



Gleaning museum visitors' behaviors by analyzing questions asked in a mobile app

Luis E. Pérez Cortés¹ · Jesse Ha² · Man Su³ · Brian Nelson³ · Catherine Bowman³ · Judd Bowman³

Accepted: 8 February 2023 / Published online: 21 February 2023
© Association for Educational Communications and Technology 2023

Abstract

This study explores the feasibility of forming detailed inferences about museum visitor behavior based on analysis of data collected via *Dr. Discovery*—a mobile question-and-answer app. We analyzed 5656 questions asked by 795 visitor groups recorded by *Dr. Discovery* across two museums in the American Southwest. Analysis of this data supported the act of intuiting visitor movement through museum exhibit halls without the use of costly tracking or location technology by leveraging question keyword content, knowledge of exhibit hall layout, and question timestamp information. Additionally, data on question topic frequency enabled us to infer visitor engagement levels with specific exhibit hall content. We conclude that analysis of seemingly limited app-based data carries implications for the practice of museum evaluation since evaluators can gain evidence-based insight into visitor behaviors as well as illustrate helpful and promising technology-supported alternatives for conducting affordable, dependable, and scalable evaluations.

Keywords App-based data collection · Data-driven behavioral analysis · Museum evaluation

Introduction

The word “museum” may, for some, conjure images of a building-sized shelf where artistic, scientific, or historical products are displayed for passive and silent admiration. This notion of museums is known as *didactic* where “the curator teaches, the visitors learn... unquestioningly, in the order and manner intended” (Black, 2005, p. 130). Such didactic views contrast with notions of museums as interactive, informal, and important learning spaces as well as heritage preservation powerhouses. For instance, museums are interactive learning spaces where visitors can engage with curated content and expert staff in various

✉ Luis E. Pérez Cortés
LEP101@pitt.edu

¹ University of Pittsburgh, Murdoch Building, 3420 Forbes Ave., Pittsburgh, PA 15260, USA

² University of California, Los Angeles, USA

³ Arizona State University, Tempe, USA

ways (Black, 2015; Koreneva, 2015) and have long been considered centers for informal learning (e.g., Durbin, 1996; Hein, 1998; Hooper-Greenhill, 1999) that leverage motivating, spontaneous, learner-centered, open-ended, and non-assessed activities (Eshach, 2007; Vadeboncoeur, 2006; Wellington, 1990). Furthermore, the American Alliance of Museums (2018) reports that 97% of Americans believe museums represent an important educational asset for their communities. In addition to varied learning purposes, the National Museum Director's Council (2015) argues that museums preserve, promote, and protect a nation's heritage, which is fundamental to maintaining a healthy and prosperous civil society. Thus, museums represent popular and growing places where a wide range of people with varying levels of interest can *experience* them—not just admire what is within them in didactic ways—to enrich their knowledge of a topic as well as maintain a unified society (Falk & Dierking, 2013). For these reasons, understanding how and to what extent visitors engage with displays in museums is imperative for ensuring the continued high-quality convergence of public interest with the curated knowledge and learning experiences available in these spaces.

To this end, practitioners and researchers have leveraged diverse approaches for enhancing and exploring museum visitor experiences to meet varied goals associated with research, documentation, education, communication, outreach, and evaluation. These have ranged from relatively low technology-dependent approaches such as conducting visitor observations or interviews (East of England Museum Hub, 2008) to relatively higher technology-dependent approaches that use mobile technologies (Chivarov et al., 2013) to, among other purposes, determine visitor positioning for enhancing museum experiences (Rubino et al., 2013) and conducting museum evaluation. Unfortunately, high-quality museum evaluation can be expensive to conduct.

Making changes to the *holistic museum experience*—the physical displays as well as more flexible elements such as docents, multimedia, public events, temporary signage, and webpages—is often prohibitively costly to undertake (Adams, 2012). To evaluate these elements, timely and accessible insight into the minds and behavior of visitors becomes invaluable actionable information that may justify associated costs. To attain this insight, museum evaluators have long pursued *naturalistic evaluation*—a form of evaluation that collects “descriptive data in situ, with the evaluators themselves serving as the inquiry instrument” (Bonner, 1990, p. 211). This approach, however, possesses important practical limitations such as being comparatively more extensive and potentially yielding edited or filtered responses (Leinhardt & Knutson, 2004). However, visitor-facing apps provide a cost-effective yet extensible and effective data collection tool to approximate elements of naturalistic museum evaluation. An additional strength of app-based data collection, compared to traditional (or formal) museum visitor evaluation methods such as interviews, surveys, or in-person timing and tracking studies, is that apps can unobtrusively and anonymously capture all visitors' in-app interactions. Such ongoing in-app interactions contrast with the potentially filtered responses that visitors might provide to an interviewer or the self-edited comments by visitors wearing a microphone (Leinhardt & Knutson, 2004). Additionally, technology-dependent approaches have been so increasingly favored to meet multiple goals—including evaluation—that “it is inconceivable to imagine a museum or heritage site not making any use of technology” (Damala et al., 2019, p. 2).

Our project (Bowman et al., 2019; Ha et al., 2021; Nelson et al., 2017, 2020), is a National Science Foundation-funded study (NSF#1438825) addressing the need for affordable, ongoing, large-scale museum evaluation while investigating innovative ways for enabling museum visitors to engage deeply with museum content. For this project, we developed a mobile app called Dr. Discovery through which visitors can ask—by either typing

or speaking—their questions and receive vetted answers about museum content. The app and its backend data visualization portal function as a platform for research, museum evaluation, STEM informal education, and data-driven decision-making by museum personnel.

Our aim in the current study is to draw implications for the practice of museum evaluation by investigating the viability of leveraging relatively “simple” and easily collected data through our Dr. Discovery app to make more complex inferences about visitor behavior. To achieve this aim, we analyzed question data collected by our project app from 795 visitor groups across two partner museums: a natural history museum (Museum A) and a science center (Museum B). This data includes the questions' content, frequency, and timestamped information of 5656 individual queries. Using these three points of visitors' question data, we explore the feasibility of intuiting visitor movement through museum exhibit halls without the use of potentially more expensive tracking or location technologies such as global positioning systems (GPS) or near-field communications (NFC). We also leverage this data to preliminarily infer visitors' overall levels of engagement, confusion, and/or interest in displays inside museum exhibit halls and investigate how such data can be productively leveraged for museum evaluation purposes. Thus, this work was guided by the following guiding questions:

- How can data on question keyword content, knowledge of exhibit hall layout, and question timestamp information be leveraged to intuit visitor behaviors across different museum layouts?
- To what extent can such data be leveraged to draw inferences about visitors' engagement with specific exhibit hall content in ways that are relevant for museum evaluation?

Background and related work

Understanding museum visitors' behaviors and motivations is a crucial concern in museum evaluation and is critical for maintaining high-quality visitor experiences and learning. Efforts to study museum visitors have adopted multiple forms over the years, with Bitgood and Shettel (1996) noting five major areas of study in their overview: (1) Audience Research and Development, (2) Exhibit Design and Development, (3) Program Design and Development, (4) General Facility Design, and (5) Visitor Services. Taken together, these areas point to the importance of a holistic view regarding museum evaluation. An important (see Bitgood, 2006) and longstanding (e.g., Porter, 1938) component in this holistic view of museum evaluation is the timely and accessible insight into visitor behaviors because these help evaluators directly infer the effects of any changes to the museum contexts. The increasing availability of diverse and innovative digital technologies has expanded the opportunities to attain such access to visitor behaviors in museums. Some recent efforts include that of Ch'ng et al. (2019) who have explored how users interact with different technologies across multiple museum contexts in China; De Angeli et al. (2020) who have explored how the emoji can help document visitors' emotional experiences that go beyond the happy-or-not dichotomy; and Emerson et al. (2020) who have deployed computational models and multimodal learning analytics of museum visitor behaviors to better understand their engagement. Taken together, these studies help illustrate that the opportunity to use innovative digital technologies for visitor behavior-based evaluation is greater than ever before.

