

U!Scientist: Designing for People-Powered Research in Museums

Mmachi God'sgloy Obiorah
Northwestern University
Evanston, IL
obiorahm@u.northwestern.edu

Will Granger
Zooniverse & Adler Planetarium
Chicago, IL
will@zooniverse.org

James K.L. Hammerman
TERC
Cambridge, MA
jim_hammerman@terc.edu

Haley Margaret West
Northwestern University
Evanston, IL
haleywest2019@u.northwestern.edu

Becky Rother
Zooniverse & Adler Planetarium
Chicago, IL
becky@zooniverse.org

Michael Horn
Northwestern University
Evanston, IL
michael-horn@northwestern.edu

Laura Trouille
Zooniverse & Adler Planetarium
Chicago, IL
lTrouille@adlerplanetarium.org



Figure 1: Visitors classifying galaxies with U!Scientist.

ABSTRACT

Scientists have long sought to engage public audiences in research through citizen science projects such as biological surveys or distributed data collection. Recent online platforms have expanded the scope of what people-powered research can mean. Science museums are unique cultural institutions that translate scientific discovery for public audiences, while conducting research of their own. This makes museums compelling sites for engaging audiences

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '21, May 8–13, 2021, Yokohama, Japan
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8096-6/21/05...\$15.00
<https://doi.org/10.1145/3411764.3445334>

directly in scientific research, but there are associated challenges as well. This project engages public audiences in contributing to real research as part of their visit to a museum. We present the design and evaluation of U!Scientist, an interactive multi-person tabletop exhibit based on the online Zooniverse project, Galaxy Zoo. We installed U!Scientist in a planetarium and collected video, computer logs, naturalistic observations, and surveys with visitors. Our findings demonstrate the potential of exhibits to engage new audiences in collaborative scientific discussions as part of people-powered research.

CCS CONCEPTS

- Human-centered computing → Field studies; Collaborative and social computing systems and tools.

KEYWORDS

Citizen science; museums; interactive tabletop displays

ACM Reference Format:

Mmachi God'sglory Obiorah, James K.L. Hammerman, Becky Rother, Will Granger, Haley Margaret West, Michael Horn, and Laura Trouille. 2021. U!Scientist: Designing for People-Powered Research in Museums. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3411764.3445334>

1 INTRODUCTION

The practice of science is changing. Computational tools and methods have empowered scientists to collect massive amounts of data in the pursuit of groundbreaking new research. At the same time, connectivity has distributed the practice of science across larger teams around the globe. However, even with the power of technology to collect and process large amounts of data, there are many types of analysis for which humans must still be involved (or are necessary to generate sufficiently large training sets for classification algorithms). This has led to crowdsourcing strategies in which large numbers of distributed volunteers use online platforms to collaborate on research tasks. For example, the Galaxy Zoo project on the popular platform, Zooniverse.org, asks volunteers to classify telescope data containing images of galaxies. With the help of non-scientist volunteers distributed around the world, the project has been able to classify hundreds of thousands of galaxies in a fraction of the time that it would have taken teams of trained scientists. Not only was the project successful from a research perspective, it also helped everyday people learn about and contribute to cutting-edge science.

As the potential of platforms such as Zooniverse grow, organizers are faced with the new challenge of engaging broader audiences in citizen science research [32, 35]. In this paper, we take up this challenge by focusing on museums and other institutions of informal science learning. Museums are not only tasked with communicating science to the public, they also have scientific missions of their own. By bringing together scientists, collections, historical artifacts, exhibits, and other educational programming, they can be ideal sites to invite new participants into new science-themed activities. At the same time, interaction with exhibits in museums looks much different than it does for online web content. Dwell times at exhibit elements tend to be very brief (from a few seconds to a few minutes [19]), and it can be difficult to convey the same level of information that would typically be available on a website. Learning in museums is also a social activity, with visitors collaborating around exhibits and discussing their shared experiences.

Thus, this project explores the following research questions:

- (1) How can we design a museum exhibit to engage visitors in people-powered research?
- (2) How can the exhibit support collaborative discussions around scientific ideas?

To answer these questions, we developed an interactive exhibit called U!Scientist for visitors to contribute to research as part of their visit to a planetarium. We built U!Scientist to run on an interactive tabletop display that we installed in a busy planetarium gallery. In crafting our design, we drew from existing research that indicates that large digital displays in museums have the potential to engage users, create new forms of audience participation, support collaboration, and support learning through exploration

[2, 3, 16, 18, 25, 30, 30]. We based our exhibit on the Galaxy Zoo project from Zooniverse [36]. The Zooniverse platform provides a robust infrastructure for people-powered research, and the Galaxy Zoo project was ideal for our planetarium context.

To help motivate engagement, we carefully designed informational text to communicate that visitors were participating in real scientific research, and that they were collaborating with many other volunteers to come to consensus about each galaxy's classification. We carried out usability testing with low-fidelity prototypes to refine both our messaging and user interaction flow.

We then conducted two studies of visitor interaction with the exhibit. The first study involved 4.2 hours of video recording of 82 consented visitors as they interacted with the exhibit. Participants completed a post-survey, and we kept computer logs of participants' interaction with the exhibit. We analyzed these data qualitatively and quantitatively. The second study was a naturalistic observation of 192 visitors who interacted with the touch table over the course of about 5.5 hours on two consecutive days. We collected arrival and departure times, noted approximate demographics and correlated these with the automatic table logs of visitor behavior. Finally, we conducted surveys of additional visitors before (N=132) and after (N=182) experiencing the exhibit, eliciting views about science and citizen science, and their ability to classify a galaxy. Our analyses suggests that our design was able to engage users and convey to them that they were contributing to real science. Our findings further show that museums can be effective spaces to both engage younger audiences and to support co-located collaboration in people-powered research activities.

In the following sections of this paper we review background research, provide an overview of our design process, and present the final design of U!Scientist. We then describe data collection and analysis of video recordings, naturalistic observations, and surveys. We organize our findings based on visitors' engagement and collaboration. Finally, we reflect on how our findings might inform the future of people-powered research in museums.

