretical framework is more appropriate than the use of coding frameworks. Furthermore, the coarser-grained analysis of the ICAP framework is better suited for this study since we are not examining the specific nuances of visitor behavior that are afforded by the other coding frameworks. For the purposes of this study, the ICAP framework is utilized to evaluate the design of our museum app to ascertain whether our app can support the facilitation of improved learning.

CSCL experiences in museums

Many CSCL researchers working in museums have investigated how to design engaging museum exhibits (Davis et al., 2015; Roberts & Lyons, 2017; Tissenbaum et al., 2017; Yoon et al., 2012), which have explored technologies as varied as tabletops (Davis et al., 2015; Horn et al., 2016; Tissenbaum et al., 2017), full-body interactives (Meisner et al., 2007; Roberts, & Lyons, 2017; Tscholl & Lindgren, 2016), and other emerging and novel technologies (Vermeeren et al., 2018; Yoon et al., 2012). Another thread of work has explored how mobile devices can collaboratively engage visitors with the museum itself (Cabrera et al., 2005; Cahill et al., 2011; Melero et al., 2015).

To further investigate computer-supported collaborative learning experiences in museums, researchers not only need to be familiar with the frameworks applied in previous studies that characterize the learning mechanisms of museum collaboration (Fleck et al., 2009; Tissenbaum et al., 2017) but also need to know what research topics have been discussed or are in ongoing discussion within the research field, as well as understanding the approaches taken (Dillenburg et al., 1996; Janssen et al., 2010; Stahl, 2007) to analyze the effect of computer technologies on fostering visitor engagement, meaning-making among visitor groups, and collaborative learning processes and performances (Davis et al., 2015; Rahm, 2004; Roberts, & Lyons, 2017).

Prior efforts towards illumination of different mechanisms of collaborative learning in museums gave rise to several research themes. For instance, some researchers have explored visitors' identity-related motivations and investigated the processes through which visitors construct, maintain, and adapt their personal identities while engaging in museum exhibits. (Falk et al., 2008; Rounds, 2006). Other researchers have specifically focused on family learning in and from museums and have identified families as a major unit of analysis for museum research (Crowley et al., 2001; Ellenbogen et al., 2007; Zimmerman et al., 2008). These researchers realized the complex nature of family interactions and provided evidence that families bring a broad array of personal and collaborative learning strategies to their shared experiences in museums. Much of this research was descriptive in nature, depicting family conversations, the roles parents and children assume, gender differences in parent–child interactions, and the similarities and differences between the conversations of families and all adult groups.

As computer-based exhibits proliferated over the past two decades, researchers and museum curators also started to investigate social interaction and collaboration that arose with and around novel computer technologies and the extent to which these computer-based exhibits do indeed facilitate engagement and new forms of participation (Heath et al., 2005; Meisner et al., 2007). Particularly, many studies have recently focused on providing computer support to facilitate collaborative visitor talks or exploring visitors' collaborative interactions with and around computer-based exhibits (Davis et al., 2015; Melero et al., 2015; Roberts & Lyons, 2017; Tissenbaum et al., 2017; Yoon et al., 2012). To foster productive visitor-talk and understand how collaborative interactions influence visitors' shared experiences, researchers have explored various approaches and presented empirical evidence for such approaches. Specifically, Davis et al. established and applied an analytical framework to focus their analysis on the way dyads negotiate their moment-to-moment exploration of the exhibit and how they develop their understanding. Roberts & Lyons introduced the SQuILD as a metric for determining how their designs are supporting productive learning talk. Proponents of emerging technologies (Yoon et al., 2012) also systematically investigated and produced evidence for using digital augmentations and knowledge-building scaffolds to enhance collaborative learning processes and learning outcomes in a science museum.

Although there is ample prior work in CSCL that investigates different kinds of computer support in fostering visitor talk and productive collaborative learning in museums, few studies have placed collaborative question-crafting or group-inquiry as a central activity in their examination. Nevertheless, some researchers have been concerned with addressing research questions related to inquiry behaviors among visitor groups. For instance, Allen and Gutwill's studies (Allen & Gutwill, 2009; Gutwill & Allen, 2010, 2012) have demonstrated effort to extend and deepen inquiry by explicitly coaching visitor groups in the skills of scientific inquiry at interactive exhibits. The app designed for our study also placed question-asking as the central activity, but our work differed from Gutwill & Allen's in two aspects. First, the "Inquiry Games" used in Gutwill and Allen (2010, 2012) were "games" in the sense of steps, rules, and a structure for social interaction. Technically speaking, their work did not contain the use of computer technologies. Second, Gutwill and Allen (2012) manipulated experimental conditions by adjusting different pedagogical methods while we generated conditions based on different app modes. In a nutshell, our current study uniquely speaks to collaborative question-crafting and has the potential to fill the gap of addressing the aforementioned dearth of research providing computer support that encourages museum visitor groups to ask questions collaboratively.

The role of question-asking in collaborative learning

Previous research on question-asking in museums has taken the Vygotskian perspective that question-asking can serve as a form of stimulation for learning. Researchers (Ash, 2004; Hohenstein & Tran, 2007) found that parents demonstrated different styles of questioning, which in turn influenced children's learning behaviors and perceived learning in museums. For instance, parents who took on a teacher-like role and followed a structured question-asking pattern would initiate questions themselves, the children would respond to such questions, and then the parents would evaluate responses. In contrast, other parents were observed to use more of a conversational style, using open-ended questions as an invitation to further dialogue and reflect with their children. It was found that parents who used open-ended questions tended to have children who said they learned more than parents who did not use open-ended questions. Another line of research that focused on object-based learning in museums treated museum exhibits as objects set in a context designed to promote learning. According to Borun (2002), object-based information is often new to adults as well. Therefore, adults and children are both learners in museums. Museums engage everyone in the visitor group to work together in investigating questions of mutual interest. This emphasizes generalizing knowledge through collaborative inquiry (Scardamalia & Bereiter, 1999) in museum contexts.

Given the fact that our app centralizes its main feature on question-asking, it is important to clarify that visitors who used this app could ask different sorts of questions depending on their familiarity with exhibit topics. Ideally, participants in our study could collectively generate questions and ask their questions by typing them down in the app during their visit. In Scardamalia and Bereiter's (1992) seminal study, knowledge-based questions were distinguished from text-based questions. Knowledge-based questions occur spontaneously and are picked up by non-intrusive observations, whereas text-based questions are typically generated on demand, such as instructing students to generate questions in response to certain cues. Furthermore, knowledge-based questions can be categorized as either "basic information questions" or "wonderment" questions (Scardamalia & Bereiter, 1992, p. 186). Basic information questions are questions asked in need of information to orient towards a topic. Wonderment questions are questions asked in need of explanation or resolving knowledge discrepancies, which often reflect "curiosity, puzzlement, skepticism, or a knowledge-based speculation" (Scardamalia & Bereiter, 1999., p. 188).

Through our app, we non-intrusively collected a full range of visitor groups' questions, which were presumed to occur spontaneously, presumably knowledge-based. Unlike Scardamalia and Bereiter (1992), who focused on children's question-asking quality and patterns, we did not primarily focus on diagnosing question-asking behaviors among visitors. The main focus of the current study was to compare the influence of two app modes (Game Mode and Ask Mode) on group discussion frequency, question generation responsibility, and collaborative discussion patterns. At the same time, in order to elucidate whether the types of questions asked by visitors in our app would promote CSCL, we analyzed visitor questions using Scardamalia and Bereiters' taxonomy of question types in order to examine whether visitors were generating deeper learning knowledge-based wonderment questions.

Games and gamification in museums

Broadly defined, gamification is the use of elements that are traditionally thought of as game-like or "fun" to promote learning and engagement in non-game contexts (Deterding et al., 2011). Gamification may be achieved through a wide variety of techniques and design strategies adapted for multiple goals and contexts (Dignan, 2011). In the context of learning and education, gamification is a growing phenomenon (Dichev & Dicheva, 2017). Here, it has primarily included implementing extrinsic reward features typically associated with games (e.g., reward systems, narrative structures, aesthetics) to nongame learning environments with the purpose of making such environments more "fun" or motivating to promote learning and engagement (Kapp, 2012; Plass et al., 2019). Gamification, however, is far from flawless. It has often been critiqued as a "missed opportunity to rethink and redesign the learning task" (Plass et al., p. 4). Additionally, previous work on gamification (defined as game elements applied to non-game settings) in learning reveals mixed conclusions. For instance, Hanus and Fox (2015) pointed to adverse effects on constructs such as motivation, satisfaction, and empowerment in gamified classrooms, while Tan and Hew (2016) found that learning, engagement, and affective outcomes were higher in gamified learning contexts. Similarly, even larger-encompassing reviews of the literature on gamification reflect such mixed results, with some (e.g., Dichev & Dicheva) concluding that there is insufficient evidence to support the long-term benefits of gamification in educational contexts while others (Sailer & Homner, 2020) concluding that gamification is an effective method for learning.

Despite the mixed findings of gamifications' impacts on learning, which point to how not all gamification techniques make sense for all contexts, gamification nevertheless continues to be a popularly implemented strategy across formal and informal learning contexts such as museums. Because interactions (e.g., collaboration, competition) can be a particularly engaging feature of games in general (Rigby & Ryan, 2011) as well as gamified learning (Arnab et al., 2016), work on game-based and gamified learning in museums has often emphasized collaboration.

Gamified work in museums has found that interactions between visitors have taken various forms depending not just on the number of visitors but on the roles that each have taken on. For instance, Cabrera et al. (2005) found that collaboration between team members has taken various forms, including the expected (designed for) pattern of exchanging pieces of images or texts through the game's interface, but it also included verbal collaboration, helping others find relevant information in the physical space, and using others' mobile devices to find the desired piece of information. Similarly, Klopfer et al. (2005) implemented a digital handheld game in which users played multiple roles (e.g., technologist, biologist, and detective) to solve a museum theft and found that these interdependent roles fostered collaboration within groups and encouraged users to think about the museum's exhibits more deeply. Additionally, observations of families using an interactive portable guide system in a museum show how unexpected usage (e.g., demands of parenting) can occur through particularities of interaction between family members (Hope et al., 2009).