Due, in part, to these technology-based opportunities, larger museums such as the American Museum of Natural History (n.d.), can regularly conduct evaluation of educational programs and exhibitions, including temporary exhibits. Unfortunately, at under-resourced museums, there is rarely staff available for carrying out extensive evaluations nor funds for hiring external evaluators. Despite evaluation being perceived as a core of change in institutions by museum educators (Adams, 2012), the Smithsonian Institution (2004) found that only 13% of museums reported having staff that spends all or part of their time on evaluations. This is because, as Adams indicates, “evaluation is often seen as a nice thing to do but not a necessary activity, especially in, but not always dependent upon, tighter economic times” (p. 28). To increase affordable evaluation and improve visitor learning, many museums have adopted mobile experiences to enhance and extend their on-site and off-site presence, with the goal of “improving the ability to measure impact using new digital technologies” (Johnson et al., 2011, p. 6).

Cellphone and smartphone ownership have both experienced consistent growth. Data from the Pew Research Center reports that 97% of Americans owned a cellphone of some kind in 2021, which has increased from 62% cellphone ownership in 2002. Similarly, Pew reports that 85% of Americans owned a *smartphone* specifically in 2021, an increase from 35% smartphone ownership in 2011 (Pew, 2021). Furthermore, data from our study suggests that smartphone ownership among museum visitors may be greater still. During the data collection period for this study (2016–2017), Pew reported that 77% of Americans owned a smartphone, which contrasts with the 97% of the roughly 2300 museum visitors who participated in our study that reported owning a smartphone.

Determining how digital technologies can be employed to improve museum goals, such as enhancing visitor experiences, is a necessity at all levels of museum education (Freeman, et al., 2016). For decades, museums have attempted to improve, bringing more interest and engagement to their exhibits and displays (Anderson, 2012; Black, 2012). These attempts have ranged from purposely designing heuristic tools such as the Ideas-People-Objects-Physical (IPOP) framework (Beghetto, 2014; Pekarik et al., 2014) to providing free admission opportunities (Bowman et al., 2019). Due, in part, to their early promise (Economou & Meintani, 2011), another popular attempt has been the design of digital apps that leverage the growing presence of portable devices across the U.S.

Ever since the Pew Research Center began tracking smartphone ownership in 2011, mobile apps have been deemed as “the most relevant features of mobiles for museums” (Johnson et al., 2011, p. 7). As a result, there is a growing number of mobile apps developed for varied uses within museums. A non-exhaustive list of these includes the American Museum of Natural History’s *Explorer*, the Brooklyn Museum’s *Ask*, the Cleveland Museum’s *ArtLens*; the Children’s Museum of Houston’s *More*; the Smithsonian’s *Infinity of Nations*, and the Museum of Natural History’s *Skin & Bones*. Each of these apps represents a distinct attempt to improve upon the museum visitor’s experience employing mobile technology and has been explored by scholars and researchers (e.g., Bickmore et al., 2013; Katz et al., 2011; Lane et al., 2013; Proctor, 2015; Røtne et al., 2013; Valente-Marques, 2017). The push to infuse museum experiences with some form of mobile technology is so prevalent that, for over a decade, few museums have been without some degree of the interactional mobile-based app (Marty & Jones, 2012). Often, these mobile-based apps 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). These uses of mobile apps in museums not only serve to further demonstrate the ubiquity of mobile devices, but also reflect 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. 285, emphasis added).

Practitioners and researchers have leveraged the multiple tools made available through mobile technologies. These include Quick Response (QR) codes (Chivarov et al., 2013), Wi-Fi (Ruíz et al., 2011), Radio-frequency Identification (RFID) (Hsu & Liao, 2011; Huang et al., 2011), Near Field Communication (NFC) (Blöckner et al., 2009), augmented reality (AR) (Carci et al., 2019; Hammady et al., 2018), databases (Cameron & Robinson, 2007), museum websites (Chong & Smith, 2017), digital signage systems (Braehmer et al., 2019), and digital interactive exhibits (Stratton et al., 2017) among others. Each has been used to meet varied goals associated with research, documentation, learning, communication, outreach, and evaluation.

For the specific purpose of museum evaluation, much research has focused on the ability to accurately and consistently track visitor movements. This is because museums are appealing and challenging environments for indoor positioning research (Kuflik et al., 2011) and, as Lanir et al. (2017) noted, “tracking of visitors can provide objective information about the way visitors interact with the museum space and content, providing valuable information about exhibition placements and design” (p. 313). To this end, a popular approach has focused on Indoor Positioning Systems (IPS) that use beacon-based technology’s wireless signals (e.g. Wi-Fi, RFID, Bluetooth) to track museum visitors (e.g., Handojo et al., 2018, 2020; Rubino et al., 2013; Spachos & Platniotis, 2020).

However, technologies such as GPS, RFID, and NFC all have their unique set of advantages and disadvantages for analyzing visitor circulation. For example, GPS is a widely available feature of smartphones, but it may not function properly in indoor locations, while RFID and NFC function well indoors but they can only cover short ranges (Ng, 2015). The employment of advanced sensing technologies such as GPS or NFC might present greater accuracy when visualizing and understanding visitor movement patterns, but they also present increases in costs and higher demands on maintenance, as well as practical concerns such as having reliable signals and potentially ensuing privacy issues.

Our project app, Dr. Discovery, was designed to collect and visualize data from museum visitors to help museum staff engage in evaluation without the need for highly specialized positioning software or technical knowledge. Using log files has long been identified as a powerful research tool in e-learning, virtual environments, and computerized agent research (e.g., Cockburn & McKenzie, 2001; Ingram & Northcote, 1999). The growing proliferation of smartphones (Pew, 2021) and similar portable devices allows informal learning settings such as museums to leverage the data made available with these devices to glean their visitors’ understanding and their experiences during the visit. Log files from mobile apps can provide snapshots and longitudinal data on visitors’ thinking, use, and engagement with displays in museum exhibit halls. Flexible and affordable apps providing such data could be useful to museums of all sizes and resources. Such apps could be used to supplement museum evaluation efforts in instances where they are unable to engage evaluation staff. Our Dr. Discovery app makes this needed evaluation data available via log files that contain records of visitor interactions and inquiries as well as the path and evolution of those inquiries.

Methods

App description

Our project used a mobile app called Dr. Discovery with two parts: (1) a front-end Question and Answer (Q&A) interface through which visitors can ask—by either typing or speaking—their questions and receive vetted answers about museum content and (2) a back-end analytics portal that visualizes recorded visitor interactions. As part of the umbrella project, we developed two modes of the app: Ask and Game (see Fig. 1). The baseline treatment for Dr. Discovery is Ask Mode. It presents a Q&A interface through which museum visitors can input questions about displays in a museum exhibit hall (or any science-related topic). Game Mode incorporates game elements such as challenges (e.g., “Help the doctor return to base”), points, levels, and badges by adding them to the baseline Q&A mechanic to create a gamified question-asking experience for users. The app is connected to a curated database of more than 13,500 verified questions and 5000 answers (many questions have the same answer) developed and related to the museum content by our project’s undergraduate students and museum partners.

Data sources

Our project app recorded a range of visitor information but in this study, we focus on visitors’ question-asking data as gleaned through keyword content, frequency, and time-stamped records with their logged entry time into the museum’s exhibit hall. We examine data collected from June 2016 to January 2017 at both of our partner museum sites



Fig. 1 Dr. Discovery’s Ask Mode [left] and Game Mode [right]

in the American Southwest: a natural history museum (Museum A) and a science center (Museum B). Both museums were partners of the larger NSF study this paper reports on, with staff members and volunteers forming part of the iterative development cycles by providing feedback and ideas on prototypes of our project app so that it best functioned as a platform for STEM evaluation, research, and data-driven decision-making by museum staff (Nelson et al., 2017). Both museums were chosen as partners, in part, because they serve large and diverse populations. For example, 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.