2 BACKGROUND

2.1 Galaxy Zoo

Zooniverse.org is an online platform that hosts a wide variety of citizen science research projects [36]. Volunteers with no special expertise or training are able to contribute to these projects by answering questions about authentic research data such as images of galaxies or videos of animals in their natural habitats [14]. Zooniverse makes small tasks in which humans outperform machines available to a large number of volunteers thereby harnessing the power of the crowd and accelerating research. By having a number of volunteers do the same work, the overall accuracy of the group's contribution is improved; and results are used to train computers to be better at the same or similar tasks [9, 24].

Galaxy Zoo is one such project hosted on the Zooniverse platform. In Galaxy Zoo, volunteers are presented with images of various sky surveys and are asked to answer a series of short questions about the shape of galaxies to help researchers learn about astrophysical processes in galaxy evolution (see figure 2). Multiple volunteers look at each image, and the consensus result (the modal response) generally matches an expert's conclusion about 98% of

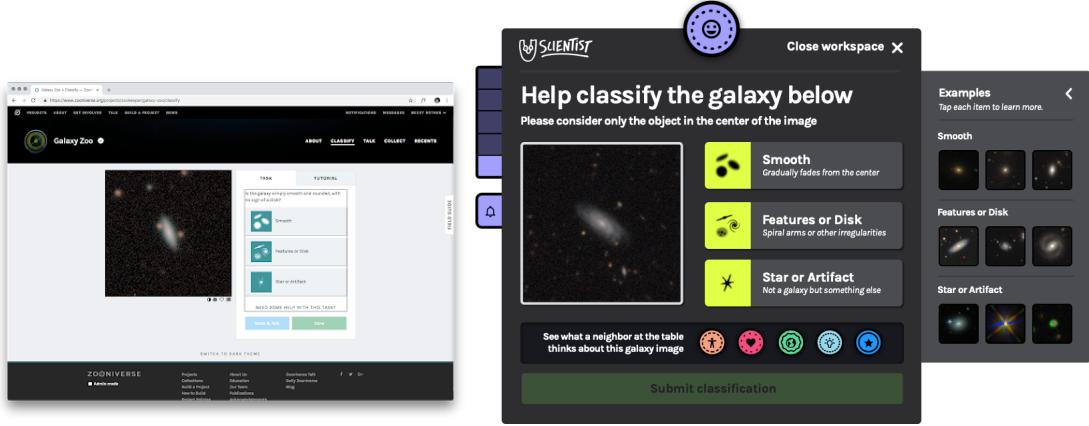


Figure 2: GalaxyZoo.org. In the online version of Galaxy Zoo (left panel), volunteers are presented with images from various sky surveys and are asked to answer a series of short questions about the shape of a galaxy to help researchers learn about astrophysical processes in galaxy evolution. U!Scientist Single Station. In a shortened version of the online task (right panel), users can classify galaxies as "smooth", "features or disk" or, as a "star or artifact". They can also request help from other stations by tapping on one of the icons corresponding to the station color from which they want help. Each station has an example panel to the right and a game level indicator to the left.

the time [23]. The task lends itself to crowd work because of the very large sample of galaxies usually required for understanding galactic processes.

Because Galaxy Zoo can be accessed online, it can offer detailed information about the classification tasks that people are asked to perform. There is also instructional information intended to ease concerns about making mistakes or contributing incorrect classifications. Volunteers carry out up to six classification tasks for a single galaxy. For instance, one path of questions would be "Is the galaxy simply smooth and rounded, with no sign of a disk?", "Could this be a disk viewed edge on?", "Does the galaxy have a bulge at its center? If so what shape?", "Is the galaxy merging or distributed?", "Do you see any of these rare features?" Galaxy Zoo also has several examples to help users classify a galaxy. As an online project, Galaxy Zoo has attracted thousands of volunteers who have classified hundreds of thousands of galaxies, which would have taken individual scientists dozens of years [13]. Galaxy Zoo images are compelling enough to rouse the curiosity of users, thus suggesting potential advantages for use in museum spaces.

Galaxy Zoo project contributors are motivated by contributing to real science and by personal interest [11, 29]. Research has shown that volunteers can be incentivized to contribute to crowd work if they feel like co-owners [35]. This motivation to volunteer for Galaxy Zoo and citizen science projects in general is well aligned with shifting perspectives of museum visitors as participants in the creation of knowledge and not only consumers of information [19, 22, 27, 34]. As a result, we were optimistic about introducing this citizen science project as a museum exhibit.

Although Zooniverse volunteers are able to talk to each other virtually [31], typically a citizen science project like Galaxy Zoo with well-structured and low-complexity tasks does not require volunteers to collaborate in order to complete a task [28]. A citizen science project as a museum exhibit brings volunteers side by

side and face to face with each other. We were interested in the interactions that would arise as a result of this new configuration.

2.2 Multi-Touch Tabletops

Digital media in public spaces have been studied extensively. Research has shown that interactive digital media in museums can support collaborative interaction and learning through exploration [2, 3, 5, 16, 17, 25, 26]. In addition, digital media in museums have the potential to extend the ways in which museum visitors are able to interact with museum artifacts and knowledge. For example, they can include gamified elements [2, 10, 16], novel forms of interaction like virtual or augmented reality [3, 4], or simply allow for different paths of exploration of artifacts. In spite of the many affordances of digital media in museum spaces, they are sometimes at odds with parents'/ guardians' intent in bringing their children/ wards to museum spaces. Many parents intend for their time in museum spaces to be a break from digital media [16]. As such, researchers have investigated how to best design digital exhibits in museums [30]. For example, Horn, et al. measured visitors' engagement and learning using a gamified exhibit against APE (Active Prolonged Engagement) recommendations [16, 19]. They suggest that their gamified exhibit allowed visitors to employ social practices and *learning arrangements* that result in successful collaboration.