Gamified work in museums has shown that personalizing "quests", for instance, to add elements of collaboration showed no significant difference in museum enjoyment (van Dijk et al., 2012). Comparisons of the complexity in user interfaces showed that more complex interfaces promote fewer visitor-visitor interactions but increase the quality of conversations, while simpler interfaces encourage competitive behaviors but show disproportionate impacts on female visitors (Lyons, 2008, 2009), a finding that points to the worth of determining the affordances of mobile technologies in museums for specific purposes and people, such as parent–child dynamics (Lyons et al., 2010).

Gamified work in museums has also found that having either mobile or fixed devices distributed throughout a museum impacts cooperation among visitors (Dini et al., 2007). Similarly, input style (e.g., directional keypad, multi-touch, device tilt) for such devices has been found to impact learnability, museum feasibility, and sociability and collaboration (Jimenez & Lyons, 2010). Additionally, mobile devices have been found to support children's collaborative learning, provide opportunities for interactions, and give them opportunities to pay more attention to less interactive and less attractive exhibits (Yatani et al., 2004; Yiannoutsou et al., 2009).

In summary, the work overviewed in this section illustrates some of the strategies and choices that have impacted collaboration in gamified work in museums. Our app's design is informed by these previous findings on increasing engagement and interest in museum exhibits through the use of mobile technology and gamification. For instance, our app runs on a hand-held device that can be carried throughout the exhibit hall and be used by groups of visitors, in which any of the group members could input questions. In our previous work (Nelson et al., 2020), where we operationalized engagement as the number of questions asked, we expected (and found) that gamified elements heightened visitor engagement. Given these findings, we expect to find differences in the way users interact collaboratively between gamified and non-gamified modes. In the following section, we discuss our app's design in more detail.

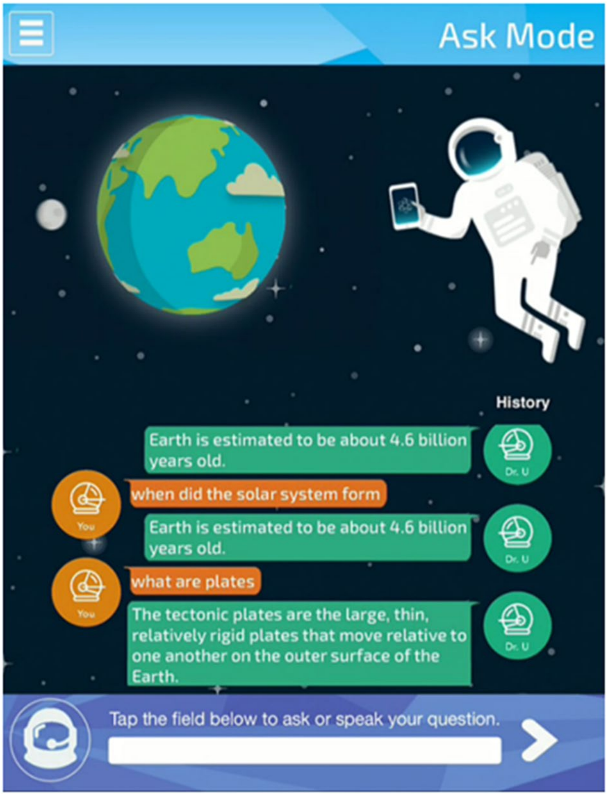
Methodology

Our mobile app

For our research project, we developed and implemented a mobile app that presented a virtual astronaut character, Dr. Discovery, for use by museum visitors that doubled as (a) a front-end museum expert and an unobtrusive data-gatherer and (b) a back-end analytics data bank that can be mined by museum staff, evaluators, and researchers. As part of the larger museum evaluation study, Nelson et al. (2017) designed the mobile app with two modes, Ask Mode and Game Mode (see Figs. 2 and 3), to assess gamification's impact on question-asking rates. Our mobile app was implemented in two museum partner sites: a science center and a natural history museum. The app is connected to a database of more than 13,500 museum content-related questions and 5000 verified answers (many questions have the same answer) curated by our project team. To build this question database and develop answers, research assistants and grant faculty walked through the exhibits entering questions and then looking up answers while frequently consulting with museum personnel. During the designing phase of our app, we regularly updated the answer pool based on questions asked by users that the system could not answer. This was a multi-person team in which one person would investigate and enter answers, and one or more faculty, museum personnel, or research assistants with content area expertise would approve the provided answer. No further updates to the database were made once we began our implementation at the museums.

Ask Mode was the baseline treatment that presented a simple Q&A interface through which museum visitors typed or spoke questions about exhibit content. It presented a simple question interface with only a few interactive elements, such as a static picture of an astronaut, to not distract visitors from the exhibit. Tapping on the text entry box pulled up the built-in iOS keypad, and users simply typed in or spoke questions they had about

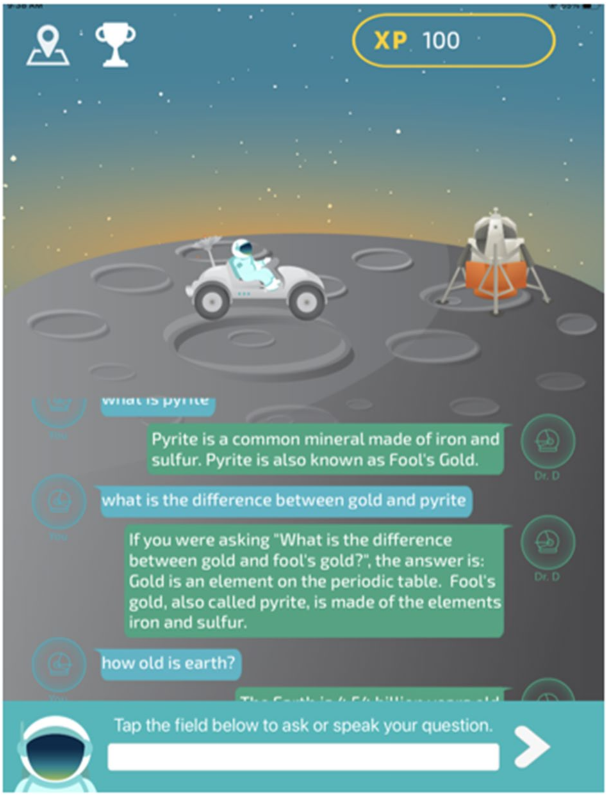
Fig. 2 Ask Mode



the exhibits. Ask Mode had no points, goals, nor badges. Game Mode incorporated several simple gamified elements. The main game mechanic was asking questions to move the astronaut's vehicle toward a goal (moon lander), with each question asked moving the vehicle a small amount toward the goal. This included a simple animation of the astronaut driving closer to the moon lander and earning experience points (100 points per question; Fig. 3). In addition, badges were awarded for reaching milestones of asking questions (e.g., "Advancing Adventurer" for asking ten questions) and for maintaining question asking over time (e.g., "Quick Questioner" for asking five questions in 5 min), see Figs. 4 and 5.

Our previous findings have shown that Game Mode elicited more questions on average than Ask Mode (Nelson et al., 2020). However, our mobile question-asking app aimed to go beyond data collecting, hoping to incentivize museum visitors to engage with exhibitions collectively and collaboratively through generating questions. According to the ICAP framework, generating and asking questions would engage users at the higher levels of engagement, constructive if individually generated, and interactive if co-generated (Chi, 2009; Graesser & Person, 1994). Considering the design of the user experience elicits higher levels of ICAP engagement modes, the app may support improved informal learning for group visitors utilizing the device. In both the Ask and Game Modes, visitors were regarded as a team of explorers and co-constructors of scientific questioning. The question-driven media platform was designed to allow participants to form stronger mental connections with the exhibits' information and allow for interactions with other group members during the collective experience (Clark & Sengupta, 2019; Honey & Hilton, 2011). Another affordance of the group interactions provided by our mobile app was

Fig. 3 Game Mode



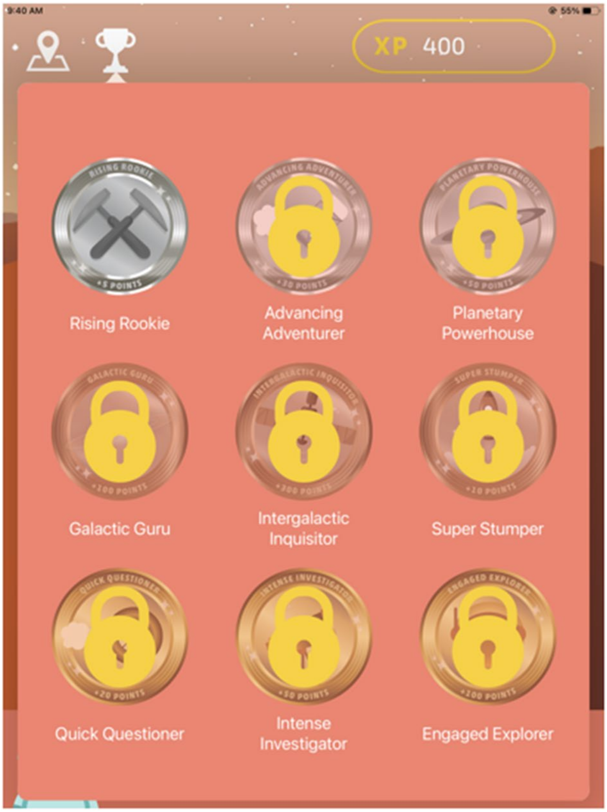


Fig. 4 Game Mode Badges

that it was designed to allow open inquiry (Banchi & Bell, 2008), a process of proposing and testing questions in a self-paced and visitor-centered learning environment.