Across both partner sites, 1296 visitor groups participated during the study's 7-month time period. A visitor group consisted of at least one adult, aged 18 or older. We assigned these visitor groups across three conditions, Ask (387), Game (408), and Control (501). In our current analyses, however, we focus on the combined data from the two treatment groups (Ask and Game) and exclude the control groups from our analyses because these did not use the Dr. Discovery app and thus did not generate question data. After accounting for this exclusion of control participants, our total was 795 visitor groups, distributed as follows: Museum A = 332 and Museum B = 463. These groups generated 5656 individual queries. It is also important to note that we have analyzed distinctions between the use of the two app modes (Game and Ask) elsewhere (Nelson et al., 2020). For this reason, we do not focus on drawing comparisons between these conditions in this paper and instead focus on our question-asking data across both modes.

Data collection

At both museums, we conducted implementations using Ask Mode, Game Mode, and a Control condition on an alternating schedule such that, on a given day, all participants were assigned to one of the three conditions. On days when Ask or Game Mode were used, the research team explained our project app's purpose and invited participants to use the app while exploring a specifically targeted exhibit hall at either museum site. On control days, the research team explained that the museum was seeking feedback about the particular exhibit hall and invited visitors to complete surveys about their visit. In all conditions, participants explored the exhibit hall at a pace and order of their choosing. In the treatment conditions, participants decided when and where to use the project app while in the exhibit hall. Their in-app interactions were recorded in the online database. All groups completed a time- and device-stamped demographic and background survey before participation as well as completing a time- and device-stamped feedback survey after participation, which allowed the two surveys to be matched. For all groups, the surveys were identical except for the omission of app-related questions in the surveys for the control participants.

Data analysis

We analyzed the total exhibit hall-related questions asked by visitors and explored the extent to which the content of these questions, coupled with timestamps for when they were asked during the visit, allows us to interpret visitor behavior given our knowledge of the exhibit halls' layouts. The intent was to determine whether we could intuit visitors' movement patterns through the exhibit hall as well as make inferences about their engagement with displays in the exhibit. To make these interpretations, we first conducted a computer-assisted word count analysis in which we identified the frequency of words and short

Table 1 Gender distribution of primary participants

Gender	Museum A (%)	Museum B
Female	61	63%
Male	35	37%
Gender not disclosed	4	n/a

Table 2 Age distribution of primary participants

Age range	Museum A (%)	Museum B (%)
18–25	12	14
26–35	31	32
36–45	33	33
46–55	11	11
56+	10	7
Did not report age	3	3

phrases that appeared in questions asked to our project app. We then identified those that were relevant to the exhibit hall's content. For example, at Museum A the word “is” was the most common word of all, but “earth” was the most common word that is relevant to the exhibit's content. Using the generated word list, we identified a set of keywords that frequently appeared in questions asked; these words were associated with specific topics and content located at different dedicated sections of each exhibit hall. We plotted how often visitors asked questions related to these keywords and correlated that data to their time-in-exhibit on a normalized scale (i.e., rescaling the time to be between 0 and 1).

Results

Visitor demographics

Our pre-survey asked users to self-report a range of demographic information about their visitor groups. In this section, we offer the results of this pre-survey.

General group composition

Across both museum sites, the average group contained nearly four people, with more women than men, and the most frequent underage member of a group was in the “child” age range (5 to 12 years old).

Age and gender distributions

Across both sites, nearly two-thirds of those who answered the survey were female (see Table 1).

Similarly, nearly two-thirds of the primary participants reported their age range between 26 and 45 (see Table 2).

Race/ethnicity

Across both sites, nearly two-thirds of primary participants identified as White. The second largest group was Hispanic/Latinx and the rest of the racial and ethnic denominations lingered around five percent or less each (see Table 3).

Question data

We collected and analyzed data about the questions asked while using the Dr. Discovery app. In this section, we offer the results of the question data.

Number of questions asked and time in exhibit

A total of 5656 questions were asked across both of our project app's treatment modes, averaging 6.8 questions amongst all visitor groups. At Museum A, there were 1693 questions asked, averaging 5.1 questions per group. At Museum B, there were 3963 questions asked, averaging 8.6 questions per group. Across both sites, Game Mode participants asked significantly more questions on average than Ask Mode participants, a finding we discuss elsewhere (Nelson et al., 2020).

At Museum A, visitors spent an average of 10 min in the exhibit hall, where the minimum amount of time spent was 1 min and the maximum was 38 min. At Museum B, visitors spent an average of 17 min in the exhibit hall, where the minimum amount of time spent was 2 min and the maximum was 64 min.

Question keywords content

We identified 34 total keywords in the questions asked at Museum A and 17 total keywords at Museum B (see Table 4). These are listed in alphabetical order, with keywords found at both sites listed in bold print.

We took the number of questions containing the keywords and compared them to their timestamp data, plotting them on a normalized scale of visitors' time-in-exhibit. Based on in-app data alone, we were able to interpret visitor movement through an exhibit hall with varying consistency by the site. We show 3 of the 34 topics identified for Museum A ("Snowball Earth", "Meteors", and "Quartz") as an illustrative, rather than exhaustive,

Table 3 Race/ethnicity of primary participants

Race/ethnicity	Museum A (%)	Museum B (%)
White	66.0	62.0
Hispanic/Latinx	16.7	16.1
Black	1.9	4.7
Asian	4.1	6.4
Native American	2.3	1.7
Pacific Islander	0.4	0.3
Two or more races	5.1	5.5
Did not report race	2.9	3.3

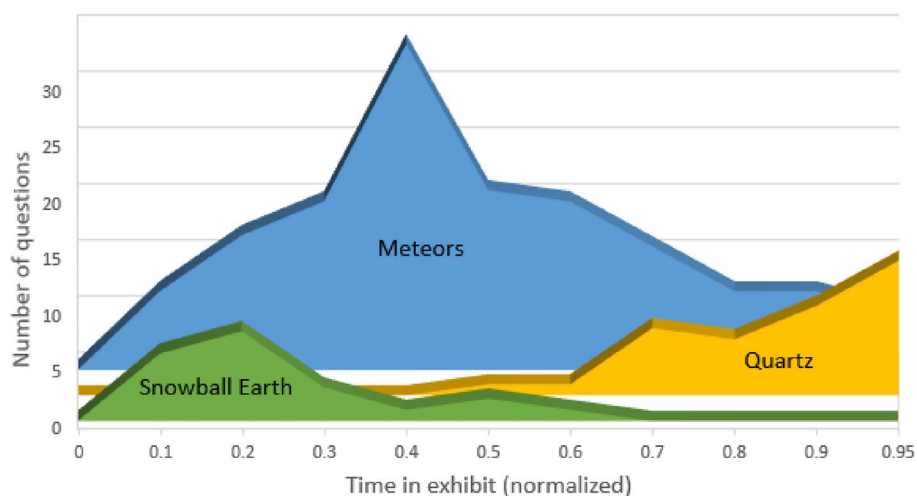
Table 4 Keywords at each museum site

Museum A		Museum B
Ape	Meteor Crater	Arizona
Ardi	Million	Carbon
Arizona	Minerals	Clouds
Astronomy	Moon	Dam
Cave	Painted Desert	Earthquake
Comet	Planets	Erosion
Copper	Pyrite	Grand Canyon
Crater	Quartz	Planets
Desert	Rock	Season
Dinosaur	Rushmore	Size
Galaxy	Selenite	Temperature
Geode	Size	Time
Gold	Snowball	Volcano
Grand Canyon	Temperature	Vortex
Iron	Time	Water
Life	Trex	Waves
Meteor	Weather	Weather

example of how we leveraged the data collected by our project app to intuit visitor's general movement patterns through the exhibit (see Fig. 2).