Multi-touch tabletops allow several users to interact with a digital exhibit at a time. Though actions of visitors on the table might be independent or collaborative, multi-touch tabletops are popular for their ability to support collaboration. However collaborative exhibits on multi-touch tables require careful design. Allen, et al. found that in order for collaborative exhibits to be successful, they should be designed so that multiple interactive elements do not overwhelm users, different user interactions are not disruptive to other users, and secondary interactive elements do not overwhelm primary ones [1].

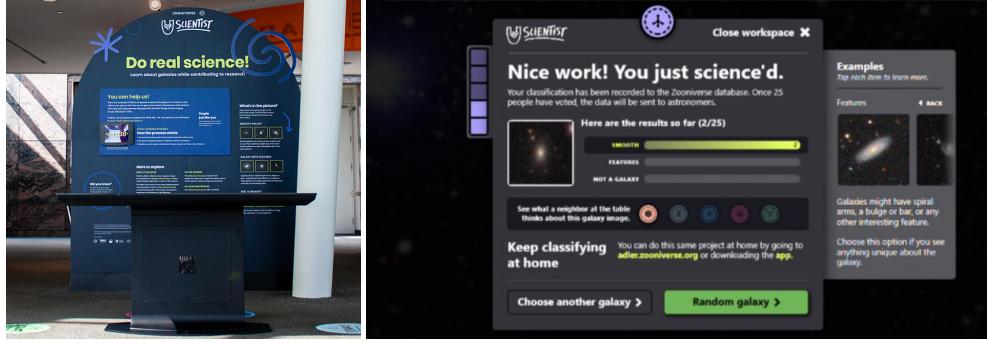


Figure 3: Exhibit Setup in the Planetarium (left). U!Scientist Summary Page. A summary page is displayed to the user with bars that indicate how other people classified the same galaxy (right).

In this study, we draw on the collaborative and engaging affordances of a multi-touch tabletop to design a people-powered research exhibit. We also include subtle game elements to take advantage of social practices associated with them as well as engage users.

2.3 Engaging Museum Visitors

In this study, we draw from the concept of Active Prolonged Engagement (APE) with museum exhibits [19]. APE exhibits support prolonged engagement by evoking inquiry and response from visitors with the intent to support them in making connections that lead to learning. Large multi-touch tabletops support multiple users on an exhibit thereby providing a platform for collaboration which can give rise to the question asking and response behaviors that result in learning. As such, we designed our exhibit on a multi-touch tabletop. In addition, the task of classifying galaxies through Galaxy Zoo lends itself to collaboration as ten volunteers are needed to come to a consensus on a classification before it is accepted into the data analysis. The planetarium where this study was carried out also provides a physical space for volunteers to interact and collaborate on classifying galaxies. APE recommendations are also geared towards shifting the role of museum visitors from recipients of instruction and explanation to that of participants. The task of classifying galaxies seems suitable for active prolonged engagement in that it is somewhat open-ended, there are no right or wrong answers, the task takes the form of inquiry, and visitors are directly participating in authentic research.

3 DESIGN OF U!SCIENTIST

In this section, we describe our preliminary studies, insights from them, and our final U!Scientist design.

3.1 Preliminary Studies

To better understand the design space we were working in, we undertook a series of preliminary studies with our target museum audience at the planetarium.

3.1.1 Messaging Study. Messaging was an essential part of the design of this exhibit. It was critical to convey that interacting with the exhibit meant participating in real research and helping

real scientists, but also that it would be fun and rewarding for museum guests. To that end, we created a poster with four possible messages and asked guests to place a sticker on the message that most resonated with them. The four messages were: "Participate in Real Research", "Learn about Galaxies", "As a community, we can classify galaxies", and "You can help researchers". We tested with about 50 museum guests over two days in July 2018, with the result a tie between "Learn about galaxies" and "Participate in real research". With that result, we decided to combine the messaging: "Learn about galaxies while contributing to research." This message would be used in future studies as well as on the wall exhibit accompanying the touch table.

3.1.2 Paper Prototyping. To determine the most successful user flow for the interaction, we used a series of paper prototypes to test with guests. We built a cardboard replica of the touch table and created sketched versions of the proposed interface on paper. Guests were asked to "interact" with the paper prototypes in the way they might interact with a touch screen, moderated by two of our team members. One team member would "drive" the prototype, moving the paper "screens" to simulate the on-screen interactions, while the other team member observed and asked questions. This method was highly effective to quickly test multiple user flows, and through observation we made several crucial adjustments that contribute to the efficacy of the exhibit.

First, we found that while visitors to a planetarium might have an interest in space science, they do not necessarily have a familiarity with the associated terminology. Many guests declined to participate because they did not feel confident in their knowledge of galaxy types. To mitigate this trepidation, we added more visible help text and example images. This is consistent with studies that show that newcomers struggle with language necessary to communicate in citizen science discussion forums [15].

Additionally, we were keen to test the second screen in the user flow: After the user submits their classification ("Smooth", "Features or Disk", or "Star or Artifact") they are shown a bar chart showing how other users classified the galaxy. Would guests be able to interpret the chart and understand that they were submitting a "vote" rather than answering a quiz question? Through observation during the testing phase we found that this visual representation of that concept was fairly successful.