Alignment with CSCL support dimensions

Recently, Rummel (2018) proposed a taxonomy framework designed to support future CSCL researchers and designers to "make their assumptions about effects of CSCL support explicit" (p. 124). Although the current study did not design the app with these principles in mind, the app can be categorized using the taxonomy framework. Table 1 lists the types of CSCL dimensions the app supports.

The goal of our app was to support domain knowledge by providing answers to visitor questions and to increase engagement (operationalized as the number of questions asked) via gamification features in the Game Mode such as badges, experience points, and simple challenges. The timing of the support was during the collaboration since the app gave real-time answers to questions. The implementation of the app was adaptive to the types of questions, the number of questions, and the timing of user questions. The delivery agent was a digital persona in the form of an astronaut. Our two app modes differed on the target support dimension. The target of the Ask mode was cognitive by providing information based on user questions, whereas the Game mode supported motivation via gamification features in addition to the cognitive support of the

Table 1 CSCL Support Dimensions of our App

Category	Support Dimensions
Goal	Domain knowledge (and Learning motivation for Game mode)
Timing	During the collaboration
Implementation	Adaptive
Delivery agent	Digital Persona
Target	Cognitive (and Motivational for Game mode)
Granularity	Turn level
Availability	On demand
Directivity	No advice
Foundation	No information on the collaboration state
Addressee	Individual or Group
Meditation	Direct
Coercion	No action required

Ask mode. Granularity was at the turn level since the support our app provided was only given whenever users took a turn inputting a formulated question. Support from our app was available on-demand, with users only receiving support whenever they chose to input a question. The directivity of our app had no guidance on how users should behave or interact with the device. Our app provided no foundation information about the state of collaboration. The addressee of the support given to the group was directed towards individuals or the group depending on whether users individually or collectively asked questions. The mediation of our app was given directly to the learner or group by directly providing answers to questions. Finally, there was no coercion since users had the choice to follow up on the support provided by the app.

Research questions

In examining to what extent the Ask and Game app modes may foster the interactive mode of engagement among visitor groups:

(a) What is the frequency of self-reported visitor group discussions about question-asking topics? (group discussion frequency).
(b) How do groups report primary question-asking responsibility? (question generation responsibility).
(c) Does the group discussion frequency vary by who has the primary question-asking responsibility? (collaborative discussion patterns).
(d) Does group discussion frequency, question generation responsibility, and collaborative discussion patterns vary by app mode?

Data

This paper examines data collected over seven months by our project partners located in a large urban area in the American southwest: a natural history museum (Museum A) and a science center (Museum B). In this study, 5632 participants in 1539 visitor groups took part in one of two conditions: Ask (614 groups) and Game (925 groups). Though we collected a range of other visitor information, e.g., visitor interest in science and access to smart devices, the current study focused on data most relevant to the Learning Sciences and CSCL communities. This study focuses on groups' post-questionnaire self-reported question-asking behaviors, specifically exploring if and how groups discussed the questions they wanted to ask using the app, both overall and by condition (Ask or Game), and how groups divided question-asking responsibilities. More specifically, the survey items of interest were:

(a) "How often did your group discuss what questions to ask the app?", with forced-choice answer options of "Not at all," "A little bit," "Some of the time," "Most of the time," and "Every time."

(b) "Who in your group asked the app the most questions?", with forced-choice answer options of, "I asked the most," "We asked together as a group," "Another adult (male)," "Another adult (female)," "A kid (male) age ____," and "A kid (female) age: ___."

Procedures

Museum A has served over 2 million visitors since 2000, including thousands of students from hundreds of schools annually, and Museum B serves 500,000 guests per year, including nearly 140,000 students. At Museums A and B, we conducted implementations using Ask Mode and Game Mode using 13 iPad minis preloaded with either the Ask or Game Mode. On alternating implementation days, all participants were blindly assigned to either Game or Ask Mode. We scheduled the days in an alternating manner in order to ensure each mode had a similar number of implementations for each day of the week (e.g., Thursday, Friday, Saturday, or Sunday). Although we scheduled the number of implementation days for each mode to be equivalent, outside factors of our in-situ study skewed the data to have more participant groups for the Game Mode. Two major factors could be attributed to the fluctuation in the number of visitors for certain days, including vacation days for children and teenagers as well as days that are proximal to major holidays. For example, there was a spike in the number of visitors during Christmas break and the days leading up to Independence Day. In contrast, very few visitors showed up for Halloween weekend.

Any willing participants or participant groups with an adult over the age of at least 18 were recruited from the floor using a standardized script and blindly assigned to the Ask or Game condition depending on the implementation day. There were no limitations to group size. The research team explained the app's purpose, lent out an iPad mini preloaded with the app and pre-survey, and invited participants to use the app while exploring a targeted exhibit hall. A primary participant, a self-selected adult from each group, completed a time- and device-stamped demographic and background survey before—and feedback survey after—participation. The background survey asked questions about interest in science and the frequency of museum visitation. Participant groups were told that anyone in the group could use the tablet during their visit. Participants explored the exhibits at their own pace, and their app interactions were recorded in the online database. Implementation days generally ran three to four hours a day.

On the back end of the app, inputted questions are matched to the database by string matching. The more strings of text matching an existing question in the database, the more certain the app is of what question is being asked. On questions where it is uncertain what is being asked, the user will be given a choice of the closest matching questions in the database, at which point the user may also indicate that none of the questions offered represent their posed question.

To answer the current study's research questions and decode museum visitor groups' question-asking patterns (group discussion frequency, question generation responsibility, and collaborative discussion patterns) under two modes (Ask and Game) of our mobile app, we analyzed two survey items from the combined data set from the two museums. To ensure the legitimacy of combining data from these two sites, we conducted an ANOVA for responses to the two survey items mentioned in the data section with our museums as conditions. We found no significant differences in either app mode ($p_a = 0.558$ & $p_b = 0.407$).

Participant demographics

There were 1539 groups for the sample of this study. We defined "groups" to include adults who came to the museum by themselves. Of all 1539 groups, 1064 groups were family groups, groups with at least one adult as well as at least one child or teen, 453 groups were adult groups, groups with only two or more adults, and 24 groups were solo adults. In addition, 352 of the groups were duos, and 1084 of the groups were trios or greater.

Groups averaged 3.65 people. There were more women (1.35), on average, in a given group than men (1.27). Most participant groups (82.1%) included at least one woman, while 64.4% had at least one man. Groups averaged at least one pre-teen child (1.81), and roughly a quarter of participant groups included a child under five years old (27.2%). The average group was unlikely to have a teenager (15.4%). Overall, 76.0% of visitor groups included at least one minor.

Of all 1539 groups, the gender distribution of the group leader was 58.7% female, 38.1% male, and 3.2% did not disclose gender. Group leaders' ages were distributed as follows: 14.2% were 18–25 years old; 29.9% were 26–35 years old; 33.9% were 36–45 years old; 10.8% were 46–55 years old; 8.6% were 56 years old or older; 2.6% did not report their age.

Group leaders reported their ethnicity with the following proportions: 65.8% as White, 19.7% as Hispanic/Latino, 5.7% as Asian or Asian American, 4.5% as Black or African American, 3.3% as Native American, and 1.0% as Native Hawaiian/Other Pacific Islander.

A total of 27.9% of group leaders reported having a college degree, 25.4% reported having a graduate degree, and 30.2% of the respondents reported having attended some college or having an associate degree. Less than 10 percent of respondents (8.2%) reported having completed no higher than a high school degree or GED.

Quality of questions

At Museum A, there were 560 groups with 266 Ask Mode groups and 294 Game Mode groups. Groups asked a total of 939 questions with an average of 3.5 questions for Ask Mode and a total of 1900 questions with an average of 6.5 questions for Game Mode. In addition, groups spent an average time of 10.5 min for Ask Mode and 12.5 min for Game Mode.

At Museum B, there were 979 groups with 348 Ask Mode groups and 631 Game Mode groups. Groups asked a total of 2154 questions with an average of 6.2 questions for Ask Mode and a total of 6916 questions with an average of 11 questions for Game Mode. In addition, groups spent an average time of 18 min for both Ask and Game Modes.

In other work (Su et al., 2021), we explore how the different exhibits have impacted the content of the questions. Users were able to ask questions on any topic (including off-topic questions) in both museums and in both modes. Our answer database included answers related to the specific exhibit content at both museums (and not to anything else). If the app didn't know the answer to an inputted question, it would indicate that to the user.

We analyzed the full sample of 11,909 questions visitors asked during the study. We checked the quality of questions by flagging incomplete questions, e.g., "who is," "where do," and "how," off-topic questions, e.g., "hi" and "who are you," and nonsense inputs, e.g., "jsogvsjdkj." Visitors at the museums asked incomplete, off-topic, or nonsense questions 590 times (19.1%) in Ask mode and 967 times (11.0%) of questions in Game mode, see Table 2. Using Scardamalia and Bereiter's (1992) taxonomy of question types, we rated visitor questions as either basic information or wonderment. For Ask Mode, visitors asked 591 (19.1%) wonderment questions and 1912 (61.8%) basic information questions. For Game Mode, visitors asked 1635 (18.5%) wonderment questions and 6214 (70.5%) basic information questions. Chi-squared tests resulted in significant proportional differences, $p < 0.001$, for incomplete, off-topic, or nonsense questions as well as basic information questions by app mode. As compared to that of the Ask Mode, the Game Mode elicited fewer nonsense, incomplete or off-topic questions as well as a greater proportion of basic information questions. No significant proportional differences were found for wonderment questions across app mode, $p = 0.56$. Interrater reliability was found to be very high (Cohen's $k = 0.90$; $p < 0.001$).