Plotted keywords on a normalized time scale

By looking at question content and related timestamp data, we can observe that visitors in Museum A engaged with exhibit displays related to “Snowball Earth” mainly during

**Fig. 2** Three illustrative topics and normalized time at Museum A

the first half of their visit while engaging with displays with content related to “Quartz” mainly during the last half of their visit. Additionally, visitors asked about “Meteors” throughout their entire time (and movement) through the exhibit hall. We arrived at this inference by observing that questions containing keywords associated with topics of Snowball Earth and Quartz were almost exclusively asked during their respective halves of normalized time.

At Museum A, we implemented our project app in an exhibit hall called “Origins” that occupies a long hallway that visitors predominantly travel through in one direction. The exhibit’s displays consist mainly of photos and physical objects of various topics related to the geological history of the Earth, space, planets, minerals, and rocks. The steady progression in the questions’ keyword content, relative frequency, and time-in-exhibit illustrated in Fig. 2 similarly occurred across the rest of the topics we identified through an analysis of word frequency and topic-display content matching. To illustrate this point further, we plotted when visitors asked questions on normalized time and compared it to normalized question number scales (0–1) at Museum A (see Fig. 3). A non-normalized figure would display larger and smaller “peaks” for each topic.

At Museum B, our project app data yielded question-asking patterns that were less clearly emphasized at specific timestamps than was evident at Museum A. For example, although the topics of “vortex” and “dam” had particular spikes (moments when participants engaged most with them), on average, visitors consistently engaged with these topics throughout their entire normalized time in the exhibit. This comparably steadier pattern is observable across the 17 topics we identified for the exhibit hall of Museum B (see Fig. 4).

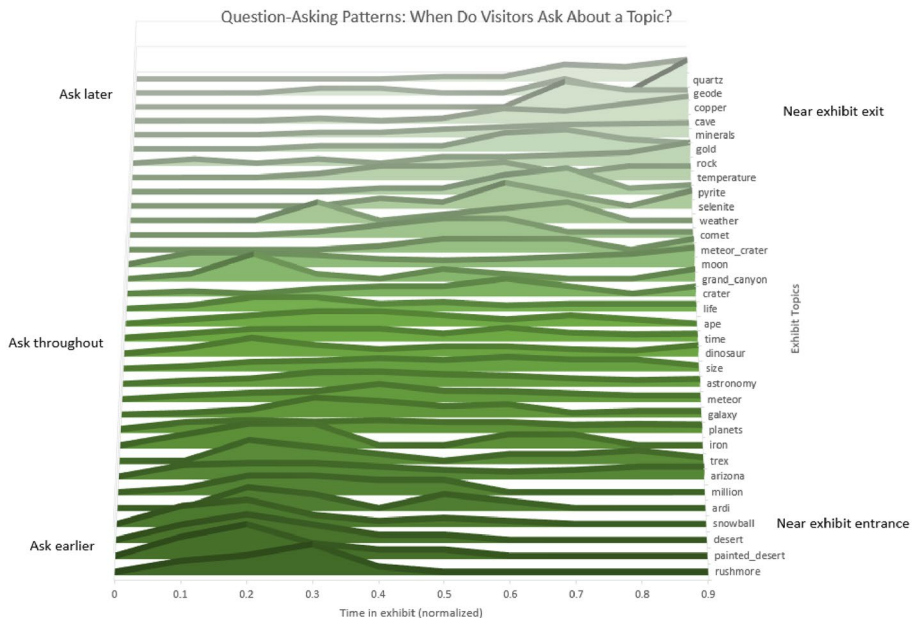


Fig. 3 Normalized question number and time at Museum A

Question-Asking Patterning: When Do Visitors Ask About a Topic?

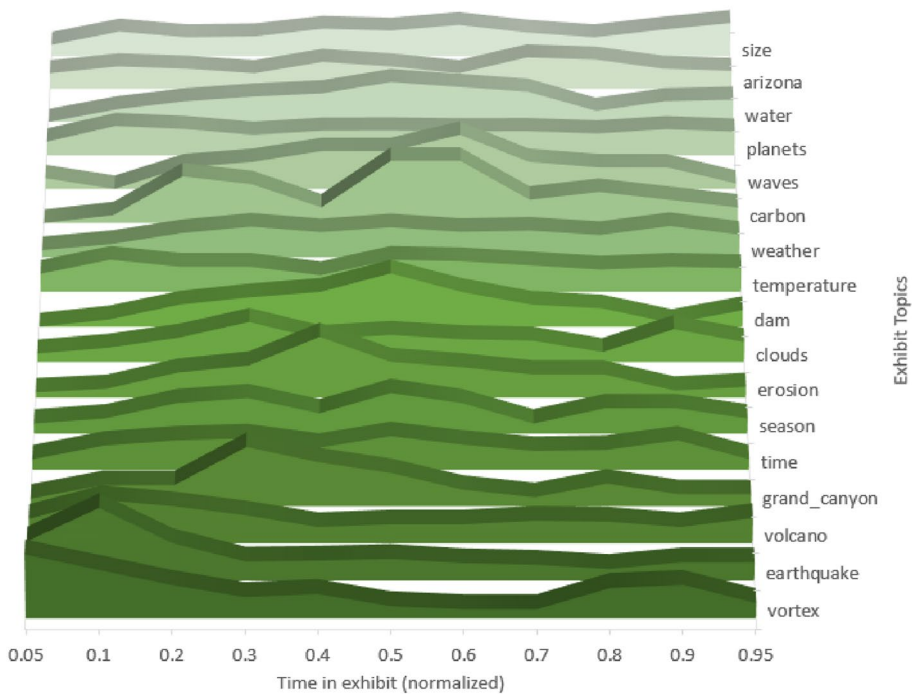


Fig. 4 Normalized question number and time at Museum B

Discussion and implications for practice

We leverage our results to develop two main points of discussion and implications for practice: (1) the varying degree to which we can intuit visitor paths and (2) inferring an exhibit hall's "success". We also discuss our findings in relation to the literature we reviewed and affordable technology-supported museum evaluation.

Intuiting visitor paths

By examining the back-end portal of Dr. Discovery, we were able to extract keyword content of questions asked, normalize question number and time in exhibit, and display this information in graphical format. Doing so revealed that our app data could help intuit visitor behaviors through an exhibit hall. However, we were able to intuit visitor's movements at each museum with differing levels of consistency by using the app data in isolation, as is evident when comparing Fig. 3 with Fig. 4, which show a more clearly defined tendency for a path at Museum A than at Museum B.

Although there could be multiple combinations of factors that explain the variations in our ability to intuit visitor paths across museums, such as group interaction effects (Ha et al., 2021) or instrumental factors such as whether participants were in the Game or Ask

condition (Nelson et al., 2020), we focus here on the differences in the exhibit halls' physical layouts as a way to draw more immediate implications for practice. There is a comparably more uniform progression from one keyword topic to another in the question-asking data throughout visitors' time in Museum A (Fig. 3) as opposed to Museum B (Fig. 4). This may reflect differences in linearity of the physical layouts between each museum's exhibits.

Museum A

We implemented our project app at Museum A in a linear exhibit hall called “Origins” that occupies a long corridor and slopes downward from the museum's entrance (see Fig. 5).

Near the entry to the exhibit hallway that is closest to the front of the museum (which is near the check-in table for tablets containing our project app), there is an eye-catching three-dimensional picture on the wall of “Snowball Earth” accompanied by scarce interpretive text. As visitors move down the hallway, they learn more about the planets and Earth, encounter physical meteorite specimens, and then end in a series of cases containing rocks and minerals. Consequently, the shift from asking questions about “Snowball Earth” early in the record of data to questions about “Quartz” later in the data matches the physical location of those objects in the exhibit hall, revealing both the general direction of visitor movement through the exhibit hall, and the general duration of time spent at a given location within the exhibit. In other words, Museum A's pattern of question keywords (Fig. 3) maps to the nominal flow of its linear exhibit hall (Fig. 5). This suggests that it is possible to use keyword matching from log files to provide accurate insights into visitor actions in a linear space. Further, it also suggests that similar processes can provide insights into visitor pathways even in less linear museum spaces. The exhibit at Museum B represents one such less linear space.