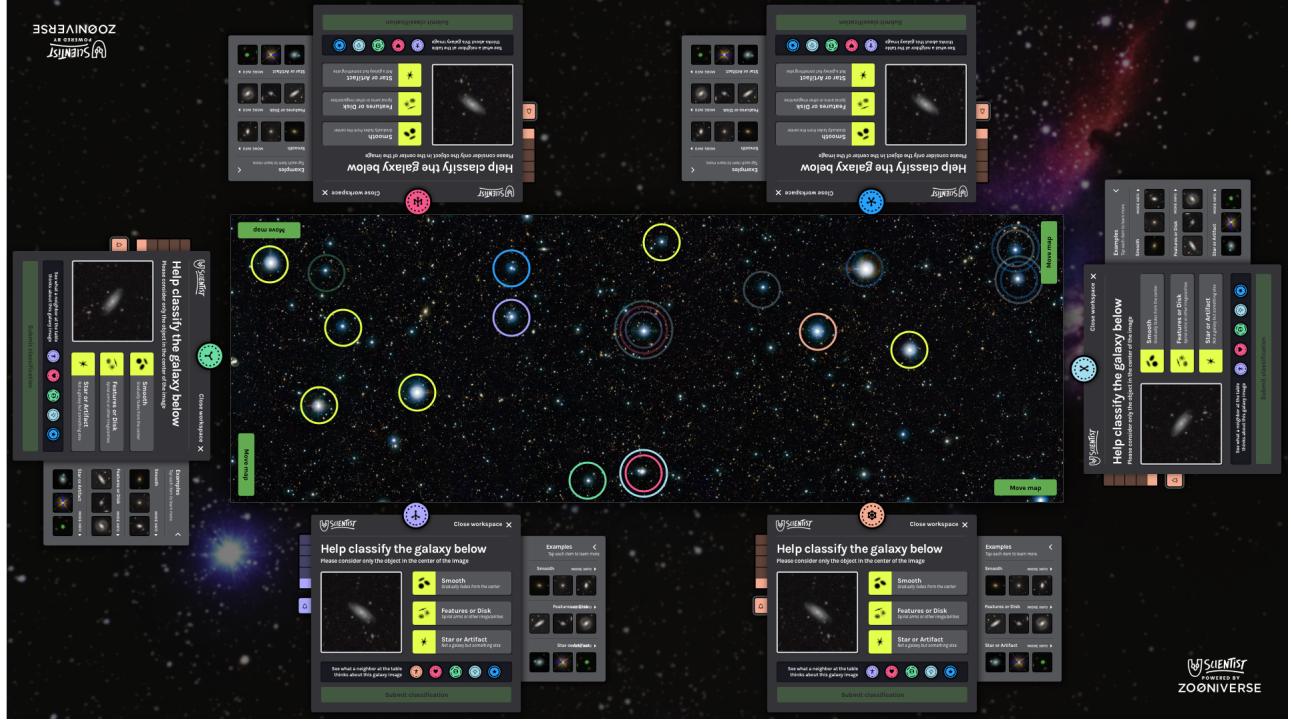


Figure 4: U!Scientist's Interface. The interface has 6 stations. Galaxies are enclosed within circles in the center of the screen and visitors can drag them from the center of the screen into their stations to classify them. Any user can tap the "Move Map" button to retrieve a new set of galaxies to classify. Colored circles around the galaxies indicate two things: 1. That the galaxy is still available for classification and; 2. that another visitor with the corresponding workspace color is classifying the same galaxy.

Finally, we determined that the original Galaxy Zoo workflow would be too lengthy to be effective in a museum setting. The online workflow has multiple questions and can take several minutes to classify a single image. In the museum, we wanted to be respectful of guests' time – there are many things in the planetarium to draw attention – so creating a task that could only take a few seconds was important. After speaking with the Galaxy Zoo research team, we determined that the first question in the workflow ("Is the galaxy simply smooth and rounded, with no sign of a disk?") would still be useful for research purposes and also be short enough to give guests a sense of accomplishment in a short period of time. Using the data from these preliminary prototypes, we moved into high fidelity prototyping.

3.1.3 Prototype Testing. Our team then built an interactive prototype using an online design tool and loaded it onto the touch table using a Chrome extension that simulated a real museum interactive by preventing guests from navigating away from the prototype to other websites or apps on the touch table. Over the course of about a month in 2018, the team observed museum guests interacting with the prototype, validating decisions from the initial design phase. This prototype also helped communicate design intent with the developer, who was able to replicate visual style and animations.

Through early discussions, we determined that the target age range would be about 12 years and up, so we paid special attention to be sure the language was simple enough and the interactive

made sense and was easy to use. Another goal of the project was to encourage museum guests to engage in evidence-based conversation with one another. Through the reference images and help text, we hoped to provide useful language for those conversations. We also took advantage of the touch table's large physical dimensions and designed the exhibit to accommodate many people or groups of people at once. The interface has six stations (see figure 4) arranged around the table and distinguished by six different colors and icons. Keeping accessibility and colorblindness in mind, we made sure to not rely solely on color to distinguish work spaces.

3.2 Final Design

Since the final museum exhibit would be used by visitors on their own without an external guide or instructor, it was essential that the interactive was easy to use for the majority of museum guests without verbal explanation. In the final version of U!Scientist, the visitor was presented with a single task (see figure 4) and many examples to help them quickly become comfortable with the interactive exhibit. From any of its six stations, users could drag a galaxy from the center of the table and drop it in their station to classify it. Galaxies could be classified as "Smooth", "Features or Disks" or "Star or Artifact". The exhibit had an example panel that showed users what the different classification categories looked like.

Users could also collaborate with other people at the table by clicking one of the colored icons (see figure 2) in their stations.

The colored icons were representative of the different stations at the table. Users' requests for help would pop up as a message on the receivers' station. The receiver could either be busy, accept the help request, or decline the help request. If they agreed to help, they would be shown the galaxy in their own station and make a classification. Their classification would then be sent back to the requester. The user could also change the set of classifiable galaxies by tapping the "move map" buttons (see figure 4). Gamification of the exhibit was minimal - we added five "levels" each representing five galaxies classified. The transition to the next level was indicated by a meter (see figure 2).

After classifying a galaxy, a summary screen displayed a chart indicating how many people selected the different classification categories for that galaxy (see figure 3), so that users could see how their answers were being used to reach an agreement. We also used textual information on the screen that would similarly convey the users' contribution to science. For instance, when users approach the table, they are invited to "help scientists classify galaxies", and after a user classifies a galaxy, one of five messages is displayed. The messages are "Nice work! You just science'd!", "Thanks for your help!", "You're a real scientist!", "Don't stop now!", and "Do this at home on uscientist.org."

The exhibit included messages to inform the user about other ways to engage with the project through the online and app versions of Galaxy Zoo. In addition, we included a button the user could tap to indicate their intent to continue to contribute to Galaxy Zoo in the future.

