Education Tech Research Dev (2020) 68:345–362
https://doi.org/10.1007/s11423-019-09696-x

RESEARCH ARTICLE

Ask Dr. Discovery: the impact of a casual mobile game on visitor engagement with science museum content

Brian C. Nelson[1] · Catherine D. D. Bowman[2] · Judd D. Bowman[2] · Luis E. Pérez Cortés[1] · Adrianna Adkins[3] · Edgar Escalante[4] · Brooke L. Owen[5] · Jesse Ha[1] · Man Su[1]

Published online: 16 July 2019

Abstract
This study examines the impact of a mobile game app on science museum visitors' level of engagement with exhibit content, compared to a non game-based version of the same app. Ask Dr. Discovery (Dr. D) is a question-asking app containing two versions: a Game Mode employing casual game mechanics and an Ask Mode providing a baseline version. We implemented both versions of Dr. D at two science museums located in the Southwestern United States with 1539 participants. In both conditions, participants could type or speak questions to receive vetted answers about museum content, but only Game Mode embedded question-asking within a simple game. Participants' level of engagement was represented by the number of questions asked about exhibit content in Dr. D. Additionally, we explored the relationship between app engagement and participants' self-reported level of interest in science. All participants completed pre- and post-questionnaires with questions related to science interest, impressions of the Dr. D app, and demographic information. Results in both museums indicated that users of Game Mode asked nearly twice as many questions on average as participants using Ask Mode. Science interest predicted engagement at one of two sites. Demographic variables, including gender, age, and race/ethnicity were not found to influence the rate of question asking in either mode. These results indicate that employing simple game mechanics in apps for museum visitors may lead to strong positive impacts on visitor engagement with museum content.

Keywords Museum evaluation · Games and user engagement · Informal learning

✉ Brian C. Nelson
Brian.Nelson@asu.edu

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

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

[3] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, USA

[4] Department of Physics, Arizona State University, Tempe, USA

[5] School of Life Sciences, Arizona State University, Tempe, USA

Introduction

Ask Dr. Discovery (Dr. D) is a National Science Foundation-funded study addressing the need for affordable, ongoing, large-scale museum evaluation while investigating innovative ways to encourage museum visitors to engage deeply with museum content. Conducting ongoing museum evaluation is imperative because, while the "hardscape" (physical exhibits) of museums is often prohibitively costly to change, comprehensive exhibit experiences also depend on flexible elements that can change over time, such as docents (location, training, and contents of their experience carts), multimedia, public events, temporary signage, and webpages. To improve these elements, timely and accessible insight into the minds of visitors becomes invaluable. To this end, visitor-facing apps provide an attractive and potentially effective data collection method in support of evaluation efforts. Besides reach and data volume, one strength of app-based data collection, when compared to traditional museum data collection methods, is that all visitor in-app interactions may be unobtrusively and anonymously captured. These 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.

To realize the Dr. D project aims, we designed and developed a mobile app with two parts: (1) a front-end Q&A interface through which visitors can ask questions and receive vetted answers about museum content, and (2) a back-end analytics portal that visualizes recorded visitor interactions with the app to be mined by museum staff, evaluators, and researchers. Through a partnership between Arizona State University (ASU), the Arizona Science Center (ASC), and the Arizona Museum of Natural History (AZMNH), we developed the Dr. D app primarily to support museum evaluation by collecting and analyzing data from visitor questions about exhibit content for use by museum personnel, while providing visitors with vetted answers to their questions.

In the current study, we focus on Dr. D's two modes of asking questions: Ask Mode and Game Mode. Ask Mode situates visitor question-asking in a non-game-based version of the app while Game Mode situates visitor question-asking within simple, casual game mechanics. In this study, we describe the impact of these two modes on visitor engagement. For our research questions, we operationalize "engagement" as the number of questions asked in the app about science content. We also explore "time spent in the exhibit" as an alternative measure of visitor engagement. Additionally, we explore the relationship between visitor reported levels of interest in science and app engagement.

The goals, design, and implementation of the Dr. D mobile app and the current study are situated within the literature on mobile apps in museums, museum evaluation, informal learning in museums, educational games, and games for engagement.

Background

Mobile apps in museums

The ownership of smartphones in the U.S. has grown rapidly. The Pew Research Center (2018) reports that 77% of American adults own a smartphone (including 67% of adults earning below $30,000), up from 35% in 2011. Smartphone prevalence is even higher among young adults aged 18–29 (94%). Our Ask Dr. Discovery (Dr. D)

study suggests that ownership among museum visitors may be greater still: Of the more than 2300 museum visitors who have participated in our study overall, more than 97% of participants reported owning a smartphone.

Museums (via apps, e.g., AZMNH's *Explorer*; Brooklyn Museum's *Ask*; Cleveland Museum's *ArtLens*; CMH's *More*; Smithsonian's *Infinity of Nations*) and researchers (e.g., Bickmore et al. 2013; Katz et al. 2011; Lane et al. 2013; Proctor 2015; Røtne and Kaptelinin 2013) have been quick to realize the opportunities provided by the ubiquity of mobile devices. In 2011, the New Media Consortium Horizon Report (Museum Edition) found that, "Mobile Apps are the most relevant features of mobiles for museums right now" (Johnson et al. 2011, p. 7). Only five years later, the 2016 version of the same report found apps to be such a reality in museums that the report's focus shifted from *whether* to *how* (Johnson et al. 2016). Proctor (2015) affirms, "Mobile is no longer a 'nice to have': we hear museum professionals echoing the sense that the pervasiveness of mobile makes it an essential part of the museum experience" (p. 501).