In Table 3, we sampled ten random visitor questions and rated the questions as wonderment or basic information questions using the question taxonomy from Scardamalia and Bereiter (1992). Though this limited sample of questions is not sufficient for making broad generalizations, it paints a picture of the types of questions asked by visitors.

Group discussion frequency

For group discussion frequency, we excluded the 24 groups of solo adults in the analysis for a total sample size of 1515. Chi-square tests revealed that in the overall data, there was a significant difference ($p < 0.001$) between the responses by mode. We found significant differences for all responses except for 'some of the time' (see Fig. 6).

To pinpoint the differences, we performed a post-hoc chi-square analysis of adjusted standardized residuals for each cell (Beasley & Schumacker, 1995; García-Pérez & Núñez-Antón, 2003). García-Pérez and Núñez-Antón (2003) found that correcting for the family-wise error rate in cellwise residual analyses of two-way contingency tables is a matter of preference with only small departures from the nominal error rate once the sample size

Table 2 Quality of Questions

n = 11,909	Ask Mode	Game Mode
Nonsense, Incomplete, or Off-topic Questions	590* (19.1%)	967* (11.0%)
Basic Information Questions	1912* (61.8%)	6214* (70.5%)
Wonderment Questions	591 (19.1%)	1635 (18.5%)

* $p < .001$

Table 3 Random Sample of Visitor Questions

Visitor Question	Question Rating
What type of rock is quartz?	Basic Information
Why is the sky blue?	Wonderment
How far away is the moon?	Basic Information
What color is the earth?	Basic Information
Why are volcanoes around the Pacific rim?	Wonderment
When did man first roam the earth?	Basic Information
What are the benefits of dams?	Wonderment
What is the name of the largest sun?	Basic Information
Was there life in snowball earth?	Basic Information
In environments in which sand dunes are located what type of climate change occurs year round?	Wonderment

reaches 500 participants. Considering our study had $n > 500$, we chose to base our analysis on an unadjusted α significance of 0.05 (see Table 4).

In response to how frequently they discussed what questions to ask the app, 37.6% of participant groups in the Ask Modes and 25.4% of participant groups in Game Mode selected "a little bit," a difference of 12.2%, representing the widest margin of significant

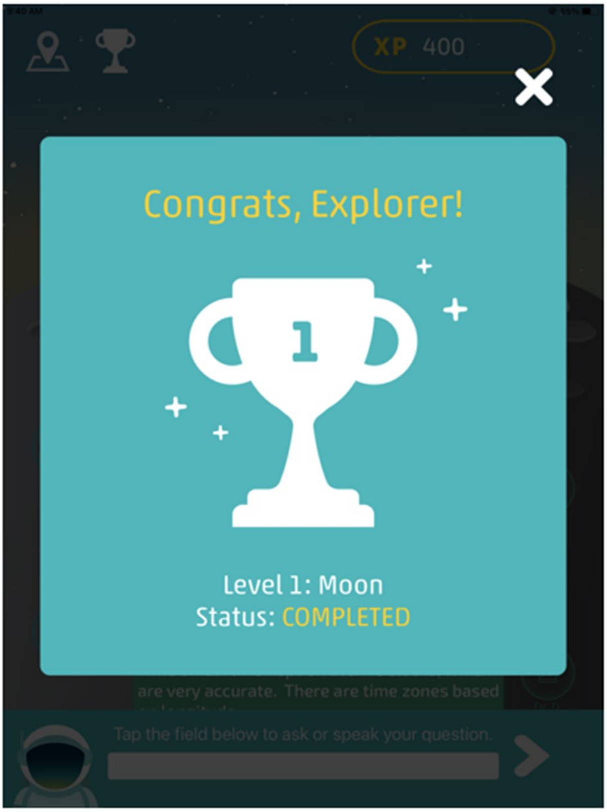


Fig. 5 Game Mode Award

difference between modes ($p<0.001$). For 'most of the time,' participant groups in the Ask and Game Modes reported significant differences ($p<0.001$) with 14.2% and 25.4%, respectively, a difference of 11.2%. Results also showed significant differences between Game and Ask Mode groups on 'not at all' and 'every time' ($p<0.05$), which favored the Game Mode.

In a further analysis, we separated family groups and adult groups to see if there were any differences based on the types of group members. Figures 7 and 8 show group discussion frequency for family groups and adult groups, respectively. Chi-square tests for both types of groups showed significant differences ($p<0.001$). The standardized adjusted residuals for family groups revealed greater differences for "a little bit" ($Z=3.9$) and "most of the time" ($Z=4.3$), whereas adult groups had respective differences of $Z=3.2$ and $Z=2.9$.

Question generation responsibility

For question generation responsibility, we excluded 453 adult groups in addition to the exclusion of 24 adult solo groups for a total sample size of 1062. Ask and Game Mode participant groups showed relatively similar results regarding question generation responsibilities (Fig. 9). Chi-square tests revealed that overall, there was no significant difference ($p=0.211$) between app mode responses. The proportion of adults reporting as the primary question-asker was higher in Ask Mode groups vs. Game Mode groups, 38.8% and 34.2%, respectively. In addition, Game Mode participant groups were more likely to report asking questions as a group (18.5%) than Ask Mode groups (14.3%).

Table 4 Group Discussion Frequency by App Mode

		Mode:		Overall Percent	Total Count
		Ask	Game		
Not at all	Count (%)	98 (16.2%)	115 (12.6%)	14.1%	213
	Z	2.0	−2.0		
	χ^2 Sig	.046*	.046*		
A little bit	Count (%)	227 (37.6%)	231 (25.3%)	30.2%	458
	Z	5.1	−5.1		
	χ^2 Sig	.000*	.000*		
Some of the time	Count (%)	176 (29.1%)	270 (29.6%)	29.4%	446
	Z	−.2	.2		
	χ^2 Sig	.841	.841		
Most of the time	Count (%)	86 (14.2%)	231 (25.4%)	20.9%	317
	Z	−5.2	5.2		
	χ^2 Sig	.000*	.000*		
Every time	Count (%)	13 (2.2%)	44 (4.8%)	3.8%	57
	Z	−2.7	2.7		
	χ^2 Sig	.007*	.007*		
No Response	Count (%)	4 (0.7%)	20 (2.2%)	1.6%	24
Total	Count	604	911		1515

*unadjusted significance $\alpha=0.05$

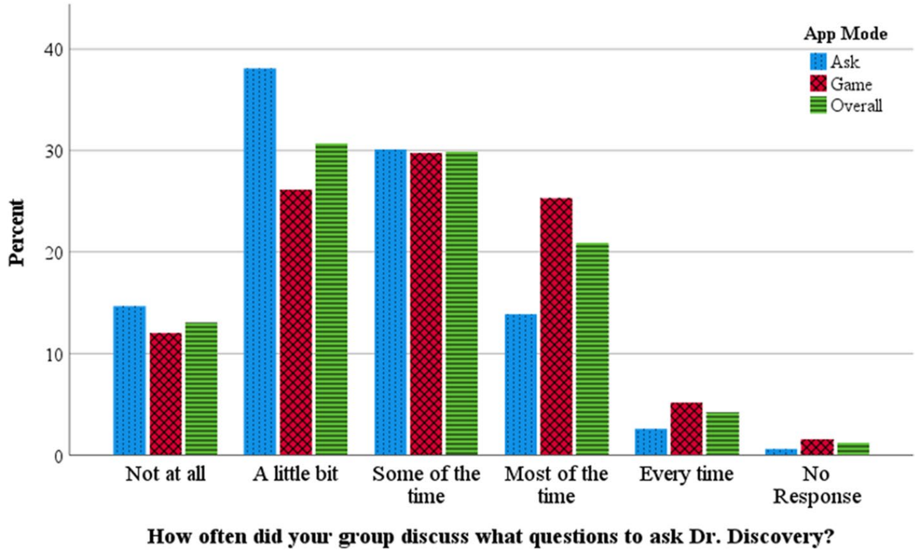


Fig. 6 Group Discussion Frequency by App Mode

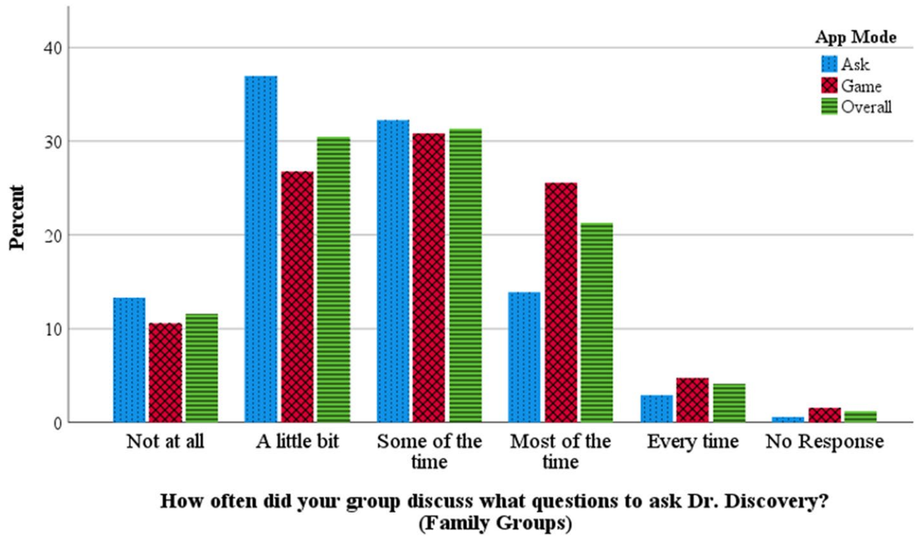


Fig. 7 Family Group Discussion Frequency

However, further post-hoc chi-square tests of adjusted standardized residuals did not result in a statistically significant difference for adults as primary question-asker between app modes ($p = 0.134$) and group question-asking between app modes ($p = 0.072$). Child and teen question-asking rates were roughly equal across app modes, with 37.6% in Ask Mode and 38.6% in Game Mode.