3.3 Data Collection

3.3.1 The Planetarium. We collected data in a midsize science museum serving around half a million visitors every year in a major American city. The exhibit is situated in a central gallery with capacity for 277 people. Surrounding exhibits are a mix of hands-on moderated experiences, interactive explainers on mounted tablets, and objects from the museum's collections.

Elsewhere in the planetarium, visitors can interact with touch tables and other digital exhibits on a variety of space-science-related topics, including designing their own constellation, discovering images of a telescope in Antarctica, or voting on whether Pluto should be considered a planet. One other exhibit in the museum is directly connected to a Zooniverse project: a collaboration with the museum's collections department where visitors to classify images of constellations.

3.3.2 Survey. During the month of August 2019, visitors to the planetarium were asked to take part in a survey about their views about science and citizen science. Surveys took place in the same gallery as the U!Scientist exhibit and included 132 people who had not (yet) experienced the exhibit (pre-survey), and 182 people who had experienced the exhibit as they were stepping away from the table (post-survey). We only surveyed visitors either before or after the touch table experience (but not both) in order to avoid taking up too much of guests' time.

Visitors rated their level of agreement or disagreement with 12 statements about science and citizen science on a five point scale, were asked to classify a galaxy, and provided demographic information about themselves (age, gender and race/ ethnicity). On

the pre-experience survey, they were also asked about their prior awareness of citizen science and the likelihood that they would try a Zooniverse or in-person citizen science project. On the post-experience survey, they rated their agreement with five additional statements about their ease of use and enjoyment of the exhibit, and were asked to describe what they thought the exhibit was about and whether astronomers would use the classifications they made at the exhibit, as well as suggestions for improvement of the exhibit.

Forty-nine percent (49%) of respondents identified as Female; 47% identified as Male and 3% identified as Non-binary or preferred not to report their gender. Visitors identified their race/ethnicity as 57% White, 11% Black, 14% Latinx, 13% Asian, and 5% mixed-race. This distribution is reflective of typical museum visitors' demographics in the U.S. [12].

3.3.3 Video Study. The exhibit was installed in a large gallery near the entrance on the main floor of the planetarium. A video recorder was mounted in front of the exhibit and an audio recorder with two microphones at opposing corners of the table was placed under the exhibit. Two of six total positions at the tabletop were blocked off to allow the camera to have a clear view of the remaining four positions, and the area around the exhibit was roped off to limit access only to consented visitors. Data were collected over a period of a month in 2019. A researcher approached visitors and informed them about the study. Visitors who agreed to participate in the study signed a consent form and were asked to step up to any of the remaining four stations. In addition to audio and video data, automatic logs of users' interaction with the exhibit were also collected.

After collection, video and audio files were merged using Adobe After Effects [8]. The video data collected consisted of ten video files. Two of the ten videos were not included in the analysis presented here because of technical difficulties with their accompanying audio files. The eight videos used for analysis resulted in a total of 254 minutes of video data. We collected data from 82 participants of whom 43 were coded by observers as male and 39 as female (acknowledging that these classifications were based on our best guess as visitors presented and may not be entirely accurate). We also coded one baby, 27 children, 11 teenagers and 43 adults. We counted 39 groups based on the percentage of visitors' overlapping times [5] with an average group size of 2.1 people.

3.3.4 Naturalistic Study. For the naturalistic study, all table positions were available to visitors and visitors were allowed to come and go freely from the touch table. Signage explained that visitors to the exhibit would be observed anonymously, and provided further information about the study and IRB approval for those who were interested. An observer stood near the touch table and used a custom Filemaker [20] database running on a tablet device to code arrival and departure times overall and when visitors switched table positions, as well as roughly coding for apparent gender and age. These data were correlated with automated log data from the touch table about behaviors.

Observations were collected over a two day period, Thursday and Friday, August 29-30, 2019, with sessions running from late morning to mid-afternoon both days, for a total of 2.9 hours on Thursday, and 2.5 hours on Friday. There were 63 people observed on Thursday (21.7 per hour), and 129 observed on Friday (52.3 per

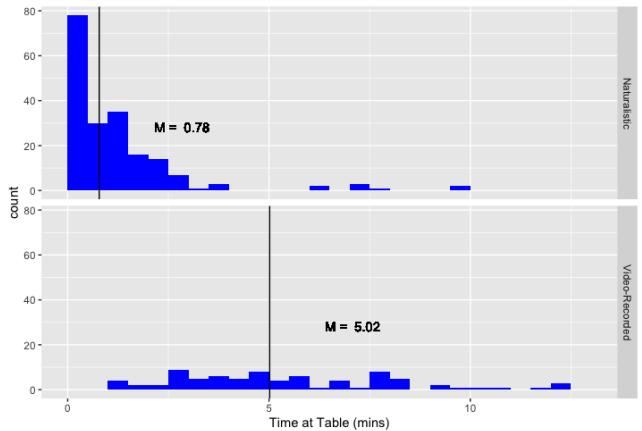


Figure 5: Dwell times (minutes) for naturalistic (top) and video data (bottom) with median dwell times marked. As in many museum studies, consented video-recorded participants spent more time at the exhibit than visitors observed in naturalistic settings.

hour) – Friday was substantially busier than Thursday. Gender was coded as female for 70 participants (49%) and as male for 72 participants (51%). Of the 143 participants whose age group was coded, 89 were adults (62%) and 54 were teens or children (38%). Fifty (50) participants were not coded for gender, and 49 were not coded for age because they stayed for a very brief time (un-coded participants had median dwell times of only 7 seconds, and all but one stayed less than 20 seconds).