The 2016 Horizon Report identifies data analytics as a key medium-term (3-to-5 years) trend coupled to mobile in museums. They note, "Today museum staff are also more willing to see themselves as learners, taking advantage of user-generated content" (p. 10), however they identify "improving digital literacy of museum professionals" (p. 21) as a key challenge and note that some thought leaders believe there are not enough best practices guidelines for technology training for current and preservice museum staff, and the most progressive examples are taking place informally. Professional development around how emerging technologies can be leveraged to further museums' interpretation goals and enhance their visitor experiences is needed at all levels of museum education (Freeman et al. 2016, p. 24).

Museum evaluation

While mobile devices offer the potential for unprecedented types and volumes of visitor data in support of evaluation, making sense of those data in practical, continuous ways remains a challenge. Communities across the U.S. host local museums and other informal learning venues, many with limited financial and staffing resources. Such sites are rarely, if ever, able to employ an evaluator. The Smithsonian found in a national study that "only a handful of institutions—13 percent—have someone on staff that spends all or part of his or her time on evaluation of exhibits or programs" (2004, p. 15). They note that museum evaluation in most institutions is tacked on to the responsibilities of other staff. A 2011 study by the Federal Institute of Museum and Library Services found that while respondents "clearly understood the utility of evaluations in refining projects, gauging their success, and informing future directions, evaluation posed challenges for museums. Some noted their own lack of knowledge of evaluation practices; others noted the cost" (Apley et al. 2011, p. III-24). As Adams (2012) points out, "While the 'want to' in the practitioner's desire to conduct evaluation is much stronger than ever before, the 'can do' is a different matter" (p. 28). The Dr. Discovery app and project, while not a replacement for full-scale evaluation, seeks to bolster museum evaluation efforts by exploring sustainable, generalizable, and transferable ways to build capacity among museum professionals for working with and learning from their own data to enable evaluation and data-driven decision making.

Informal learning in museums

The current study examines participant engagement with the Dr. D mobile app in support of data collection for museum evaluation. Because the larger NSF-funded Ask Dr. Discovery project is focused on museum evaluation, investigating participant learning was outside the scope of the study. However, the project is situated in the larger realm of informal learning, and our focus on museum visitor engagement is informed by, and may help inform, related work on informal learning.

Informal learning is characterized as motivating, spontaneous, learner-centered, and open-ended (e.g. Wellington 1990; Ramsey-Gassert 1997; Eshach 2007; Vadeboncouer 2006) factors that contribute to learning and interest by offering experiences different from traditional classroom learning. Abraham-Silver (2006) explains that those in informal learning environments show increased attention and enthusiasm, as well as a willingness to observe, question, and discuss objects around them. Furthermore, she asserts, "students who are generally reticent in the formal environment of the classroom or laboratory are more likely to be drawn out and eagerly participate when immersed in the activity and energy found in the galleries of a science center" (pg. 12). Thus, informal learning environments may serve to better engage both students and lifelong learners who do not (or did not) have a strong interest in classroom science.

Beyond increasing motivation and improving attitudes toward science, informal learning also plays a role in enhancing learning and reasoning abilities, which is particularly important when considering complex topics such as sustainability. Enriched informal activities in settings such as museums may provide experiences that stimulate cognitive conflict and promote social interaction to improve visitors scientific reasoning abilities (Gerber et al. 2001). As such, there is a two-fold need to (1) help visitors explore complex topics informally across visits and venues as the public's understanding of science is the result of varied contexts, both formal and informal (Falk et al. 2007, p. 456) and (2) support museum staff as their roles expand beyond facility guides into sources for interpretation (Johnson et al. 2016, p. 10).

Educational use of digital games

Scholars have been exploring video or computer (digital) games' affordances for engagement for decades, with Malone (1981) as an early example. However, the literature on digital games' value in teaching and learning accelerated, reignited by James Gee's *What video games have to teach us about learning and literacy* (2003). Gee challenged the perception of video game play as a waste of time and elaborated how, when played in a thoughtful context that fosters critical thinking, video games can become excellent tools which model good design for teaching and learning. The Federation of American Scientists (2006) echoed Gee's views, stating that "there are many features of digital games, including game design approaches and digital game technologies, which can be applied to address the increasing demand for high quality education" (FAS 2006, p. 4). A number of scholars have since examined video games' potential for improving education (e.g., Shaffer 2006; Squire 2011), as well as how games may be leveraged to improve business (e.g., Edery and Mollick 2009), address social problems like depression, obesity, poverty, and climate change (e.g., McGonigal 2011), and serve as platforms for art and politics (e.g., Bogost 2011).

Researchers have also argued that video games, both commercial and educational, have the potential to engage students who are typically disenfranchised by more traditional classroom curricula, encouraging them to actively participate in learning within informal contexts (e.g., Cassell and Jenkins 2000; Dickey 2005, 2006; Hayes 2005, 2008; Hartmann and Klimmt, 2006; Kafai 2006; Steinkuehler and Duncan 2008; Jenkins et al. 2003; Prensky 2001; Steinkuehler 2004). For example, Cassell and Jenkins (2000) describe the importance of video games in popular culture and in children's lives. Steinkuehler (2004) examines the power of commercial games for informal learning, demonstrating the impact of playing massively multiplayer online role-playing games (MMORPG) on shaping player literacy practices.