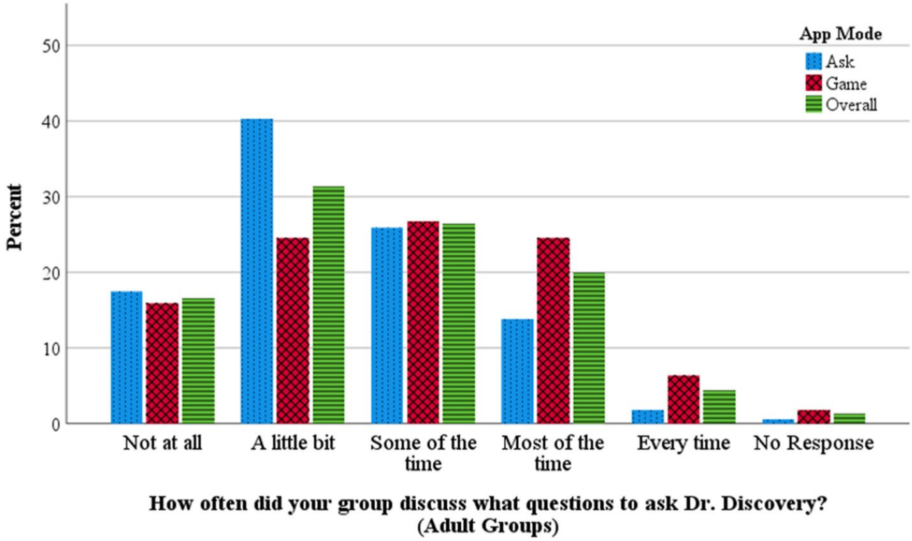


Fig. 8 Adult Group Discussion Frequency

Collaborative discussion patterns

For collaborative discussion patterns, we excluded 453 adult groups in addition to the exclusion of 24 adult solo groups for a total sample size of 1062. There were clear differences in group collaboration by both mode and primary question asker(s). Overall, in groups that reported a child asking most of the groups' questions, the group also reported discussing the questions to input into the app more frequently than in groups that reported an adult asked most of the groups' questions. In addition, groups that reported sharing the responsibility for asking questions to the app were the most likely to collaborate on their question-asking. Given there were clear differences in group collaboration by both app modes and who asked the most questions, we are interested in examining the interaction of these two variables. Generally, groups with children-led app use had higher group discussion frequency than those with adult-led app use. As expected, group-led app use had the highest group discussion frequency. Figure 10 shows the group discussion frequency based on which primary question-asker asked the most questions. The results found a hierarchy of collaborative question-asking of groups > children > adults.

In addition, these differed by app mode, with Game Mode groups reporting a higher frequency of group discussions and collaboration regardless of the reported primary question asker. Chi-square tests confirmed significant differences; Ask Mode ($p < 0.001$), Game Mode ($p < 0.001$), and overall ($p < 0.001$) at a significance level of $p < 0.05$. Figure 11 splits the Ask and Game Mode responses into clustered bar graphs that show an increase of collaboration in the Game Mode for all three types of question-askers. Further post-hoc chi-square tests of standardized residuals, Table 5, for each response revealed that the Game Mode resulted in fewer responses for "a little bit" and greater responses for "most of the time."

Groups with adult primary question asking were equally likely to respond, "not at all" (Ask = 20.3% & Game = 20.1%) regardless of mode, whereas groups

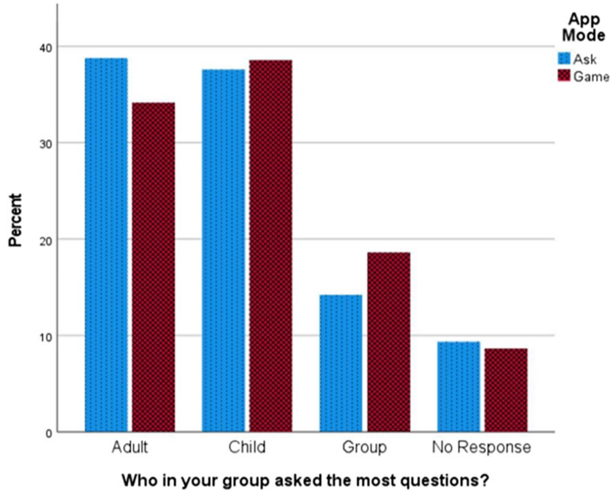


Fig. 9 Question Generation Responsibility

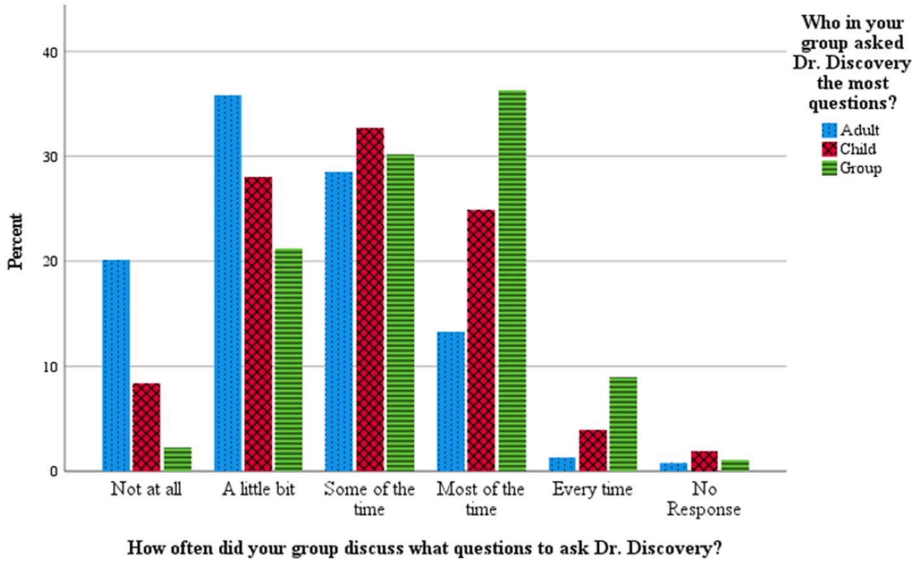


Fig. 10 Collaborative Discussion Patterns by Primary Participant (Overall)

with children primary question asking showed a decrease of responding "not at all" (Ask = 10.5% & Game = 7.1%), see Table 5. For both adult and children primary question asking, there was a decrease in the response "a little bit" from Ask to Game mode ($Z = 2.5$; $p < 0.05$). In addition, whenever children were the primary question asker, there was an increase in the response "most of the time" for Game Mode ($Z = 3.1$; $p < 0.05$). There were no statistically significant differences for group collaboration frequency by user type between modes whenever the group primarily asked questions.

Table 5 Collaborative Discussion Patterns by Primary Participant * App Mode

		Adult		Child	
		Ask	Game	Ask	Game
Not at all	Count (%)	32 (20.3%)	45 (20.1%)	16 (10.5%)	18 (7.1%)
	Z	0.0	0.0	1.2	−1.2
	χ^2 Sig	0.969	0.969	0.239	0.239
A little bit	Count (%)	68 (43.0%)	69 (30.8%)	54 (35.3%)	60 (23.7%)
	Z	2.5	−2.5	2.5	−2.5
	χ^2 Sig	0.014*	0.014*	0.012*	0.012*
Some of the time	Count (%)	41 (25.9%)	68 (30.4%)	51 (33.3%)	82 (32.4%)
	Z	−0.9	0.9	0.2	−0.2
	χ^2 Sig	0.347	0.347	0.848	0.848
Most of the time	Count (%)	16 (10.1%)	35 (15.6%)	25 (16.3%)	76 (30.0%)
	Z	−1.6	1.6	−3.1	3.1
	χ^2 Sig	0.120	0.120	0.002*	0.002*
Every time	Count (%)	1 (0.6%)	4 (1.8%)	4 (2.6%)	12 (4.7%)
	Z	−1.0	1.0	−1.1	1.1
	χ^2 Sig	0.329	0.329	0.285	0.285

*unadjusted significance $\alpha = 0.05$

Discussion

Quality of questions

Participants, for the most part, asked fully developed questions. Overall, 13.1% of questions asked were flagged as nonsense, incomplete, or off-topic. In terms of question quality for those not flagged, about two-thirds of the questions asked were knowledge-based basic information questions, and about a fifth of the questions asked were knowledge-based wonderment questions (see Table 2). We found significant differences in the types of questions asked across conditions, with the Game Mode eliciting greater basic information questions and fewer nonsense, incomplete, or off-topic questions.

We were interested to see if visitors would try to “game” the Game Mode by asking more nonsense questions in order to earn more experience points or badges; however, we found that this was not the case. On the contrary, we found that the Game Mode elicited more on-topic questions, albeit mainly basic questions. Based on these results, the gamified version of our app seemed to encourage the on-topic asking of questions but did not affect the asking of wonderment level questions. One of the goals while designing the Game Mode was to support and increase motivation by including simple gamification features that would encourage users to use the app while not distracting users from the exhibit content. The badges were simply quick animations that provided positive feedback to users asking questions to the app. Other than simple “pointsification” in the form of badges and points that are popular gamification elements (see Seaborn & Fels, 2015 for a review), there were no major rewards or animations for asking more questions. Our findings on question quality suggest that gamifying museum apps in a manner that gives positive feedback without creating an artificial need to “game” the app may produce greater on-topic usage.

Considering that about a fifth of the questions were knowledge-based wonderment questions, which would be considered generative under the ICAP framework, these results suggest that our app may be able to support visitor cognitive engagement in the Constructive or Interactive modes. Hence, our app may allow for the possibility of deeper informal collaborative learning, considering users asked generative questions.