3.4 Data Analysis

3.4.1 Survey. Because the survey was not built from an existing scale or scales, we conducted exploratory factor analysis to determine whether items could be clustered into a smaller number of coherent and meaningful factors. Items on the survey cluster into four factors with slight differences in weights on the pre- and post-survey:

- The first factor includes items “I’m good at understanding science topics”, “I’m good at explaining science to other people”, “I like thinking about science problems”, and “I could be a scientist one day if I wanted to”. These items seem to capture a sense of empowerment and identity related to science.
- The second factor includes items “Our world is nicer to live in because of science”, “I think science is important”, “I trust the process scientists use” and, particularly on the post-survey, “Museums are a place in which you can contribute in valued and meaningful ways to science”. These items seem to capture a valuing of the purposes and approaches of science.
- A third factor includes “I want to learn more about science” and “I enjoy science”. These items point to a positive attitude toward science.
- Finally, a fourth factor includes ratings for “How likely would you be to try a Zooniverse project?” and “How likely would

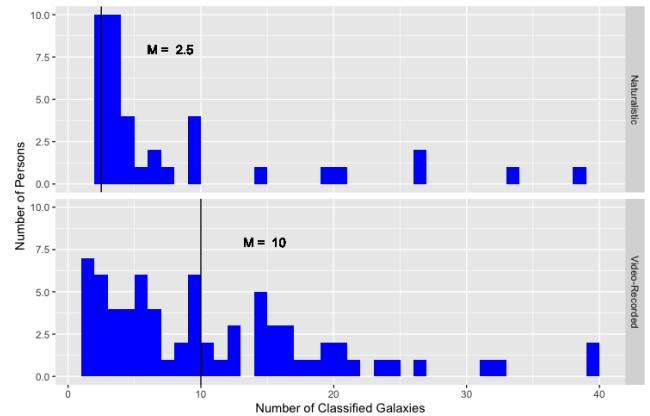


Figure 6: Number of galaxies classified by people in naturalistic and video conditions with the median number of classifications marked.

you be to try an in-person citizen science project?” These focus on the likelihood of participating in some kind of citizen science project.

All together, these four (4) factors explain 60.2% of the cumulative variation of all twelve (12) items and are sufficient to describe the data ($\chi^2 = 32.9, df = 24, p = 0.19$). Internal reliability of the factors (standardized Cronbach’s α) range from 0.72 to 0.81. Based on this, we calculated factor scores by averaging the ratings for items contributing to each factor, and then analyzing these as our outcome variables.

We went through a similar process for the five (5) post-experience feedback items and found two coherent factors:

- One factor loads on the items “I thought the exhibit was fun” and “If I came back to the planetarium, I would try the exhibit again”. These items both capture a sense of enjoyment.
- The other factor loads on items “I thought the exhibit was easy to use”, “I knew what to do right away” and “I felt confident using the exhibit”. These items all point to a sense of ease and confidence in interacting with the touch table.

Together these two (2) factors explain 69% of the total variation in these questions and are sufficient to describe the data ($\chi^2 = 0.84, df = 1, p = 0.36$).

We used descriptive statistics and OLS regression to look for pre-to post-experience differences in these factor scores, as well as the correctness of respondents’ galaxy classification.

3.4.2 Video Study. Two raters analyzed video and audio collected from participants’ visits to the U!Scientist tabletop. The tabletop positions were numbered from left to right, from one to six. Positions one, two, three and six were used for data collection. Participants themselves were also systematically given IDs, based on their arrival time, gender, and age. Each participant’s arrival and exit times were also recorded, with arrival time being determined as the first time they stood over their station and their exit times being the time they left their stations.

With participants and positions labeled, raters coded participant behaviors. Behavioral codes were built on previous observational work with multi-touch tables in museums, and included verbal,

physical, and emotional codes that were initially used in [5] and were modified to meet the specific content and design concerns of this project. The codes were refined by the researchers as they reviewed the videos, primarily to improve focus on target behaviors and inter-rater reliability (see table 1). During video review and early coder training the collaboration and social codes were collapsed to one code.

Raters coded participant behavior one position at a time, completing one position in its entirety before moving on to the next position and its set of participants. Behavioral codes were captured in intervals of twenty seconds, until all positions and participants were accounted for. We used twenty seconds as the unit of our analysis based on prior work intended to support real-time coding of visitor interaction with exhibits (see [5]). Each rater's behavioral codes were then compared using a Cohen's Kappa (κ) statistic for each position and behavioral code. Both raters trained on two videos with 36 minutes of participant interaction, coding for ten minutes of data at a time and computing Cohen's Kappas until we reached a κ of over 0.7 across all codes. This reliability was reached after 4 ten minutes segments were reviewed (there were portions of the ten minute segments when the table was unoccupied). For each behavioral feature, the Kappa values were then averaged across each position to determine its inter-rater reliability. We then combined codes for each position aligning the codes based on time and counted every occurrence of a code once per twenty seconds. We combined the codes to be consistent with our team's previous live data collection in naturalistic settings which was collected per group rather than per individual. We matched the logs collected during visitors' interaction with the exhibit to individual users from the video data using their arrival times.

We also analyzed the video data following a thematic analysis approach [6]. We generated codes based on questions raised in this work. We went through the video data identifying and extracting conversations relevant to these codes. We generated themes by reviewing, revising, collapsing and expanding our codes as well as developing memos for codes. We share results from our analysis in the results section.

3.4.3 Naturalistic Study. Observations of visitors' estimated demographic characteristics as well as their times of arrival, positions, and departures from the exhibit were recorded in a Filemaker database. These data were merged with log data from the table based on times and positions, with visitors' initial position at the table used to identify them, even if some time elapsed between arrival at the exhibit and their first logged interaction with the table. Some analyses focus solely on the observation data – e.g., dwell time and number of visitors at the same position. Group size was calculated as the time-weighted average of number of people at the table during each person's time at the table – a slightly different method than was used in the video data. Other analyses count table behaviors associated with a position, aggregating these for people based on the times they were recorded as standing at different positions. Analyses include summary statistics and OLS regression to address group differences and relationships among variables. Skewed data were log transformed as needed to normalize their shape before analyses – then transformed back to ease interpretation.

4 RESULTS

Here we assess the extent to which U!Scientist supports active prolonged engagement by examining visitors' engagement with the exhibit, their collaboration with one another, and the effect of the gamified elements on their behavior and engagement.

4.1 Visitors' Engagement

We measured visitors' engagement in several ways including the time spent at the exhibit, the number of completed classification tasks, visitors' talk that demonstrated an understanding of the exhibit tasks, and any indication of visitors' willingness to participate in people-powered research in the future.