Games and engagement

Although higher levels of engagement may lead to increases in learning, this is not always the case. Some findings suggest that use of games can bolster player engagement, even if they do not foster learning that can be directly measured. For example, Annetta et al. (2009) conducted a study in which a digital game designed to teach genetics did not lead to improved learning outcomes compared to a non-game curriculum despite showing significantly increased levels of engagement ($p < 0.05$) by learners who played the game. Similarly, Ke (2008), in a study into the impact of a math game on learning and engagement, found participants reporting more positive attitudes toward math when playing a game, but no accompanying increase in scores on a performance measure. Furthermore, in their meta-review of games and simulations used for science learning, Clark et al. (2009) found similar mixed results across an array of studies: near universal increased engagement on the part of participants playing games, but only occasional increased scores on learning measures associated with games vs. non-game-based learning environments.

Although others have made causal relationships between engagement and learning, the Dr. D project does not focus on measuring or assessing learning. Instead, our focus is on exploring how simple game mechanics may affect visitor engagement. To investigate our research questions, we use the number of in-app questions asked as a proxy for participant engagement. That is, the greater the number of questions participants asked within the app, the higher we classified their engagement.

Another approach to measuring engagement is "time on task", with the idea being that the more time a participant spends doing something, the greater their level of engagement in the task. We believe that, for the Dr. D project, question-asking level is a more direct measure of engagement, as it demonstrates ongoing, intentional visitor engagement. However, because "time on task" is a common tool for measuring engagement (e.g. Johns and Woolf 2006; Joseph 2005), and because we recorded "time in exhibit" in minutes for all participants, we include an exploratory analysis of the relationship between app mode and engagement as represented by time spent in the exhibit.

Study

For this study, we employed a two-group experimental design to examine whether museum visitors engage more with a baseline question-and-answer version of the Dr. D mobile app (Ask Mode) or a game-based version (Game Mode) of this same app. Two partner sites in the Phoenix, Arizona metro area—the Arizona Science Center (ASC) and the Arizona Museum of Natural History (AZMNH)—allowed visitor groups to be lent an iPad with

the Dr. D app pre-installed for use throughout a single exhibit at each site. On alternating implementation days, all participants were assigned to either Dr. D's Game or Ask Mode.

Research questions

1. How do casual game mechanics incorporated into a mobile question-asking app (Game Mode) impact levels of app use by museum visitors exploring a science-related exhibit compared to a version of the app lacking game mechanics (Ask Mode)?
2. Does the level of app use, overall and by mode (Game vs. Ask), vary as a function of museum (ASC or AZMNH)?
3. Does level of app use vary by self-reported level of science interest, both overall and by mode (Game vs. Ask)? Does app use vary as a function of the demographics of participants?

Dr. Discovery app

Dr. Discovery principally serves as a question-asking app for museum visitors. To increase accessibility of use, questions may be inputted through text as well as by voice in either English or Spanish. To display potential answers, Dr. D connects to a curated database of more than 13,500 verified questions and 5000 answers. Note that the total number of questions is greater than the total number of answers because many questions (or ways of asking a question) have the same answer. This curated database was developed by research project team members and related specifically to the content of the museum exhibits in the study.

Dr. D makes use of a tiered response system to answer questions. If a user's question perfectly matches one in the database, Dr. D replies with the associated answer. If the question match is close, but not perfect, the system responds, "If you meant [question from the database], the answer is [answer from the database]." With a partial match, Dr. D responds, "Did you mean to ask [question from the database] or [other question from the database]?", both of which are hyperlinked for the user to tap to select. If there is no match, Dr. D will reply that it has not yet learned that information and provide the user with a random science fact from its database. Users also can indicate whether Dr. D answered their question(s) adequately, allowing us to make continuous improvements to the database.

Ask Mode

Dr. D's Ask Mode (Fig. 1) represents the baseline treatment. In Ask Mode, visitors ask questions about the exhibits to Dr. D through a simple Q&A interface. Visually, app users see an image of Dr. D floating in space in an astronaut's suit. A familiar text entry box allows users to type or speak questions to Dr. D in English or Spanish. The answers appear in a highlighted window. The user's question history is also shown.

Game Mode

Dr. D's Game Mode (Fig. 2) represents the experimental treatment. In Game Mode, users ask questions similarly to Ask Mode, either through text or voice, but Game Mode overlays simple game mechanics onto the baseline question asking and answering process. It is

Fig. 1 Dr. Discovery Ask Mode

important to notice that Game Mode represents a central tension in implementing a casual game in a museum setting. This tension resides in the fact that the usefulness of the game for museum personnel as an evaluation tool directly relates to the number of visitor questions. That is, on one hand, higher number of questions asked results in a more nuanced and useful portrait of visitor interests, areas of confusion, and patterns of interaction with the exhibit content. On the other hand, greater number of questions asked represents time that is diverted from actual museum exhibits. Because of potential usefulness to museum staff, the project strives to create a game that encourages players to ask a lot of questions about the exhibit content. However, if the visitors are concentrating mostly on asking questions while visiting an exhibit in hopes of winning the game, they may not fully engage with the physical museum experience outside of the screen. To avoid this undesired situation, Game Mode underwent several design iterations.