Another finding of interest was the length of time users spent on our apps. At Museum A, users spent an average of 12.5 min in the Game Mode compared to 10.5 min in the Ask Mode. Museum B showed no differences in terms of user time by mode. From an ICAP standpoint, Museum A would be classified as engaging visitors at the passive mode because the linear exhibit hall does not have any hands-on elements, whereas Museum B would be classified as engaging visitors at the active mode because the open-room exhibit was designed with many hands-on elements for visitors to manipulate. Figures 12 and 13 show some of the differences in the exhibits, which we will explore more deeply in a forthcoming paper (Pérez Cortés et al., 2021). The Game Mode of the app may have encouraged visitors to slow down in walking through the linear exhibit hall in Museum A by allowing them to take the time to ask questions and receive answers via the app.

Group discussion frequency

We found that groups using the gamified version of the app were significantly more likely to report discussing questions together 'most of the time' and significantly less likely to report discussing questions together only 'a little bit.' Furthermore, an analysis of family and adult groups showed an even greater significant increase for family groups discussing the generation of questions in Game Mode than for adult groups. These findings suggest that the game version of the app supported greater levels of group discussion than that

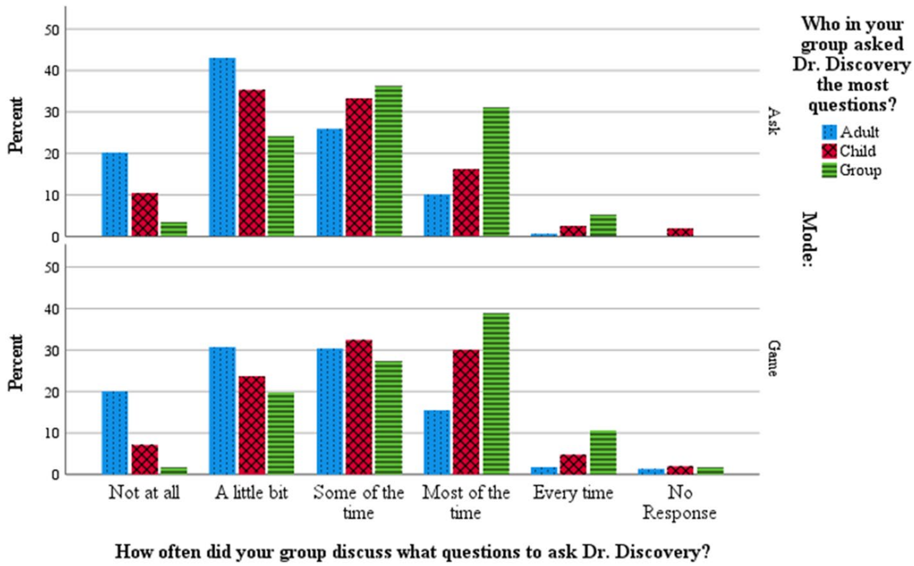


Fig. 11 Collaborative Discussion Patterns by Primary Participant * App Mode

Fig. 12 Museum A—Linear Exhibit Hall

of the non-game version, more so for family groups. The Game Mode's simple gamification mechanics, such as visually marking the progress towards a goal as well as points and simple badges awarded for asking questions, appear to have increased motivation and group discussion while at the same time not encouraging users to "game" the app in order to unlock more compelling rewards than our simple badges. In addition, the Game mode may have encouraged more collaboration, as seen in Fig. 11. This increase in motivation for using the gamified version of the app may also boost the overall use of the mobile app since there was only one tablet available per party, which may encourage members to collaborate.

Moreover, visitors may have informally learned more about the exhibit topics considering the ICAP mode of engagement. Asking questions, the main feature of our mobile app falls under the constructive mode of engagement (and, if it involved two or more group members, the interactive mode of engagement) in the ICAP framework because the act of devising a question is generative (Chi, 2009; Graesser & Person, 1994). The increased likelihood of groups discussing questions in the gamified condition predicts a higher level of engagement and informal learning for those devising questions to input into the query. This suggests that Game Mode could have had a positive impact on the informal learning that occurs within museum exhibits. An implication of the lower percentage of nonsense questions in the gamified mode may be that museums should consider employing simplified gamification in ways that give positive feedback without too many distracting elements when designing their CSCL experiences to increase both visitor learning from museum exhibits and collaboration within visitor groups. This way, such efforts can decrease the "heads-down experience" (Valente-Marques, 2017, p. 15) problem that diverts attention to the content of a screen and takes it away from the content in the museum and the potential social interactions with other visitors.

Fig. 13 Museum B—Open Room Exhibit Hall

Question generation responsibility & collaborative discussion patterns

Visitor groups tended to designate children and teens as well as multiple members of the group as the primary question asker(s) more for the Game Mode than Ask Mode though our results did not show a statistically significant difference.

We found that the group member who managed to ask the app the most questions can change the museum experience and group interactions significantly. Whenever a child or teenager was reported as generating or inputting the most questions, groups similarly reported increased discussion of questions as opposed to whenever an adult was in control. Though we cannot ascertain whether the reported data refers to either the generation or inputting of the questions or both, the results suggest that whenever children and teenagers were more involved in the process of using our app, groups reported increased discussion around the questions. When the two app modes are compared, the differences were more pronounced (Fig. 11).

A possible explanation is that children and teenagers may be eliciting help from other family members in formulating meaningful questions to ask the app. Groups that reported collaborating on questions also reported the greatest level of discussion before inputting questions into the app. This is somewhat expected, considering collaboration assumes discussion among group members. The increased discussion and collaboration based on a child or group taking control of the device would also predict higher levels of ICAP engagement with a stronger emphasis on the interactive mode. An implication of these results may be that museums should find ways to encourage participants to collaborate or have children or teenagers take control of devices such as our app in order to increase the interactions among party members and improve the informal CSCL experience.

Limitations

Although we had two modes for comparison, our study did not collect equivalent data from a control group that did not have any mobile device. Therefore, our results cannot ascertain whether our mobile apps would improve discussion and collaboration as compared to that of business-as-usual museum experiences without mobile apps.

Another limitation of the current study is the post-hoc nature of our analysis. This study analyzed data that was already gathered in a study that focused on the viability of mobile apps for museum evaluation (Nelson et al., 2017). The original study did not gather data on visitor learning. Thus, this paper makes no claims on student learning but points towards the possibility that gamified mobile apps in museums can act as a conduit for meaningful informal learning through the collaborative asking of knowledge-based questions. In addition, based on the available data of the original study, we cannot ascertain whether those who collaborated on asking questions reached the interactive mode as defined in the ICAP framework since groups could have had dominating members. At the same time, at a minimum, app users would have reached the constructive mode since someone in the group generated questions.

Assignment to app mode (Game or Ask) was not fully randomized. For each implementation, different modes were selected on different days in each museum on a rotating basis so that on any given implementation day, all participants were assigned to the same group. However, care was taken to ensure that each mode was implemented on varying days of the week to avoid having one app mode always implemented on a weekend day, for example.

Another limitation of the current study, and a key area for future research, is that our data on group interactions do not include field observations of groups as they explored exhibits or formal interviews with groups after they completed their exhibit visit. Such focused qualitative data gathering could add nuance to the current survey-based findings supportive of gamified question-asking apps reported here. Moreover, analysis of qualitative data could have validated our self-reported data as well as see the frequency of app use by children, teenagers, and adults.

A timely limitation to note is our survey's lack of gender identity options that encompass more than male or female as options. Future surveys should consider allowing for users to input their own gender as well as have a wider array of options available than the binary, male or female.

Conclusions and Future Research

Recently, museums have shifted their exhibit design from a traditional transmission model to a dialogical model (Mclean, 1999; Simon, 2010), aligned with ICAP's interactive mode of engagement. In principle, both the Ask and Game Modes of our app reached the interactive mode of engagement through the main feature requiring users to generate and ask questions to our app. In terms of Rummel's (2018) framework, the Game Mode supported two target dimensions, cognitive and motivational, whereas the Ask Mode supported only one, cognitive. We found that users generated and asked more questions in the Game Mode, implying that gamification may support increased informal learning. In addition, mobile apps with multidimensional target support, such as our Game Mode, may improve CSCL experiences instead of single target support. Based on these findings, museums may promote increased productive collaboration by gamifying their CSCL experiences. In terms of the ICAP framework, considering the Game mode elicited a greater number of questions and greater collaboration in asking questions, the Game Mode may have supported a more interactive mode of engagement than that of the Ask Mode. Additionally, our analysis of types of questions visitors asked using Scardamalia and Bereiter's (1992) taxonomy revealed that visitors generated and asked wonderment questions indicating that visitors are likely to ask some questions that would elicit constructive and interactive modes of engagement.

The ICAP framework can be a valuable tool for designing the cognitive aspects of exhibits. At the same time, our results reveal that the ICAP framework does not address motivation in informal learning settings. Simple motivational differences elicited by our Game Mode had significant effects on how visitors collaborated and utilized the app. Thus, future research should consider how the ICAP framework may inform the cognitive design of museum experiences while at the same time considering ways of addressing motivational aspects of informal learning.

Considering that the ICAP framework may help explain and predict informal learning, future studies should consider measuring learning in museums using ICAP as a framing model for designing museum informal learning experiences. Moreover, museums should design devices, exhibits, and experiences to encourage higher ICAP modes (Constructive, Interactive) or, at a minimum, improve exhibit design from engaging visitors in the passive mode to the active mode. Mobile apps may prove to be a relatively affordable way to do so when changes to exhibit hardscape are not possible.