4.1.1 Dwell Time. For video recorded participants, we calculated the dwell time by keeping track of the first time a visitor stood over a station at the tabletop to the time the visitor left the table. The average time spent by each participant for the video study was 6.2 minutes with a median of 5 minutes and standard deviation of 4.8 minutes (see figure 5).

However, these dwell times during consented, videotaped observations may not reflect dwell times under more naturalistic conditions [5] (see figure 5). To examine this, we also collected dwell times over a period of two days for 192 visitor-users of U!Scientist when access to the exhibit was *not* limited by the need to get consent for videotaping.

During these more naturalistic times, median visitor dwell time was 47 seconds (mean of log-transformed distribution, 37 seconds) though this varied by date, with shorter dwell times (28 seconds) when the table was busier (Friday), and longer dwell times (62 seconds) when it was less busy (Thursday) ($t=4.1$, $p<.0001$).

Prior research also shows that group size has an impact on dwell times at multi-touch table exhibits [5]. We found dwell times for medium sized groups (2.2 people) averaged 46 seconds, whereas larger groups spent slightly less time (44 seconds for a group twice the size), and smaller groups spent much less time (22 seconds for a group half the size) at the exhibit ($F=11.9$, $df=(3,188)$, $p<.0001$).

There were 78 observed gap times (periods when the table was empty) during the two days of naturalistic observation, ranging from as little as 2 seconds to as much as 22.2 minutes, with median gap time of 54 seconds (IQR from 26 to 151 seconds). The table was empty for 1.6 of the 2.6 hours on Thursday (63% of the observed time) and 1.4 of the 2.3 hours on Friday (58% of the time). Overall, the table was occupied 39% of the time.

4.1.2 Galaxy Classification. Log data showed that video recorded participants classified a total of 901 galaxies, with an average of 12 galaxies per person, a median of 10 galaxies, and a standard deviation of 14 galaxies 6.

Visitors in naturalistic observations classified fewer galaxies, with a total of 398 galaxies classified for 192 visitors. While 78 (41%) classified at least one galaxy, median number of galaxies classified for this group was 2.5 (mean = 5.1, SD = 7.4) with a maximum of 39 galaxies classified. Adults were less likely to classify at least one galaxy (36% of them) than teens or children (52% of them; $\chi^2 = 3.31$, $df=1$, $p=0.07$). Visitors who classified at least one galaxy stayed significantly longer (median 96 seconds, mean 138 seconds)

Code	Comment
Verbal Codes	
Make claim	Comments with elements of a scientific claim.
Provide evidence	Comments that back up a claim with evidence or an observation.
Astro talk	Any talk about galaxies, stars, space etc.
Collaboration/Social	Verbally inviting or engaging others in the activity.
How to/Technical	Comments on how to interact with the table user interface.
Read aloud	Reading the table content verbatim.
Refer to other activity	Talking about another exhibit or experience in the past.
More citizen science	Reference to engaging in citizen science in the future.
Physical Codes	
Point/Indicate	Any physical gesture to the table element for someone else.
Prevent/Control	Physically controlling or preventing others' access to the table.
Emotion Codes	
Enjoyment	Expression of happiness accompanied with a verbal confirmation.
Achievement	Comments including any reference to gamified elements.
Dislike/Boredom	Explicitly expressing dislike or boredom at the exhibit.
Confusion	Expressing a lack of knowledge of how to engage with the exhibit.

Table 1: Description of coding scheme used for video analysis.

than the 114 (69%) who didn't classify any galaxies (median 21 seconds, mean 31 seconds; $t=7.5$, $df=84$, $p<0.0001$).

Estimating how long it takes for visitors to classify a galaxy is tricky because not all their time is spent classifying – they need to get oriented, and there are other ways for them to interact with the table and one another. Overall, average time spent between galaxy classifications for those who classified at least one was 27 seconds. However, log files indicate that the actual process of making a classification decision once a galaxy example has been placed in a visitor's classifier is much faster – median time for naturalistic observations is 4.7 seconds (mean = 7.2, SD = 7.5; IQR from 2.9 to 8.0 seconds; range 1 second to 1 minute).

4.1.3 Real Galaxies, Real Work and Real Scientists. One of our design goals was to convey to users that their interactions with the exhibit were contributing to work done by astronomers. Visitors' talk revealed that they perceived their interaction with the exhibit as "real work" and recognized that they were classifying "real" galaxies. This was expressed in a number of ways by participants while using the exhibit (see excerpts 1 and 2 below). This attribution of authenticity to their interaction with the exhibit was prompted mostly by text on the screen inviting visitors to help scientists and text displayed after the visitor classified a galaxy that referred to the visitor as a "real scientist." Expressions of enjoyment and achievement sometimes accompanied reading the "You're a real scientist!" text out loud (as in excerpt 3). Some visitors also attributed authenticity to their activities after they moved the map to display a new set of galaxies to classify.

excerpt 1:

Boy: Whoa! Wait! is this going to astronomers?!
Man: Yeah, we are doing real science ... Picking things out and classifying them

excerpt 2:

Man: These are actual pictures of galaxies and stars...this is like actual work.

excerpt 3:

Woman: It says I'm a real scientist.
Man: Yeah baby!