As we have described elsewhere (Nelson et al. 2017), we followed a design-based research process in creating the Game Mode of the Dr. D app. We iterated through three designs, focusing on the central challenge of making the game engaging enough for visitors

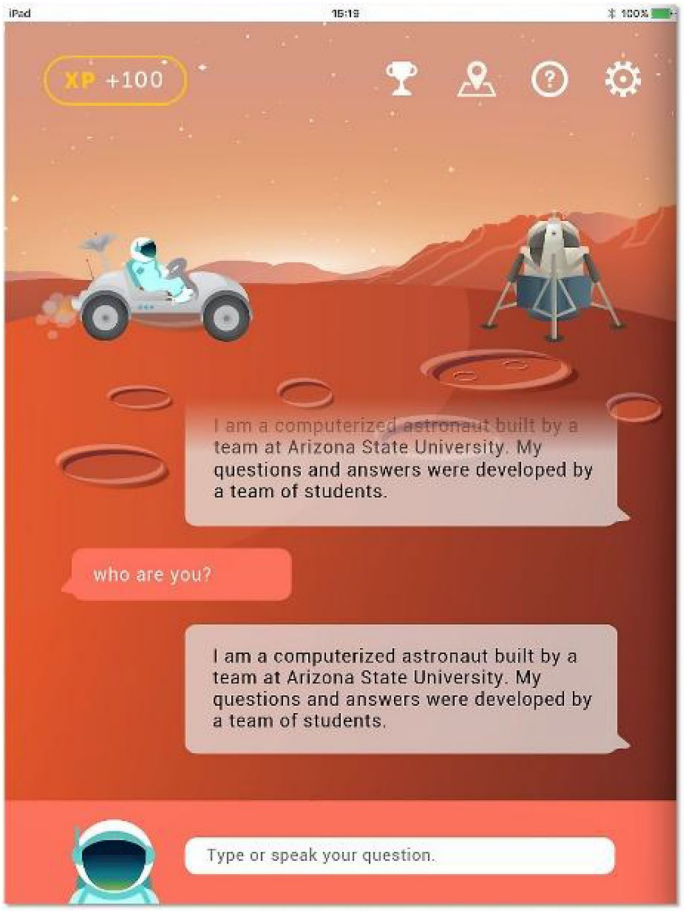

Fig. 2 Dr. Discovery Game Mode

to want to use it while keeping their primary focus on museum exhibit content. The game positions museum visitors to help Dr. D return to the safety of a space lander. Visitors see Dr. D in a mobile vehicle with the space lander on the opposite side of the screen. There are three levels in the game app-identical except for their visual setting (Moon, Mars, and Titan). With each question a player asks, Dr. D's vehicle moves forward a short distance across the screen before friction brings it to a halt. In response to feedback from visitors, we provided a small collection of badges and trophies in the game that are designed to reward players for completing specific sets of question-asking behaviors. For example, the "Rising Rookie" badge is awarded upon asking 5 questions.

Participants

There were 1539 participants in the final treatment group sample for this study. We defined "a participant" as one adult from a visitor group (of any group size, including those visiting on their own). While experimentally "messy" since an individual's responses are surely impacted by those in their group, this reflects the reality of doing research in an informal

museum setting. The data collected are analogous to traditional museum evaluation methods, which ask questions of an individual visitor, but cannot "control" for that visitor's responses or experiences being mediated by interactions within their group during the visit.

The majority of the participants were female (59%) and between 26 and 45 years of age (64%). The sample included 1495 participants who answered an ethnicity question, and described themselves as Native American/American Indian/Alaska Native (N = 23; 1.5%), Native Hawaiian/Other Pacific Islander (N = 9; 0.6%), Asian/Asian American (N = 69; 4.6%), Black or African American (N = 50; 3.3%), Hispanic or Latino (N = 259; 17.3%), White (Non-Hispanic) (N = 1000; 66.9%), and participants who selected two or more races (N = 85; 5.7%). Because of the uneven sample sizes, the race and ethnicity variable was examined as "White (Non-Hispanic)" participants and "Other Race".

Overall, 60% of participants were assigned to Game mode and 40% to Ask Mode. However, this breakdown was not consistent by museum as there were more Game Mode participants at ASC (64%) than at AZMNH: (52%), due to different scheduling availability for implementations across the two sites (Table 1).

Procedures

At each partner museum, we conducted implementations using Ask and Game Modes on an alternating schedule such that on a given day all visitors would be offered participation in only one of the conditions. Implementation periods generally ran 3–4 h each day. At each site, 13 iPad minis were available to lend to participants. The data collection team set up two tables in each museum, one at the entrance to the exhibit, where visitors were solicited to take part in the study and complete the pre-questionnaire and one at the exhibit exit, where visitors completed the post-questionnaire.

Each table contained a questionnaire (pre or post depending on the table), coloring sheets, crayons, informed consent forms, stickers, tattoos, project business cards, a "time in" log, a "time out" log, and a decline sheet. Researchers at the entrance table approached every visitor group containing at least one adult over the age of 18 to offer participation, following a standardized script. If the group declined, the general composition of the group was recorded (number of adults, youth, and babies/toddlers) and any relevant spoken comments were noted. If the group accepted, one adult in each group filled out the pre-questionnaire. After the paperwork, participants received a numbered iPad, with the number recorded on their questionnaire along with the time noted as they entered the exhibit. Participants were informed that the research team was testing an iPad app, designed to answer their questions as they go through the exhibit. The adult participant who completed the pre-survey was instructed to be the "question enterer" for the group. Other members of the group were invited to contribute questions to the question enterer. Participants were told they could stay in the exhibit as long as they wished. After the exhibit, participants

Table 1 Participant numbers by venue and app mode

	Ask Mode	Game Mode	Total
Arizona Science Center (ASC)	348 (36%)	631 (64%)	979
Arizona Natural History Museum (AZMNH)	266 (48%)	294 (52%)	560
Total	614	925	1539

returned the iPad to the exit table. Researchers recorded the exit time on the post-questionnaire before handing it to the adult participant for completion. Participant groups were provided with stickers and temporary tattoos in appreciation of their participation.