Future studies of question-asking and data-gathering apps for visitors in venues such as museums and science centers should explore a more in-depth analysis of the relationship among gamified CSCL experiences, group interactions, and informal learning to elucidate and expand upon the results of this study. In terms of gamification, future studies may consider exploring the effect of different gamification strategies to clarify whether our app's low-attention strategy (e.g., badges and points) is indeed more effective than other higher-attention gamification features (e.g., animations and unlockable content). Revisiting Rummel's (2018) framework's target support dimension, future studies may explore the effect of multidimensional target support. For example, future research may consider targeting the social support dimension to facilitate group discussions, much like the role of the museum staff moderator in Gutwill and Allen's (2010) inquiry games, in addition to targeting the motivational support dimension via gamification. To elucidate Robert and Lyons' (2017) findings, future research may also consider varying conditions by the number of devices available in a party.

Additionally, future studies may want to consider investigating whether children and teenagers are more likely to ask questions to other group members when given the opportunity to input them into an app, as well as whether they are less likely to ask questions to other group members when not given a chance. Researchers may also consider investigating whether children and teenagers elicit help from family members in formulating meaningful questions. In addition, researchers may consider exploring why gamifying a museum question-asking app may elicit fewer off-topic questions. Finally, future studies should consider adding pre- and post-learning measures to further investigate gamified mobile devices' impact on informal learning and to verify the ICAP framework's predictions. These learning measures should be designed to assess deeper learning via inferential questions, considering the ICAP framework predicts deeper learning and not shallow recall.

The goal of our work was twofold, to see if gamification can increase user engagement with our museum question-asking app and to explore the role of the ICAP framework for evaluating the design of exhibits and exhibit technology. Though our in situ study had several limitations, our findings contribute to the body of CSCL literature and future research on informal learning and museum mobile app design by exploring the effect of gamification on visitor group interactions as well as exploring the utility of the ICAP framework for the design of museum exhibits and informal learning.

Funding This work was supported by the National Science Foundation's Directorate for Education and Human Resources [Grant Number 1438825].

References

Ahn, M. L., Cha, H. J., Hwang, Y. J., & Kim, H. J. (2011). An exploration of various evaluation methods to improve usability of museum mobile device. *Journal of the Ergonomics Society of Korea, 30*(6), 765–773. https://doi.org/10.5143/JESK.2011.30.6.765

Allen, S. (2002). Looking for learning in visitor talk: A methodological exploration. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.), *Learning conversations in museums* (1st ed., pp. 259–303). Hillsdale: Lawrence Erlbaum.

Allen, S., & Gutwill, J. P. (2009). Creating a program to deepen family inquiry at interactive science exhibits. *Curator, 52*(3), 290–306. https://doi.org/10.1111/j.2151-6952.2009.tb00352.x

Anderson, G. (2004). *Reinventing the museum: Historical and contemporary perspectives on the paradigm shift*. Lanham: AltaMira Press.

Arnab, S., Bhakta, R., Merry, S. K., Smith, M., Star, K., & Duncan, M. (2016, October). Competition and collaboration using a social and gamified online learning platform. In *10th European Conference on Games Based Learning: ECGBL* (Vol. 2016, pp. 19–27).

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

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children, 46*(2), 26.

Barriault, C. (1999). The science centre learning experience: A visitor-based framework. *The Informal Learning Review, 35*(1), 14–16.

Barriault, C., & Pearson, D. (2010). Assessing exhibits for learning in science centers: A practical tool. *Visitor Studies, 13*(1), 90–106. https://doi.org/10.1080/10645571003618824

Beasley, T., & Schumacker, R. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *The Journal of Experimental Education, 64*(1), 79–93. https://doi.org/10.1080/00220973.1995.9943797

Bickmore, T. W., Vardoulakis, L. M. P., & Schulman, D. (2013). Tinker: A relational agent museum guide. *Autonomous Agents and Multi-Agent Systems, 27*(2), 254–276. https://doi.org/10.1007/s10458-012-9216-7

Borun, M. (2002). Object-based learning and family groups. *Perspectives on object-centered learning in museums*, 245–260.

Bowman, C. D. D., Adkins, A., Owen, B. L., Rogers, K. J., Escalante, E., Bowman, J. D., et al. (2019). Differences in visitor characteristics and experiences on episodic free admission days. *Museum Management and Curatorship*. https://doi.org/10.1080/09647775.2019.1578992

Cabrera, J. S., Frutos, H. M., Stoica, A. G., Avouris, N., Dimitriadis, Y., Fiotakis, G., & Liveri, K. D. (2005, September). Mystery in the museum: collaborative learning activities using handheld devices. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 315–318). ACM. https://doi.org/10.1145/1085777.1085843

Cahill, C., Kuhn, A., Schmoll, S., Lo, W. T., McNally, B., & Quintana, C. (2011, June). Mobile learning in museums: how mobile supports for learning influence student behavior. In *proceedings of the 10th international conference on interaction design and children* (pp. 21–28). ACM Press. https://doi.org/10.1145/1999030.1999033

Chen, G., Xin, Y., & Chen, N. S. (2017). Informal learning in science museum: Development and evaluation of a mobile exhibit label system with iBeacon technology. *Educational Technology Research and Development, 65*(3), 719–741. https://doi.org/10.1007/s11423-016-9506-x

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Clark, D. B., & Sengupta, P. (2019). Reconceptualizing games for integrating computational thinking and science as practice: Collaborative agent-based disciplinarily-integrated games. *Interactive Learning Environments, 28*(3), 328–346. https://doi.org/10.1080/10494820.2019.1636071

Crowley, K., Callanan, M. A., Jipson, J. L., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. *Science Education, 85*(6), 712–732.

Davis, P., Horn, M., Block, F., Phillips, B., Evans, E. M., Diamond, J., & Shen, C. (2015). "Whoa! we're going deep in the trees!": Patterns of collaboration around an interactive information visualization exhibit. *International Journal of Computer-Supported Collaborative Learning, 10*, 53–76. https://doi.org/10.1007/s11412-015-9209-z

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). ACM. https://doi.org/10.1145/2181037.2181040

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

Dierking, L. D., & Falk, J. H. (1994). Family behavior and learning in informal science settings: A review of the research. *Science Education, 78*(1), 57–72. https://doi.org/10.1002/sce.3730780104

Dignan, A. (2011). *Game frame: Using games as a strategy for success*. New York: Free Press.

van Dijk, E. M. A G., Lingnau, A., & Kockelkorn, H. (2012). Measuring enjoyment of an interactive museum experience. In *Proceedings of the ACM International Conference on Multimodal Interaction, ICMI '12*, (pp. 249–256). ACM Press. https://doi.org/10.1145/2388676.2388728

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189–211). Elsevier.

Dini, R., Paternò, F., & Santoro, C. (2007, September). An environment to support multi-user interaction and cooperation for improving museum visits through games. In *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services* (pp. 515–521). https://doi.org/10.1145/1377999.1378062

Ellenbogen, K. M., Luke, J. J., & Dierking, L. D. (2007). Family learning in museums: Perspectives on a decade of research. *In principle, In Practice: Museums as Learning Institutions*, 17–30.

Falk, J. H., & Dierking, L. D. (1992). *The museum experience*. Kittery: Howells House.

Falk, J. H., Heimlich, J., & Bronnenkant, K. (2008). Using identity-related visit motivations as a tool for understanding adult zoo and aquarium visitors' meaning-making. *Curator: the Museum Journal, 51*(1), 55–79. https://doi.org/10.1111/j.2151-6952.2008.tb00294.x

Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., & Bonnett, V. (2009, November). Actions speak loudly with words: unpacking collaboration around the table. In *Proceedings of the ACM international conference on interactive tabletops and surfaces* (pp. 189–196).

García-Pérez, M. A., & Núñez-Antón, V. V. (2003). Cellwise residual analysis in two-way contingency tables. *Educational and Psychological Measurement, 63*(5), 825–839. https://doi.org/10.1177/0013164403251280

Gilbert, S. (2016, October). *Please turn on your phone in the museum.* The Atlantic. https://www.theatlantic.com/magazine/archive/2016/10/please-turn-on-your-phone-in-the-museum/497525/

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137. https://doi.org/10.3102/00028312031001104

Griffin, J. (1998). Learning science through practical experiences in museums. *International Journal of Science Education, 20*(6), 655–663. https://doi.org/10.1080/0950069980200604

Gutwill, J. P. (2005). Observing APE. In T. Humphrey & J. Gutwill (Eds.), *Fostering active prolonged engagement: The art of creating APE exhibits.* Walnut Creek: Left Coast Press.

Gutwill, J. P., & Allen, S. (2010). Facilitating family group inquiry at science museum exhibits. *Science Education, 94*(4), 710–742. https://doi.org/10.1002/sce.20387

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

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Heath, C., Lehn, D. V., & Osborne, J. (2005). Interaction and interactives: Collaboration and participation with computer-based exhibits. *Public Understanding of Science, 14*(1), 91–101. https://doi.org/10.1177/0963662505047343

Hein, G. (1996). Constructivist learning theory. In G. Durbin (Ed.), *Developing museum exhibitions for life-long learning* (pp. 30–34). The Stationery Office.

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

Honey, M. A., & Hilton, M. (2011). *Learning science through computer games and simulations*. National Academies Press.

Hope, T., Nakamura, Y., Takahashi, T., Nobayashi, A., Fukuoka, S., Hamasaki, M., & Nishimura, T. (2009, April). Familial collaborations in a museum. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1963–1972). https://doi.org/10.1145/1518701.1519000

Horn, M. S., Phillips, B. C., Evans, E. M., Block, F., Diamond, J., & Shen, C. (2016). Visualizing biological data in museums: Visitor learning with an interactive tree of life exhibit. *Journal of Research in Science Teaching, 53*(6), 895–918. https://doi.org/10.1002/tea.21318

Humphrey, T., & Gutwill, J. P. (Eds.). (2005). *Fostering active prolonged engagement: The art of creating APE exhibits*. Walnut Creek: Left Coast Press.