Museum B

At Museum B, we implemented our app in an open-room exhibit hall called “Forces of Nature” that occupies a large and semi-circular space (see Fig. 6).

This exhibit is positioned at one corner of the building and invites visitors to engage with hands-on displays on water, air, and land that are located throughout the open room. Alternatively, visitors could also engage with the display at the center of the hall, which is an interactive stage that visitors can choose to walk on to; this stage simulates effects of



Fig. 5 Linear exhibit hall at Museum A—origins



Fig. 6 Open-room exhibit hall at Museum B—forces of nature

harsh weather or natural disasters such as hurricanes by sprinkling water on visitors and volcanic eruptions by blowing warm air on visitors. The aggregated keyword data from visitors in this exhibit does not show strong peaks by normalized time-in-exhibit, which suggests that visitors explored the open room layout in varied orders. Given the open layout of Museum B's exhibit hall, it is perhaps unsurprising that the pattern of question keywords had less-pronounced peaks per normalized time-in-exhibit (Fig. 4). However, if the exhibit hall's design was intended for visitors to explore in a specific order, such visualization from keyword matching from log files could provide actionable insights into visitor pathways even in this less linear museum space.

Comparing the app's question data between Museum A and Museum B suggests that the spaces are utilized differently. On one hand, Museum A's exhibit hall has one entrance from which visitors progress in the same general direction, encountering the displays in a predetermined order. On the other hand, there is one, albeit large, entrance to Museum B's exhibit hall, but from that point on, visitors can choose to walk left, right, or straight to the center stage without an obvious sense of progression from one display to the next. Even though visitors at both museums were free to explore, skip ahead, or backtrack as they wished, their paths—as demonstrated by the app data and supported by the differing physical layouts of each exhibit hall—were comparatively less linear at Museum B than they were at Museum A. Though relatively simple, this type of data has important implications for museum staff conducting museum evaluation. For example, it demonstrates that app-based data collection methods can provide a relatively affordable and unobtrusive means for collecting valuable data on visitor paths through a museum. Such data illustrates that museum layouts carry with them important affordances and constraints on the paths that visitors tend to take to explore an exhibit hall.

While museums may have *designed* (or intended) paths for visitors to follow, visitors may or may not follow these paths. Having access to unobtrusive and unfiltered app data such as questions' keywords content and time-in-exhibit offers the potential to chart the *actual* paths (deviations from a museum's designed paths) that museum visitors follow through an exhibit hall. The ability to track such keyword and time data may prove to be important because museums possess designed layouts for educational and/or practical reasons, such as facilitating logical transitions from one topic to the next or attempting to avoid pile-ups of visitors at certain stations. However, if visitors follow their own paths in an exhibit, a potential consequence might be that visitors unintentionally follow a visitor-made "path of confusion" leading to eventual disengagement with the museum experience. Additionally, the museum's designed exhibit layout itself may contain points at which visitors tend to disengage with the exhibit for a variety of reasons. These issues could remain

unaddressed if not for unobtrusively collected data of visitor movement. By unobtrusively tracking the sheer number of questions asked and mapping them to time-in-exhibit, patterns revealed can act as a proxy for visitor interest or confusion and may help identify when and where changes need to be made to the exhibit materials. The same technique may be applied, after exhibit changes have been made, to help evaluators assess an exhibit's "success" in directing visitor pathways or engaging visitors with content.

Assessing exhibit "success"

The number of questions asked by visitors about a particular exhibit display, and any noticeable changes (increases or decreases alike) over time, can serve as an indicator of exhibit success in helping visitors learn something from the display or signal a need for adjustment.

Increase of questions

If museum visitors log increased numbers of questions into our app at particular parts of an exhibit hall, this may indicate that: (a) visitors are confused either by a museum display or by the app's provided answer, or they are not satisfied with a provided answer from the app, leading visitors to repeatedly input a similarly worded question; or that (b) they are interested in the museum display or received answers to their initial question from the app that then intrigued them to ask other probing or follow-up questions. In both cases, the museum staff is made aware of a particular display in an exhibit that may need closer scrutiny to include more information (i.e., higher *quantity*) or more detailed information (i.e., higher *quality*). Following up with in-person techniques, such as observations or visitor interviews, could provide greater detail for such museums evaluation while maintaining an affordable approach to doing so. As a response, museums might either clarify the display's content that is already provided (e.g., interpretive text) to address confusion or adjust how the exhibit delivers the content of its display by including additional opportunities (e.g., expert speakers, special docent engagement) to meet the heightened interest and improve museum visitor learning of a topic.

Decrease of questions

Similarly, if the data shows *decreases* in questions asked at particular displays, this may indicate that: (a) the visitors are uninterested or (b) the visitors are satisfied with their understanding of—and learning from—the display. This could point to a display that is either in potential need of revisions to heighten visitor interest or to a display's informational features that can serve as an exemplar to be emulated with other displays in the museum. Again, additional attention by museum staff through, for example, visitor interviews, could explore the characteristics of that display that make it effective. In all these cases, the Dr. Discovery app and the information we have gleaned from it illustrate that unobtrusive insight into the question-asking behaviors of museum visitors can provide timely, actionable information about specific displays and identify targets for more extensive museum evaluation.

Affordable technology-supported museum evaluation

In addition to the practical implications discussed above, these findings and points of discussion contribute to the literature surveyed in this paper. For instance, this work aligns with what Bitgood and Shettel (1996) called the General Facility Design area of research on museum visitor studies, which explores the type of information people require to reduce visitor confusion or feeling lost. More broadly, the work described here contributes to the holistic view of museum evaluation. In this view, the timely and accessible insight into visitor behavior helps evaluators directly infer the effects of any changes to the museum contexts, which is an important and a longstanding component of museum evaluation research (see Bitgood, 2006; Porter, 1938).

In service of such insight into museum visitor behaviors, promising and increasingly available (Economou & Meintani, 2011; Pew Research Center, 2021) technological means have been widely explored by ongoing research (e.g., Ch'ng et al., 2019; De Angeli et al., 2020; Emerson et al., 2020). Our work with the Dr. Discovery mobile app herein described (i.e., a front-end Q&A interface and a back-end analytics portal that visualizes recorded visitor interactions) aligns with such ongoing technology supported efforts to attain direct and data-based insight into museum visitor behaviors through approaches that use mobile technologies in museums (e.g., Chivarov et al., 2013; Damala et al., 2008; Tesoriero et al., 2014). Among other purposes, a popular goal of incorporating such mobile technologies is to conduct museum evaluation that is based on the objective information on visitor interactions with exhibits (Lanir et al., 2017) that indoor positioning technologies such as barcodes, RFID, Wi-Fi, Bluetooth, and infrared technologies may provide (Handojo et al., 2020; Handojo, et al., 2018; Hsi & Fait, 2005; Kuflik et al., 2011; Naismith & Paul Smith, 2009; Rubino et al., 2013; Spachos & Platniotis, 2020). Unfortunately, high-quality museum evaluation that can sustainably use such location-based technologies can be prohibitively expensive and present their own set of limitations and security risks such as GPS not functioning properly in indoor locations, or RFID and NFC covering only short ranges (Ng, 2015), and security or privacy concerns with using public Wi-Fi connections.