Measures

The pre-questionnaire contained a multi-part item about free-time science behaviors; smart device ownership, use, and comfort; museum membership and visitation; group size and composition; and standard demographic items. The post-questionnaire included questions about visitor interest in the exhibit, impressions of the Dr. D app, and three questions related to group participation in question-asking. The questionnaires were timestamped and collated by number, providing the total time spent in the exhibit by each visitor group. There were seven items that made up the free time science activity variable (adapted from Shavelson and Towne 2002): read news about science, read books about science, visit websites about science, watch TV about science, talk to others about science, visit science-related museums, and do science activities/projects. Reliability analysis was conducted to ensure that the seven items fit reliably to the scale, with results indicating a reliable fit ($\alpha = .86$) and therefore a "free time science" scale was created by averaging the means across all seven items.

The pre- and post-questionnaires were developed using closed- and open-ended items adapted from existing sources (e.g., Bickmore et al. 2011; Gehlbach et al. 2008; Shavelson and Towne 2002) and items developed by the team to inform the project's research questions. The questionnaires were piloted following standard testing and refinement procedures (e.g. Converse and Presser 1986; Dillman et al. 2014) including expert review and two rounds of pilot testing. Expert review (Olson 2010) was provided by a member of the project advisory board with expertise in museum research, by the project external evaluator, and by an external questionnaire design expert at another university. The first version of the questionnaire was piloted simultaneously with testing the project's data collection procedures for the first time, facilitating collection of feedback from the target population (museum visitors) ($n = 42$). From the responses, areas of misinterpretation were identified, and changes were made to the problematic items. Following changes, a second pilot of the questionnaire and data collection procedures was conducted ($n = 95$) to confirm the clarity of the items.

Analysis

For the current study, quantitative questionnaire data were analyzed within and across samples, by individual item and considering the relationship between items. Descriptive and inferential statistics inform the findings. Along with descriptive statistics, two-sample t-tests assuming equal variances were used to determine the significance of the collected data. The value of alpha used for all t-tests is .05. Categorical data were assigned numerical codes (for example, male coded to 1, female to 0).

Three different iterations of Game Mode were used in the data collection at different times. A one-way ANOVA test was conducted to see if there was a statistical difference in the data between the different game versions. The variable that was compared in the ANOVA test was "number of questions asked". The two museums (ASC and AZMNH) were compared separately for each game version (alpha equal to .05). The *p* value for ASC was 0.7181; the p-value for AZMNH is 0.1215. In both cases we fail to reject the null

Table 2 Correlation analysis for all study variables

Measure	1	2	3	4	5	6
1. App use	–					
2. Mode	.32**	–				
3. Museum	− .29**	− .12**	–			
4. FTS	.06*	.03	.00	–		
5. Race	− .01	.00	− .01	.09*	–	
6. Gender	− .02	.02	.05	− .07*	.03	–
Mean	7.74	.60	1.36	3.51	.67	.61
SD	6.90	.49	.48	.99	.47	.49

SD standard deviation, *FTS* free time in science, *Race* dummy coded variable

**p < .001 (2-tailed)

Table 3 Means and standard deviations of mode and museum

	Ask Mode M (SD)	Game Mode M (SD)
Arizona Science Center (ASC)	6.19 (6.13) N = 348	10.96 (7.56) N = 631
Arizona Natural History Museum (AZMNH)	3.53 (3.27) N = 266	6.46 (5.51) N = 294

hypothesis that the means of "number of questions asked" for each game version are relatively the same. This allows us to combine data for all three game versions.

Results

Correlation analyses were run to assess the overarching associations between all study variables (Table 2). Overall, the dependent variable (use of the app) was significantly associated with mode (Ask or Game), museum site, and free time spent in science activities. Thus, higher levels of app use were associated with Game Mode (as compared to Ask Mode), with participants at the Arizona Science Center (as compared to the Arizona Museum of Natural History), and with higher levels of reported free time spent engaged in science activities. Further, the race and gender variables were significantly associated with reported free time science activities, with higher levels of engagement in free time science activities reported by White (Non-Hispanic) and male participants.

To investigate the first research question about the difference in question-asking levels between Ask and Game Mode participants, descriptive statistics were examined, and a one-way ANOVA was conducted to determine the mean differences in the number of questions asked across Ask and Game Modes. The results showed a significant difference across modes, $F(1,1537) = 173.78$, $p < .001$, indicating that, on average, a higher number of questions were asked by participants who used the Game Mode (M = 9.53) compared to Ask Mode (M = 5.04).

To investigate the second research question exploring whether level of app use, overall and by app mode, varies by museum, descriptive statistics were examined (Table 3),

indicating that Game Mode participants asked more questions on average than Ask Mode participants at both museums. A two-way ANOVA showed a statistically significant interaction between the effects of venue and mode on app use, $F(1,1535) = 7.345$, $p = .007$. Simple main effects analysis showed that, although Game Mode participants asked significantly more questions than Ask Mode participants at both museums, the strength of the difference was somewhat stronger at ASC, $F(1977) = 101.697$, $p < .0001$, than at AZMNH, $F(1558) = 57.113$, $p < .0001$.

To explore the third research question on the extent to which app use varied by self-reported level of engagement in science-related activities, both overall and by mode (Ask and Game), a regression analysis was conducted to determine whether there was a significant interaction between the free time spent in science-related activities and app mode as they relate to the number of questions asked.

The first step in this analysis included centering the independent variables. Next, the regression analysis was run to determine whether there were significant main effects and/or interaction effects, while controlling for race and gender. The results indicated two significant main effects, but no interaction ($\beta = 0.00$, $p = .96$) and no associations between race and gender and the amount of questions asked. However, both amounts of reported free time science activity ($\beta = 0.06$, $p < .05$) and mode ($\beta = 0.31$, $p < .001$) positively predicted the amount of questions asked (see Table 4).