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational Psychology Review, 22*(2), 139–154. https://doi.org/10.1007/s10648-010-9131-x

Jewitt, C. (2012). Digital technologies in museums: New routes to engagement and participation. *Designs for Learning, 5*(1–2), 74–93. https://doi.org/10.2478/dfl-2014-0005

Jimenez, P. F., & Lyons, L. B. (2010, April). Studying different methods of providing input to collaborative interactive museum exhibit using mobile devices. In *2010 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education* (pp. 225–227). IEEE. https://doi.org/10.1109/WMUTE.2010.34

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education.* Pfeiffer.

Katz, J. E., Goldman, K. H., & Foutz, S. (2011). Mobile phones for informal science center learning: A socio-technical analysis. In J. E. Katz, W. LaBar, & E. Lynch (Eds.), *Technology and creativity: Social media, mobiles and museums* (1st ed., pp. 346–379). Edinburgh: MuseumsEtc.

Klopfer, E., Perry, J., Squire, K., Jan, M. F., & Steinkuehler, C. (2005). Mystery at the museum – A collaborative game for museum education. In *Proceedings of the International Conference on Computer Supported Collaborative Learning 2005*, (pp. 316–320). https://repository.isls.org//handle/1/3697

Laursen, D. (2012). Co-participation among school children around a computer-based exhibit. *Social Studies of Science, 43*(1), 97–117. https://doi.org/10.1177/0306312712455114

Livingstone, D. W. (1999). Lifelong learning and underemployment in the knowledge society: A North American perspective. *Comparative Education*. https://doi.org/10.1080/03050069927955

Lohr, S. (2014, October 23). *Museums morph digitally.* The New York Times. https://www.nytimes.com/2014/10/26/arts/artsspecial/ the-met-and-other-museums-adapt-to-the-digital-age.html

Lyons, L., Becker, D., & Roberts, J. (2010). Analyzing the affordances of mobile technologies for informal science learning. *Museums & Social Issues, 5*(1), 89–104. https://doi.org/10.1179/msi.2010.5.1.87

Lyons, L. (2008). *Designing for the Cooperative Use of Multi-user. Multi-device Museum Exhibits* [Doctoral Dissertation, University of Michigan].

Lyons, L. (2009). Designing opportunistic user interfaces to support a collaborative museum exhibit. *In Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning - Volume 1* (pp. 375–384) International Society of the Learning Sciences. https://doi.org/10.3115/1600053.1600110

Manabe, M., & Lydens, L. (2007). Making connections: Using mobile phones as a museum tool. *Journal of Museum Education, 32*(1), 27–34. https://doi.org/10.1179/jme.2007.32.1.27

Marty, P. F., & Jones, K. B. (2012). Museum informatics: People, information, and technology in museums. *Routledge*. https://doi.org/10.4324/9780203939147

McLean, K. (1999). Museum exhibitions and the dynamics of dialogue. *Daedalus*, *128*(3), 83–107. https://www.jstor.org/stable/20027568

McManus, P. (1994). Families in museums. In L. Miks & A. Zavala (Eds.), *Towards the museum of the future* (1st ed., pp. 81–118). Routledge.

Meisner, R., vom Lehn, D., Heath, C., Burch, A., Gammon, B., & Reisman, M. (2007). Exhibiting performance: Co-participation in science centres and museums. *International Journal of Science Education, 29*(12), 1531–1555. https://doi.org/10.1080/09500690701494050

Melero, J., Hernández-Leo, D., & Manatunga, K. (2015). Group-based mobile learning: Do group size and sharing mobile devices matter? *Computers in Human Behavior, 44*, 377–385. https://doi.org/10.1016/j.chb.2014.11.078

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*(3), 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., et al. (2020). Ask Dr Discovery: The impact of a casual mobile game on visitor engagement with science museum content. *Educational Technology Research and Development, 68*(1), 345–362. https://doi.org/10.1007/s11423-019-09696-x

Parry, R., & Ortiz-Williams, M. (2007). How shall we label our exhibit today? Applying the principles of on-line publishing to an on-site exhibition. In J. Trant & D. Bearman (Eds.), *Museums and the Web 2007: Proceedings.* Archives & Museum Informatics.

Parry, R., & Sawyer, A. (2005). Space and the machine: Adaptive museums, pervasive technology and the new gallery environment. In S. Macleod (Ed.), *Reshaping museum space* (1st ed., pp. 39–52). Routledge.

Pérez Cortés, L. E., Ha, J., Su, M., Nelson, B. C., Bowman, C., & Bowman, J. D. (2021). *Gleaning detailed museum visitor behavior via basic mobile app.* [Manuscript submitted for publication]. Mary Lou Fulton Teachers College, Arizona State University.

Pew Research Center (2019, June 12). *Mobile fact sheet.* https://www.pewresearch.org/internet/fact-sheet/mobile/

Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2019). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning*. Cambridge: MIT Press.

Poce, A., Re, M. R., Amenduni, F., De Medio, C., & Valente, M. (2019). Developing a web App to provide personalized feedback for museum visitors: a pilot research project. *Form@re - Open Journal Per La Formazione in Rete*, *19*(3), 48–59. https://doi.org/10.13128/form-7703

Proctor, N. (2015). Mobile in museums: From interpretation to conversation. *The International Handbooks of Museum Studies, 4*(22), 499–525. https://doi.org/10.1002/9781118829059.wbihms322

Rahm, J. (2004). Multiple modes of meaning-making in a science center. *Science Education, 88*(2), 223–247. https://doi.org/10.1002/sce.10117

Ramey-Gassert, L., Walberg, H. J., III., & Walberg, H. J. (1994). Reexamining connections: Museums as science learning environments. *Science Education, 78*(4), 345–363. https://doi.org/10.1002/sce.3730780403

Rigby, S., & Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound.* Praeger.

Roberts, J., & Lyons, L. (2017). The value of learning talk: Applying a novel dialogue scoring method to inform interaction design in an open-ended, embodied museum exhibit. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 343–376. https://doi.org/10.1007/s11412-017-9262-x

Røtne O. & Kaptelinin V. (2013) Design choices and museum experience: A design-based study of a mobile museum app. In Collazos C., Liborio A., & Rusu C. (Eds.), *Human Computer Interaction.* CLIHC 2013. Lecture Notes in Computer Science, vol 8278. Springer. https://doi.org/10.1007/978-3-319-03068-5_3

Rounds, J. (2006). Doing identity work in museums. *Curator the Museum Journal, 49*(2), 133–150. https://doi.org/10.1111/j.2151-6952.2006.tb00208.x

Rummel, N. (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 123–129. https://doi.org/10.1007/s11412-018-9273-2

Russell, T. (1995). The enquiring visitor: Useable learning theory for museum contexts. *Journal of Education in Museums, 16*, 19–21.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*, 77–112. https://doi.org/10.1007/s10648-019-09498-w
Scardamalia, M., & Bereiter, C. (1999). Schools as knowledge-building organizations. *Today's Children, Tomorrow's Society: The Developmental Health and Wealth of Nations*, 274–289.
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
Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006
Simon, N. (2010). *The participatory museum.* Museum 2.0.
Stahl, G. (2007). Meaning Making in CSCL: Conditions and Preconditions for Cognitive Processes by Groups. In Chinn, C. A., Erkens, G., & Puntambekar, S. (Eds.), *The Computer Supported Collaborative Learning (CSCL) Conference 2007, Volume 8, Part 2* (pp. 651–660). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2007.651
Stogner, M. B. (2009). The media-enhanced museum experience: Debating the use of media technology in cultural exhibitions. *Curator: the Museum Journal, 52*(4), 385–397. https://doi.org/10.1111/j.2151-6952.2009.tb00360.x
Su, M., Pérez Cortés, L. E., Ha, J., Nelson, B. C., Bowman, C., & Bowman, J. D. (2021) *Understanding visitors' question-asking quality in museums through a mobile app.* [Manuscript in preparation]. Mary Lou Fulton Teachers College, Arizona State University.
Tan, M., & Hew, K. F. (2016). Incorporating meaningful gamification in a blended learning research methods class: Examining student learning engagement and affective outcomes. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.2232
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
Tscholl, M., & Lindgren, R. (2016). Designing for learning conversations: How parents support children's science learning within an immersive simulation. *Science Education, 100*(5), 877–902. https://doi.org/10.1002/sce.21228
Valente-Marques, D. (2017). *The visitor experience using augmented reality on mobile devices in museum exhibitions* [Doctoral dissertation, Universidade do Porto]. ProQuest Dissertations Publishing.
Vermeeren, A. P., Calvi, L., Sabiescu, A., Trocchianesi, R., Stuedahl, D., Giaccardi, E., & Radice, S. (2018). Future museum experience design: crowds, ecosystems and novel technologies. In *Museum Experience Design* (pp. 1–16). Springer, Cham.
Yatani, K., Onuma, M., Sugimoto, M., & Kusunoki, F. (2004). Musex: A system for supporting children's collaborative learning in a museum with PDAs. *Systems and Computers in Japan, 35*(14), 54–63. https://doi.org/10.1002/scj.10696
Yiannoutsou, N., Papadimitriou, I., Komis, V., & Avouris, N. (2009, June). " Playing with" museum exhibits: designing educational games mediated by mobile technology. In *Proceedings of the 8th International Conference on Interaction Design and Children* (pp. 230–233). ACM. https://doi.org/10.1145/1551788.1551837
Yoon, S., Elinich, K., Wang, J., Steinmeier, C., & Tucker, S. (2012). Using augmented reality and knowledge-building scaffolds to improve learning in a science museum. *International Journal of Computer-Supported Collaborative Learning, 7*(4), 519–541. https://doi.org/10.1007/s11412-012-9156-x
Zimmerman, H. T., Reeve, S., & Bell, P. (2008). Distributed expertise in a science center: Social and intellectual role-taking by families. *Journal of Museum Education, 33*(2), 143–152. https://doi.org/10.1080/10598650.2008.11510595