The use of apps in museum learning spaces for diverse purposes is hardly a novel endeavor. For instance, shortly after the popularization and wider availability of smartphones that could run such apps, researchers were remarking on their early promise in museum settings (Economou & Meintani, 2011; Johnson et al., 2011). Since then, diverse advancements have been made, with some example including the use of social media (Charitonos et al., 2012), documentation and sharing practices after a museum visit (Hillman et al., 2012), building, curating, and sharing nature photo collections (Kawas et al., 2019), and the use of various mixed reality technologies (Sylaiou et al., 2018). However, although mobile technologies can indeed provide diversified ways of interacting with and learning from museum content in addition to richer, more nuanced data that can, in turn, lead to increasingly informative multipurpose analyses on the part of museum staff, the issues tied to the affordability of museum evaluation must still guide and temper the use of such technologies. To this effect, the present work has shown that even while using relatively simple data that does not rely on more costly and resource-intensive technologies for location tracking, museums that differ both in layout and content can still glean timely and actionable information about specific displays based on visitor data. This information can thus provide data-driven justifications not only for *what* a museum should consider changing, but also *why* and *how*. Such

information can be gleaned from intuiting visitors' locations and interpreting their relative levels of confusion or interest in that particular exhibit, as we discussed earlier.

Additionally, the approach described in this paper has the advantage of providing relatively unobtrusive insight and unfiltered responses when compared to naturalistic evaluations (Leinhardt & Knutson, 2004) where the evaluator is the instrument (Bonner, 1990) while providing data-driven insight that is accessible and affordable both for museums and their visitors. Flexible and affordable apps providing such data could be useful to museums of all sizes and resources, but it is of special concern in the case of museums that are not privileged with resources to install and maintain sophisticated technologies or have dedicated staff on hand for museum evaluation. Understanding visitor engagement in museums to then make informed decisions about those displays is imperative for maintaining the productive convergence of visitor interest with the curated knowledge and learning experiences available in these spaces. Our app makes this needed evaluation data available via log files that contain records of visitor interactions and inquiries as well as the path and evolution of those inquiries.

We offer these previous explanations as examples of the practical uses of our app-based approach for museum evaluation. However, we also note that there may be other, more app-focused, reasons for changes not only in the number of questions but also in question levels. For instance, we found that Game Mode elicited a higher number of questions asked by museum visitors instead of Ask Mode (Nelson et al., 2020). And in follow-up work (Ha et al., 2021; Su et al., 2022), we found that the Game Mode of our app encouraged visitors at Museum A (with linear exhibits) to ask higher-level questions compared to those using the Ask Mode at the same museum. This might indicate a case of app-specific aspects that help foster visitors' changes in quantity and quality of their questions.

Limitations

Our study represents an initial step toward intuiting visitor behaviors from question-asking data. Due to the nature of the study and available data, our inferences on visitor movement, behavior, and perceptions were not tested nor validated through traditional techniques such as observations, post-visit questionnaires, or interviews. As such, under our current iteration of the app, we cannot confidently determine, for example, the reason why visitors asked a particular question or what prompted an increase or decrease in the total number of questions asked. We can only hypothesize about such causal relationships based on which condition participants were assigned to (Game or Ask Modes) and their resulting question data, as we have done in this study and other work. We recognize that such causal information would be beneficial for museums to make more informed decisions based on question data.

Additionally, it is clear that visitors could ask any question at any place in the museum, so intuiting visitor paths by analyzing questions is a limited way of accomplishing such a task. This means that our inferences carry the constraints of app-based tablet technologies that did not include features associated with location-tracking or instant positioning such as GPS, RFID, and NFC, which might enhance the utility of a question-asking app for evaluation and tracking of visitor movements. In other words, museum staff would be unlikely to conclusively "pinpoint" users based on their question data alone unless such data would be accompanied by, for instance, observational data or location tracking devices/technology. However, such technologies also have their unique set of disadvantages that include increases in costs and maintenance as well as practical concerns such as having reliable

signals for such technologies and privacy concerns. It is precisely these disadvantages that the current work is responding to and is applicable for. Therefore, our app data and analysis show that informative inferences can be made even without these more complicated or costly additions.

Future directions

Future work should consider including several augmentations such as creating a network of museums that would contribute to curated sets of in-app questions and answers. Having an expanded and diverse network of museums using the same app to curate and develop questions and answers would help to iteratively improve its design for diverse populations that visit museum sites. Doing so would also maintain an expert-informed and verified set of answers to questions that avoids the potential for misinformation that is possible on already existing services such as search engines (Krutka et al., 2021).

Another augmentation would be designing an option that allows participants to quickly identify the reason for their questions. Participants could specify *why* they are asking a question by quickly selecting from multiple-choice options such as “I don’t understand the display” or “I am interested in finding out more”. This could help museum staff determine more appropriate ways forward when assessing or changing specific displays while decreasing the need to implement more hands-on and time-consuming approaches such as observations or interviews.

Another direction for future work includes employing a more sophisticated semantic and pragmatic analysis of keyword data. For instance, one alternative would be to include text-mining functions such as those found in specific packages of *R*—a language and environment for statistical computing within which statistical techniques are implemented for a variety of functions such as data mining. Doing so would allow us to conduct more nuanced and informative models for analyzing our current and other potentially new keywords.

Additionally, apps can be designed so that they identify keywords that are associated with exhibit displays and automatically tag them with timestamps. Doing so could, for example, help illustrate visitors’ movement and routes in real-time based on the keyword content of their questions, creating a type of keyword “heat map”. Future work could also include keywords that were related to each other in a relational matrix. For example, “meteors” is related to “meteorites” while “selenite” is an example or subset of “mineral”.

Conclusion

Our project aimed to address the need for affordable, ongoing, and scalable museum evaluation through unobtrusively collecting data useful for understanding museum visitors’ behaviors and interactions with museum content. By utilizing unobtrusive and data-driven approaches, we can provide multimodal evidence to support museums in their efforts to respond to the learning needs and expectations of their visitors. Museum staff may use this data on visitor behavior for multiple purposes, such as informing potential redesigns of exhibit halls and gaining a deeper understanding of visitor trends. Specifically, by examining the content and number of in-app questions asked, relative interest or confusion in certain topics can be approximated, helping identify points in an exhibit hall where visitors tend to engage or disengage. This data can be used to inform museum personnel’s decisions to change an exhibit’s displays to address visitor curiosity or confusion and justify

the need for further in-person evaluations. Executing changes to museum content based on visitor actions reflects a focus on what Bitgood (2006) referred to as an *exhibit/visitor interaction view* that assumes that both visitor factors (such as visitor interests) and exhibit factors (such as designing of exhibit elements) must be considered jointly. Additionally, by coupling question content with question timestamps, museums may leverage visitor behavior to intuit their preferred exploration paths and make adjustments as necessary based on time spent at specific displays in an exhibit or on question content.

This project provides evidence that relatively simple data, collected through widely available and unobtrusive means, can provide museums with a functional, cost-effective, and versatile way to conduct museum evaluation by leveraging insight from visitors' behavior. This, in turn, suggests that there may still be more innovative ways in which visitor data can be constructed, interpreted, explored, and leveraged for multiple purposes to improve museum visitor learning experiences. In the same vein, we believe that museums can come to make improved uses of their own museum spaces if they are informed by the insights that are efficiently gleaned from unobtrusive data collection methods such as those afforded by mobile apps such as Dr. Discovery.

Acknowledgements The authors would like to acknowledge the staff at both of our partner museums.

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

Data availability The data that support the findings of this study are available upon request from the corresponding author.

Declarations

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

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

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

References

- Adams, M. (2012). Where have we been? What has changed? And where do we need to go next? *Journal of Museum Education*, 37(2), 25–35. <https://doi.org/10.1080/10598650.2012.11510728>
- American Alliance of Museums. (2018). *Museums & public opinion*. <https://www.aam-us.org/2018/01/20/museums-and-public-opinion/>
- American Museum of Natural History. (n.d.). *Evaluation*. <https://www.amnh.org/learn-teach/evaluation-research-and-policy/evaluation>
- Anderson, G. (2012). *Reinventing the museum*. Altamira Press.
- Beghetto, R. A. (2014). The exhibit as planned versus the exhibit as experienced. *Curator: The Museum Journal*, 57(1), 1–4. <https://doi.org/10.1111/cura.12047>
- 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.
- Bitgood, S. (2006). An analysis of visitor circulation: Movement patterns and the general value principle. *Curator*, 49(4), 463–475. <https://doi.org/10.1111/j.2151-6952.2006.tb00237.x>
- Bitgood, S., & Shettel, H. H. (1996). An overview of visitor studies. *Journal of Museum Education*, 21(3), 6–10. <https://doi.org/10.1080/10598650.1996.11510329>
- Black, G. (2005). *The engaging museum: Developing museums for visitor involvement*. Psychology Press.
- Black, G. (2012). *Transforming museums in the twenty-first century*. Routledge.