The same regression analyses were run separately for each of the museums. For the ASC, there was a main effect such that Game Mode positively predicted the amount of questions asked ($\beta = .31$, $p < .001$), however, there was not a main effect for reported free time science activity, nor was there a significant interaction. The results for the AZMNH were similar to the results across all participants. Specifically, both reported free time science activity ($\beta = .10$, $p = .02$) and assigned app mode ($\beta = .31$, $p < .001$) positively predicted the number of questions asked. Neither race nor gender were significant predictors in either of these models (see Table 5).

As a post hoc analysis, a one-way ANOVA was conducted to explore the relationship between mode (Ask, Game) and time spent in the exhibit as an alternative approach to measuring visitor engagement. Overall, we saw a significant difference, with Game Mode participants spending more time on average (16.29 min) than Ask Mode participants (14.77 min), $F(1,1537) = 11.074$, $p = .001$. However, the difference in time spent in the exhibit by Game and Ask mode participants varied by museum. A two-way ANOVA showed a statistically significant interaction between the effects of Mode and Venue on time spent in the exhibit, $F(1,1535) = 5.229$, $p = .022$. Simple main effects analysis showed that Game Mode participants at AZMNH spent significantly more time in the exhibit than Ask Mode participants, $F(1558) = 11.515$, $p = .001$, but at ASC there were no differences in the time spent in the exhibit by Mode, $F(1977) = .000$, $p = .986$.

Table 4 Regression analysis for Dr. Discovery app use across all participants

Variable	Standardized coefficients		
	β	*t*	*p*
Mode	0.31	12.63	.000
FTS	0.06	2.28	.023
Interaction	0.00	0.05	.961
Race	−0.02	−0.82	.415
Gender	−0.02	−0.94	.3444

Table 5 Regression analysis for Dr. Discovery app use: by museum

Variable	Standardized coefficients			
	β	*t*	Sig. (p)	*d*
ASC				
Mode	0.30	9.44	.000	.693
FTS	0.04	1.29	.197	–
Interaction	0.01	0.03	.976	–
Race	−0.04	−1.17	.242	–
Gender	−0.02	−0.78	.437	–
AZMNH				
Mode	0.32	7.72	.000	.646
FTS	0.10	2.44	.015	–
Interaction	−0.01	−0.16	.873	–
Race	0.04	0.85	.394	–
Gender	0.02	0.46	.642	–

Discussion

Overall, exposure to Game Mode was the strongest predictor of app use, and all the variables combined accounted for approximately 25% of the variance. In both museums, participants using Game Mode asked significantly more questions on average than those using Ask Mode. While the frequency of reported free time science activities was positively associated with the number of questions asked, it seemed less influential when including app mode in the same regression model. Interestingly, frequency of engagement in free time science activities was still a positive predictor at the AZMNH.

Exploring time in the exhibit as an alternative measure of participant engagement revealed mixed results. While Game Mode participants spent more time overall in the exhibit than Ask Mode participants, the difference was only seen in one of the two museums (AZMNH). Although we recorded start and end time for each participant group, we have no way of determining how the time in exhibit was spent. In particular, we cannot know how participants divided up their time while in the exhibit: for example, how much time they spent looking at exhibit items vs. contemplating, discussing, typing questions, and reading the provided answers. In other words, time in the exhibit may not be as accurate a measure of engagement as question-asking levels. It is interesting that at ASC, the mean time spent in the museum was the same for Ask and Game Mode participants, but the mean number of questions asked was significantly higher for Game Mode participants. Even though both groups spent roughly the same time in the exhibit on average, a larger proportion of Game Mode participants' time was spent entering questions.

In considering the implications of these findings, it is important to restate the primary goal of the Dr. Discovery project: collecting, analyzing, and interpreting data on museum visitor questions to support museum evaluation efforts. It is clear that data gathered about visitor questions related to exhibit content does not, by itself, constitute a full program of museum evaluation. However, such data can bolster evaluation efforts, particularly in the majority of museums that have no staff conducting visitor evaluation (Smithsonian 2004). Further, the more data collected (i.e., the more questions visitors ask), the more useful the data will be in uncovering trends in visitor interests, areas of confusion, and interaction (Ha et al. in preparation).

As the gathering of question data is a primary concern of the Dr. D project, it is important to consider the practical "bang for the buck" impact of such an app on question-asking. Considering Game Mode participants in the current study asked significantly more questions on average than Ask Mode users, does an average of between 5 and 10 questions per user provide enough evaluative information to be worth the expense of developing and implementing such a tool? We believe it does. In aggregate, the number of questions asked adds up (more than 11,000 questions were asked overall in the current study). Given the dearth of on-staff evaluation personnel at most museums (Smithsonian 2004), data gleaned from thousands of up-to-date visitor questions, specific to exhibit content, should be very useful in supporting evaluation efforts.

The findings in this study showing the Game Mode version of the Dr. D app to be associated with significantly higher levels of question-asking leads us to conclude that designers of similar data-gathering tools in informal science venues such as museums and science centers may want to incorporate simple game mechanics into their designs to bolster use. These simple game mechanics may lead to increased visitor engagement with the content of their setting if the tool is designed to incorporate game mechanics. Our findings on increased levels of visitor engagement (as reflected through levels of question asking) associated with Game Mode echo findings from our related work on the impact of educational games in formal settings (Clark et al. 2016; Martinez-Garza et al. 2013; Nelson et al. 2011, 2016), and that of other game researchers (e.g. Annetta et al. 2009; Ke 2008; Barab et al. 2005). Educational games have been consistently shown to foster high levels of user engagement (Barab et al. 2005; Clark et al. 2009), even as the record on increased learning with games is mixed (Martinez-Garza et al. 2013).

Further, our rounds of user testing on the design of the Dr. D Game mode suggest that museum researchers and app designers should focus on keeping game elements as simple as possible, and on tightly integrating game goals with actions that maintain visitor focus on exhibit items (Nelson et al. 2017).

Limitations

Assignment to one of the two app modes (Game or Ask) was not fully randomized (different modes were implemented on different days on a rotating basis). However, care was taken to assure that each mode was implemented on varying days of the week (to avoid having one mode always implemented on a weekend day, for example). Also, as in any research with voluntary participants, there is the possibility of nonresponse error, "which occurs when the characteristics of the respondents differ from those who chose not to respond in a way that is relevant to the study results" (Dillman et al. 2014, p. 5), so the reported results are only representative of visitors who chose to participate in the study.

Future directions and conclusions

The analysis described in this paper provides an in-depth, but still preliminary step in understanding the impact of the Dr. D Game mode, and other similar software, on visitor engagement in science museums. While visitors clearly asked more questions, on average, when using the Game mode version of the Dr. D app, we have yet to analyze qualitative data collected that may clarify the extent to which use of the Game Mode may draw visitor attention away from viewing and interacting with exhibit content. As we have described

elsewhere (Nelson et al. 2017), with the Dr. Discovery project, we sought to design a simple game that was engaging enough to promote sustained, active use by visitors, but not so engaging that it became the focus of visitor attention. An ongoing and critical challenge for museum apps such as Dr. D is the impact their use could have on drawing visitors' attention out of the physical experience of a museum exhibit and directing it into staring at a screen. As Rung and Laursen (2012) describe, digital devices and apps may lead visitors to "…spend all their time looking at a poky little screen instead of the art" (p. 322), which museums and researchers must keep in mind as they introduce new mobile experiences. Our future studies will formalize our exploration of how to design casual games for use in museum settings to find the best balance between game and exhibit engagement.

Another area for study that can shed light on the role of question-asking apps such as Dr. D on visitor engagement in museums centers on understanding if and how participant groups interact with each other while using them. In a separate study (Ha et al. in preparation), we are investigating this topic. Our post-questionnaire asked participants about how their groups went about asking questions with the app, probing such issues as how often group members discussed which questions to ask, who in the group asked the most questions, and how use of the Dr. D app affected the group's interactions with each other and with museum personnel. Understanding such focused qualitative stories can add nuance to the basic findings supportive of game-based question-asking apps reported here. For example, group-level information can provide insights on how to better design similar apps to foster more and higher quality collaboration in the question-asking process.

Finally, given the focus of the Dr. D study on engagement and museum evaluation, we did not collect data on participant-visitors' learning gains. Therefore, we cannot make claims of Dr. D's effect on participant-visitor's learning. Our future studies will build on the findings reported here to investigate the relationship between engagement with Dr. D and informal learning for its users.

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

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abraham-Silver, L. (2006). Valuing informal science learning environments. *The Science Teacher, 73*(1), 1–12.

Adams, M. (2012). Where have we been? What has changed? And where do we need to go next? *Journal of Museum Education, 37*(2), 25–36.

Annetta, L., Minogue, J., Holmes, S., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*(1), 74–85.

Apley, A., Frankel, S., Goldman, E., & Streitburger, K. (2011). Supporting Museums–serving communities: An evaluation of the Museums for America Program. Full Report. *Institute of Museum and Library Services*.

Barab, S., Arici, A., & Jackson, C. (2005). Eat your vegetables and do your homework: A design-based investigation of enjoyment and meaning in learning. *Educational technology, 65,* 15–21.

Bickmore, T., Pfeifer, L., & Schulman, D. (2011). Relational agents improve engagement and learning in science museum visitors. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents, IVA'11* (pp. 55–67). Berlin, Heidelberg: Springer.