- Black, G. (2015). *Developing museum display for informal learning*. ProQuest Dissertations & Theses Global.
- Blöckner, M., Danti, S., Forrai, J., Broll, G., & Luca, A. D. (2009). Please touch the exhibits! Using NFC-based interaction for exploring a museum. In *Proceedings of the 11th international conference on human-computer interaction with mobile devices and services* (Vol. 71, pp. 1–2). <https://doi.org/10.1145/1613858.1613943>
- Bonner, J. (1990). Formal versus naturalistic evaluation in the museum context. *Visitor Studies*, 2(1), 211–224.
- Bowman, C. D. D., Adkins, A., Owen, B. L., Rogers, K. J., Escalante, E., Bowman, J. D., Nelson, B. C., & Stoltman, A. (2019). Differences in visitor characteristics and experiences on episodic free admission days. *Museum Management and Curatorship*, 35(3), 264–280. <https://doi.org/10.1080/09647775.2019.1578992>
- Braehmer, A., Dorner, W., & Weinfurter, A. (2019). Digital technologies for cross-cultural and cross-medial museum work. In *9th International conference on advanced computer information technologies (ACIT)* (pp. 487–491). <https://doi.org/10.1109/ACITT.2019.8779854>
- Cameron, F., & Robinson, H. (2007). Digital knowledgespaces: Cultural, theoretical, practical, and usage issues facing museum collection databases in a digital epoch. In F. Cameron & S. Kenderdine (Eds.), *Theorizing digital cultural heritage: A critical discourse* (pp. 165–191). MIT Press.
- Carci, G., Caforio, A., & Gamper, C. (2019). Digital technologies and museums: Augmented reality, learning and audience development. *Form@re Open Journal Per La Formazione in Rete*, 19(1), 274–286. <https://doi.org/10.13128/formare-24619>
- Ch'ng, E., Cai, S., Leow, F. T., & Zhang, T. E. (2019). Adoption and use of emerging cultural technologies in China's museums. *Journal of Cultural Heritage*, 37, 170–180.
- Charitonos, K., Blake, C., Scanlon, E., & Jones, A. (2012). Museum learning via social and mobile technologies: (How) can online interactions enhance the visitor experience? *British Journal of Educational Technology*, 43(5), 802–819. <https://doi.org/10.1111/j.1467-8535.2012.01360.x>
- Chivarov, N., Ivanova, V., Radev, D., & Buzov, I. (2013). Interactive presentation of the exhibits in the museums using mobile digital technologies. *IFAC Proceedings Volumes*, 46(8), 122–126. <https://doi.org/10.3182/20130606-3-XK-4037.00014>
- Chong, C., & Smith, D. (2017). Interactive learning units on museum websites. *Journal of Museum Education*, 42(2), 169–178. <https://doi.org/10.1080/10598650.2017.1301626>
- Cockburn, A., & McKenzie, B. (2001). What do web users do? An empirical analysis of Web use. *International Journal of Human-Computer Studies*, 54(6), 903–922.
- Damala, A., Cubaud, P., Bationo, A., Houlier, P., & Marchal, I. (2008). Bridging the gap between the digital and the physical: Design and evaluation of a mobile augmented reality guide for the museum visit. In *Proceedings of the 3rd international conference on digital interactive media in entertainment and arts (DIMEA '08)* (pp. 120–127). Association for Computing Machinery. <https://doi.org/10.1145/1413634.1413660>
- Damala, A., Ruthven, I., & Hornecker, E. (2019). The MUSETECH model: A comprehensive evaluation framework for museum technology. *Journal on Computing and Cultural Heritage*, 12(1), 1–22. <https://doi.org/10.1145/3297717>
- De Angeli, D., Kelly, R. M., & O'Neill, E. (2020). Beyond happy-or-not: Using emoji to capture visitors' emotional experience. *Curator: The Museum Journal*, 63(2), 167–191.
- Durbin, G. (1996). *Developing museum exhibitions for lifelong learning*. The Stationery Office.
- East of England Museum Hub. (2008). *Evaluation toolkit for museum practitioners*.
- Economou, M., & Meintani, E. (2011). Promising beginning? Evaluating museum mobile phone apps. In *Proceedings of the re-thinking technology in museums: Emerging experience conference* (pp. 87–101). Retrieved from <http://eprints.gla.ac.uk/104173/>
- Emerson, A., Henderson, N., Rowe, J., Min, W., Lee, S., Minogue, J., & Lester, J. (2020). Investigating visitor engagement in interactive science museum exhibits with multimodal bayesian hierarchical models. In *International Conference on Artificial Intelligence in Education* (pp. 165–176). Springer.
- Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology*, 16, 171–190.
- Falk, J. H., & Dierking, L. D. (2013). *The museum experience revisited*. Routledge.
- Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., & Yuhnke, B. (2016). *NMC Horizon report: 2016 museum edition*. The New Media Consortium.
- Gilbert, S. (2016). Please turn on your phone in the museum. *The Atlantic*. Retrieved January 28, 2020, from <https://www.theatlantic.com/magazine/archive/2016/10/please-turn-on-your-phone-in-the-museum/497525/>