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.
Bogost, I. (2011). *How to do things with videogames*. Minneapolis, MN: University of Minnesota Press.
Cassell, J., & Jenkins, H. (Eds.). (2000). *From barbie to mortal kombat*. Cambridge, MA: The MIT Press.
Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: genres, examples, and evidence. An NAS commissioned paper. Learning Science: Computer games, simulations, and education.
Clark, D. B., Virk, S., Sengupta, P., Brady, C., Martinez-Garza, M., Krinks, K., et al. (2016). SURGE's evolution deeper into formal representations: The siren's call of popular game-play mechanics. *International Journal of Designs for Learning., 7*(1), 107–146.
Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage Publications.
Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Education Training Research and Development, 53*(2), 67–83.
Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley.
Edery, D., & Mollick, E. (2009). *Changing the Game: How video games are transforming the future of business*. New Jersey: FT Press.
Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology, 16*(2), 171–190.
Falk, J. H., Storksdieck, M., & Dierking, L. D. (2007). Investigating public science interest and understanding: evidence for the importance of free-choice learning. *Public Understanding of Science, 16*(4), 455–469.
Federation of American Scientists. (2006). *Summit on educational games: Harnessing the power of video games for learning*. Washington, DC. Retrieved from https://fas.org/programs/ltp/policy_and_publications/summit/Summit%20on%20Educational%20Games.pdf.
Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., & Yuhnke, B. (2016). *NMC Horizon report: 2016* (Museum ed.). Austin, Texas: The New Media Consortium.
Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.
Gehlbach, H., Brown, S. W., Ioannou, A., Boyer, M. A., Hudson, N., Niv-Solomon, A.,… & Janik, L. (2008). Increasing interest in social studies: Social perspective taking and self-efficacy in stimulating simulations. *Contemporary Educational Psychology*, *33*(4), 894–914.
Gerber, B. L., Cavallo, A. M., & Marek, E. A. (2001). Relationships among informal learning environments, teaching procedures and scientific reasoning ability. *International Journal of Science Education, 23*(5), 535–549.
Ha, J., Pérez Cortés, L., Su, M., Nelson, B., Bowman, C., & Bowman, J. (in preparation). *Group inquiry and engagement in museums using ask Dr. Discovery.*
Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*(4), 910–931.
Hayes, E. (2005). Women and videogaming: Gendered identities at play. Presented at the Games, Learning, and Society Conference, Madison, WI.
Hayes, E. (2008). Game content creation and IT proficiency: An exploratory study. *Computers & Education, 51*(1), 97–108.
Smithsonian Institution. (2004). *The evaluation of museum educational programs: A national perspective*. Washington, DC: Office of Policy and Analysis, Smithsonian Institution.
Jenkins, H., Squire, K., & Tan, P. (2003). Entering the education arcade. *ACM Computers in Entertainment, 1*(1), 1–11.
Johns, J., & Woolf, B. (2006, July). A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the National Conference on Artificial* Vol. 21, No. 1, p. 163). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
Johnson, L., Adams, S., & Witchey, H. (2011). *The NMC horizon report: 2011* (Museum ed.). Austin, TX: The New Media Consortium.
Johnson, L., Becker, S. A., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 higher* (education ed., pp. 1–50). Austin: The New Media Consortium.
Joseph, E. (2005). Engagement tracing: using response times to model student disengagement. *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology, 125,* 88.
Kafai, Y. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture, 1*(1), 36–40.

Katz, J. E., Goldman, K. H., & Foutz, S. (2011). *Mobile phones for informal science center learning: A socio-technical analysis*. na.
Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education, 51*(4), 1609–1620.
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. AIED 2013. Lecture notes in computer science, vol 7926*. Berlin: Springer.
Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333–369.
Martinez-Garza, M., Clark, D. B., & Nelson, B. (2013). Digital games and science learning: Research across the NRC strands of science proficiency. *Studies in Science Education, 49*(2), 170–208.
McGonigal, J. (2011). *Reality is broken*. New York: Penguin Books.
Nelson, B., 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.
Nelson, B., Erlandson, B., & Denham, A. (2011). Global channels for learning and assessment in complex game environments. *British Journal of Educational Technology, 42*(1), 88–100.
Nelson, B., Kim, Y., & Slack, K. (2016). Visual signaling in a high-search virtual world-based assessment: A SAVE Science design study. *Technology, Knowledge, and Learning., 21*(2), 211–224.
Olson, K. (2010). An examination of questionnaire evaluation by expert reviewers. *Field Methods, 22*(4), 295–318.
Pew Research Center. (2018). *Mobile Fact Sheet*. Retrieved July 16, 2018 from http://www.pewinternet.org/fact-sheet/mobile/.
Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
Proctor, N. (2015). Mobile in museums: From interpretation to conversation. *The International Handbooks of Museum Studies, 4*(22), 499–525.
Ramsey-Gassert, L. (1997). Learning science beyond the classroom. *The Elementary School Journal, 97*(4), 434–449.
Røtne, O., & Kaptelinin, V. (2013). Design choices and museum experience: A design-based study of a mobile museum app. In *Human computer interaction* (pp. 9–13). Springer, Berlin.
Rung, M. H., & Laursen, D. (2012). Adding to the experience: Use of smartphone applications by museum visitors. In *The transformative museum proceedings*, (pp. 314–324).
Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave McMillan.
Shavelson, R. J., & Towne, L. (2002). Scientific research in education. Committee on scientific principles for education research. *Center for Education. Division of Behavioral and Social Sciences and Education. National Research Council*. Washington, DC: National Academy Press.
Squire, K. (2011). *Video games and learning: Teaching and participatory culture in the digital age*. New York: Teachers College Press.
Steinkuehler, C. (2004). The literacy practices of massively multiplayer online gaming. Paper presented at the American Educational Research Association, San Diego, CA.
Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science and Technology, 17*(6), 530–543.
Vadeboncoeur, J. A. (2006). Engaging young people: Learning in informal contexts. *Research in Education, 30,* 239–278.
Wellington, J. (1990). Formal and informal learning in science: Role of the interactive science centers. *Physics Education, 25,* 247–253.

Brian C. Nelson is a professor of educational technology in the Mary Lou Fulton Teachers College at Arizona State University.

Catherine D. D. Bowman is an associate research professor in the School of Earth and Space Exploration at Arizona State University.

Judd Bowman is a professor of astronomy in the School of Earth and Space Exploration at Arizona State University.

Luis E. Pérez Cortés is a PhD student in the Learning, Literacies, and Technologies program in the Mary Lou Fulton Teachers College at Arizona State University.

Adrianna Adkins is an undergraduate student of mathematics and statistics at Arizona State University.

Edgar Escalante is an undergraduate student of physics at Arizona State University.

Brooke L. Owen is an undergraduate student of biology at Arizona State University.

Jesse Ha is a PhD student in the Learning, Literacies, and Technologies program in the Mary Lou Fulton Teachers College at Arizona State University.

Man Su is a PhD student in the Learning, Literacies, and Technologies program in the Mary Lou Fulton Teachers College at Arizona State University.