- Ha, J., Pérez Cortés, L. E., Su, M., Nelson, B., Bowman, C., & Bowman, J. (2021). The impact of a gamified mobile question-asking app on science museum visitor group interactions: An ICAP framing. *International Journal of Computer-Supported Collaborative Learning*, 16, 367–401. <https://doi.org/10.1007/s11412-021-09350-w>
- Hammady, R., Ma, M., & Powell, A. (2018). User experience of markerless augmented reality applications in cultural heritage museums: 'MuseumEye' as a case study. In L. De Paolis & P. Bourdot (Eds.), *Augmented reality, virtual reality, and computer graphics* (p. 10851). Springer.
- Handojo, A., Lim, R., Octavia, T., & Kurnia Anggita, J. (2018). Museum interactive information broadcasting using indoor positioning system and bluetooth low energy: A pilot project on Trowulan museum Indonesia. In *2018 3rd technology innovation management and engineering science international conference (TIMES-iCON)*. <https://doi.org/10.1109/TIMES-iCON.2018.8621815>
- Handojo, A., Octavia, T., Lim, R., & Kurnia Anggita, J. (2020). Indoor positioning system using BLE beacon to improve knowledge about museum visitors. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, 18(2), 792–798. <https://doi.org/10.12928/telkomnika.v18i2.14886>
- Hein, G. (1998). *Learning in the museum*. Routledge.
- Hillman, T., Weilenmann, A., & Jungselius, B. (2012). Creating live experiences with real and stuffed animals: The use of mobile technologies in museums. In *Proceedings of The transformative museum* (pp. 138–149).
- Hooper-Greenhill, E. (1999). *The educational role of the museum*. Routledge.
- Hsi, S., & Fait, H. (2005). RFID enhances visitors' museum experiences at the exploratorium. *Communications of the ACM*, 48(9), 60–65. <https://doi.org/10.1145/1081992.1082021>
- Hsu, H., & Liao, H. (2011). A mobile RFID-based tour system with instant microblogging. *Journal of Computer and System Sciences*, 77(4), 720–727. <https://doi.org/10.1016/j.jcss.2010.02.011>
- Huang, Y., Wang, S., & Sandnes, F. E. (2011). RFID-based guide gives museum visitors more freedom. *IT Professional*, 13(2), 25–29. <https://doi.org/10.1109/MITP.2011.33>
- Ingram, A., & Northcote, M. (1999). Using Web server logs in evaluating instructional Websites. *Journal of Educational Technology Systems*, 28(2), 137–157.
- Johnson, L., Adams, S., & Witchey, H. (2011). *The NMC horizon report: 2011 museum edition*. The New Media Consortium.
- 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* (pp. 346–379). MuseumsEtc.
- Kawas, S., Chase, S. K., Yip, J., Lawler, J. J., & Davis, K. (2019). Sparking interest: A design framework for mobile technologies to promote children's interest in nature. *International Journal of Child-Computer Interaction*, 20, 24–34. <https://doi.org/10.1016/j.ijcci.2019.01.003>
- Koreneva, A. Y. (2015). Museum as an institute of continuous and non-formal education. In *Bulletin of the Kemerovo State University* (pp. 216–220).
- Krutka, D. G., Smits, R. M., & Wilhelm, T. A. (2021). Don't be evil: Should we use google in schools? *TechTrends*. <https://doi.org/10.1007/s11528-021-00599-4>
- Kuflik T., Lanir, J., Dim, E., Wecker, A., Corra, M., Zancanaro, M., & Stock, O. (2011). Indoor positioning: Challenges and solutions for indoor cultural heritage sites. In *Proceedings of the 16th international conference on intelligent user interfaces* (pp 375–378). <https://doi.org/10.1109/IEEEI.2012.6376935>
- Lane, H. C., Cahill, C., Foutz, S., Auerbach, D., Noren, D., Lussenhop, C., & Swartout, W. (2013). The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education Lecture notes in computer science*. (Vol. 7926). Springer.
- Lanir, J., Kuflik, T., Sheidin, J., Yavin, N., Leiderman, K., & Segal, M. (2017). Visualizing museum visitors' behavior: Where do they go and what do they do there? *Personal and Ubiquitous Computing*, 21(2), 313–326. <https://doi.org/10.1007/s00779-016-0994-9>
- Leinhardt, G., & Knutson, K. (2004). *Listening in on museum conversations*. Altamira Press.
- Lohr, S. (2014). Museums morph digitally. *The New York Times*. Retrieved January 28, 2020, from <https://www.nytimes.com/2014/10/26/arts/artsspecial/the-met-and-other-museums-adapt-to-the-digital-age.html>
- Marty, P. F., & Jones, K. B. (Eds.). (2012). *Museum informatics: People, information, and technology in museums*. Routledge.
- Naismith, L., & Paul Smith, M. (2009). Using mobile technologies for multimedia tours in a traditional museum setting. In A. Mohamed (Ed.), *Mobile learning: Transforming the delivery of education & training*. AU Press.

- National Museum Director's Council. (2015). *Museums matter: Executive summary*. Retrieved from https://www.nationalmuseums.org.uk/media/documents/publications/museums_matter/museums_matter_web.pdf
- Nelson, B., Bowman, C., & Bowman, J. (2017). Designing for data with Dr. Discovery: Design approaches for facilitating museum evaluation with real-time data mining. *Technology, Knowledge, and Learning Journal*, 22, 427–442. <https://doi.org/10.1007/s10758-017-9313-4>
- Nelson, B. C., Bowman, C. D. D., Bowman, J. D., Pérez Cortés, L. E., Adkins, A., Escalante, E., Owen, B. L., Ha, J., & Su, M. (2020). Ask Dr. Discovery: The impact of a casual mobile game on visitor engagement with science museum content. *Educational Technology Research and Development*, 68, 345–362. <https://doi.org/10.1007/s11423-019-09696-x>
- Ng, T. M. (2015). From “Where I am” to “Here I am”: Accuracy study on location-based services with iBeacon technology. *HKIE Transactions*, 22(1), 23–31.
- Pekarik, A. J., Schreiber, J. B., Hanemann, N., Richmond, K., & Mogel, B. (2014). IPOP: A theory of experience preference. *Curator: The Museum Journal*, 57(1), 5–27. <https://doi.org/10.1111/cura.12048>
- Pew Research Center. (2021). *Mobile fact sheet*. Retrieved June 30, 2022 from <https://www.pewresearch.org/internet/fact-sheet/mobile/>
- Porter, M. (1938). *The behavior of the average visitor in the Peabody Museum of Natural History* (Vol. 16). American Association of Museums.
- Proctor, N. (2015). Mobile in museums: From interpretation to conversation. *The International Handbooks of Museum Studies*, 4(22), 499–525.
- Røtne, O., & Kaptelinin, V. (2013). Design choices and museum experience: A design-based study of a mobile museum app. In C. Collazos, A. Liborio, & C. Rusu (Eds.), *Human computer interaction, CLIH 2013. Lecture notes in computer science*. (Vol. 8278). Springer.
- Rubino, I., Xhembulla, J., Martina, A., Bottino, A., & Malnati, G. (2013). Musa: Using indoor positioning and navigation to enhance cultural experiences in a museum. *Sensors*, 13(12), 17445–17471. <https://doi.org/10.3390/s131217445>
- Ruíz, B., Pajares, J. L., Utray, F., & Moreno, L. (2011). Design for all in multimedia guides for museums. *Computers in Human Behavior*, 27(4), 1408–1415. <https://doi.org/10.1016/j.chb.2010.07.044>
- Smithsonian Institution. (2004). *The evaluation of museum educational programs: A national perspective*. Office of Policy and Analysis, Smithsonian Institution.
- Spachos, P., & Platniotis, K. K. (2020). BLE beacons for indoor positioning at an interactive IoT-based smart museum. *IEEE Systems Journal*, 14(3), 3483–3493. <https://doi.org/10.1109/JSYST.2020.2969088>
- 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>
- Stratton, A., Bates, C., & Dearden, A. (2017). Quando: Enabling museum and art gallery practitioners to develop interactive digital exhibits. In S. Barbosa, P. Markopoulos, F. Paternò, S. Stumpf, & S. Valtolina (Eds.), *End-user development* (pp. 100–107). Springer.
- Su, M., Ha, J., Pérez Cortés, L. E., Bernier, J., Yan, L., Nelson, B., Bowman, J., & Bowman, C. (2022). *Question asking in museums: How much could a question-asking mobile app tell us?* [Manuscript submitted for publication]
- Sylaiou, S., Kasapakis, V., Dzardanova, E., & Gavalas, D. (2018). Leveraging mixed reality technologies to enhance museum visitor experiences. In *International conference on intelligent systems* (pp. 595–601). <https://doi.org/10.1109/IS.2018.8710530>
- Tesoriero, R., Gallud, J. A., Lozano, M., & Penichet, V. M. R. (2014). Enhancing visitors' experience in art museums using mobile technologies. *Information Systems Frontiers*, 16, 303–327. <https://doi.org/10.1007/s10796-012-9345-1>
- Vadeboncoeur, J. A. (2006). Chapter 7: Engaging young people: Learning in informal contexts. *Review of Research in Education*, 30(1), 239–278.
- Valente-Marques, D. (2017). *The visitor experience using augmented reality on mobile devices in museum exhibitions (dissertation)*. ProQuest Dissertations Publishing.
- Wellington, J. (1990). Formal and informal learning in science: The role of the interactive science centres. *Physics Education*, 25, 247–252.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Luis E. Pérez Cortés is a post-doctoral associate at the Learning Research and Development Center (LRDC) at the University of Pittsburgh. His research explores how the playing, making, and modding of digital and tabletop games helps develop young peoples' abilities to address environmental issues, benefit underrepresented and underserved students, and aids players see the world, society, and themselves as a malleable, re-designable entities.

Jesse Ha is a post-doctoral associate at the University of California, Los Angeles. His research interests focus on quantitative methodology, learning sciences, education technology, and collaborative learning.

Man Su is a doctoral student at the Mary Lou Fulton Teachers College at Arizona State University. Her research interests focus on understanding how the interaction of human and computer technologies affect how people learn, and investigating instructional methods to scaffold learning in both formal and informal learning contexts.

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

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

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