4.1.4 Contributing vs Being Correct. We assumed that allowing users to follow the process of building consensus was one way to convey to them the importance of their contribution to research. Understanding that the exhibit was collecting answers from visitors to further research rather than testing visitors' knowledge of galaxies seemed to be central to fostering a feeling of participation and co-ownership that museums strive to give visitors. Thus, we were interested in understanding whether visitors perceived the majority classification on the summary screen (see figure 3) as the correct classification, or whether they understood it as the results of a survey. Many visitors recognized that they were classifying real galaxies, but they interpreted the charts showing the distribution of classifications in several ways. Some visitors considered alignment with the majority classification a positive achievement and non-alignment, a negative achievement (see excerpt 4). Many visitors acknowledged that the majority classification was not necessarily the *right* classification, but they often looked to the majority classification to confirm their own classification, or to learn more

about how to classify galaxies, and expressed joy when their classification coincided with the majority classification. They also took advantage of the motivation afforded by the majority classification to keep themselves engaged.

excerpt 4:

Woman 1: *Yaay! I got that right!*
 Woman 2: *You ain't got nothing right!*
 Woman 1: *I did!*
 Woman 2: *Did they say it's right? (Looking over into woman 1's station).*
 Woman 1: *'cause I choose it, that means you got it right. (Pointing to the screen)*
 Woman 2: *But read it, 'cause mine said that too. They're sending the data over.*

4.1.5 Beyond the Target Users. Although U!Scientist was designed with people ages 12 and up in mind, we found that younger visitors found value in their interaction with the table. Younger children could also participate in galaxy classification with a little help. While some younger children came to understand and use U!Scientist by watching their older siblings and exploring the exhibit themselves, others explicitly asked parents or older siblings how to use the exhibit. Parents provided scaffolds for young children by reading instructions aloud, using simple language (for example, *"That one definitely has features do you see how it's all wriggly?"*), using exciting melodies as they dragged and dropped galaxies, modeling how to use the exhibit and interpreting the charts. When parents or adults did not join younger children in using the exhibit, the children had more difficulty interacting with the exhibit. Confusion codes came mostly from this group. As with most digital devices, children approached the tabletop with the expectation that they would figure out how to use the exhibit as they used it. They used the written instructions minimally and referred to them only when they got stuck or were unable to level up.

4.1.6 More Citizen Science. Although many visitors expressed that they wanted to know more about the galaxies they were classifying, and the reasons why they took certain shapes and forms, very few (just 3) used the check box on the touch table to indicate their intent to use Galaxy Zoo in the future. Verbal codes to capture speech connecting U!Scientist to other exhibits or to future exhibits were similarly scarce. We were cautious not to overload the exhibit with too much information that would be difficult to organize, though we wanted to provide sufficient information to engage the user without taking up too much user time. We considered visitors' desire for more information as a success and consistent with the goals of our design. We wanted to rouse users' interest in the Galaxy Zoo project so that they would engage with it in the future and outside the museum setting. In the excerpt below, a visitor whose talk reveals she is a teacher takes a photo of information on the exhibit that tells how to engage in galaxy classification in the future and declares that it would be a good exercise for her classroom. These are the types of connections we hoped to make by installing this exhibit in a museum space.

excerpt 5:

Woman 1: *I'd like to do this in my classroom. I don't know if this (inaudible)*

Woman 2: You can do it. It's an app.

Woman 1: I know. (Pulls out a phone from her purse and takes a picture of something on the table)

4.1.7 Survey Results. Visitors who responded to the pre-experience survey were able to successfully classify a galaxy 73.8% of the time; those who responded after engaging with U!Scientist were successful 81.9% of the time. This represents increased success of 8%, but the difference is not statistically significant ($t=1.61$, $p=0.11$).

Survey scales ranged from -2 (strongly disagree/ highly unlikely) to +2 (strongly agree/ highly likely) with 0 representing neutral values. Overall, participants agreed somewhat (mean = 0.84, SD=0.73) that the exhibit was easy to use with only 8% giving this negative ratings. They agreed more strongly (mean = 1.27, SD=0.63) that the exhibit was fun to use, with only 2% giving negative ratings.

Pre-experience visitors agreed somewhat with the Factor 1 statements about science empowerment and identity (mean=0.52, SD=.77), and with the Factor 4 statements about their likelihood of participating in citizen science projects (mean=0.50, SD=.93). They agreed more strongly with the Factor 2 statements about valuing the purposes and approaches of science (mean=1.23, SD=.49), and with the Factor 3 statements reflecting positive attitudes towards science (mean=1.17, SD=.57).

Post-experience visitors agreed more strongly than pre-experience visitors on all four factors. These differences were statistically significant for Factor 2 about valuing science ($\beta = 0.15$, $t=2.54$, $p=0.012$), and marginally significant for Factor 1 about science identity ($\beta = 0.16$, $t=1.92$, $p=0.056$) and Factor 3 about positive attitudes towards science ($\beta = 0.12$, $t=1.66$, $p=0.099$). Post-experience visitors agreed more strongly ($\beta = 0.14$) that they were likely to participate in citizen science activities (Factor 4) but these differences were not statistically significant ($t=1.31$, $p=0.19$).

4.2 Collaboration

People-powered research uses the power of the crowd to carry out tasks that would otherwise be impossible to do by a small number of people. Thus, it lends itself easily to collaboration. We tried to preserve and amplify this collaboration value (between scientists and contributors and among contributors) when implementing the touch table exhibit. The invitation to visitors to use the exhibit was to "Learn about galaxies while contributing to research" and not just simply to explore an exhibit or to classify galaxies. In addition, after classifying a galaxy, visitors got to see how other people had classified the same galaxy on the summary screen (see figure 3).

Requesting help with the tool provided in U!Scientist did not capture the amount of collaboration that transpired while using the exhibit. As other studies have shown, investigating talk in open-ended informal learning environments helps to identify moments of productive learning [21, 33]. As such, our analysis of visitors' talk while using the exhibit gave us more insight into their collaborative behaviors. The interaction logs for video recorded participants showed that 32 people out of 82 in total made 121 requests for help and only 16 of those requests were accepted. However, our verbal codes indicated a significant amount of collaborative interaction between visitors. Our video data analysis resulted in 1475 codes of which 1056 were verbal codes, 265 were emotion codes and 154 were physical codes. "Astro talk" code was 35% percent of

verbal codes. This code helped us to understand whether visitors were able to use the language supplied through our design for expressing their thoughts and collaborating on galaxy classification. The "Make claim" codes made up 26% of verbal codes and visitors' talk in general. These codes were also an indication of collaboration as they were often framed as questions and directed towards members of the group. Making claim codes were only surpassed by "Astro talk" code. We excluded the "collaboration/social code" in computing these percentages because most verbal codes were directed to someone else and coded as "collaboration/social". The number of claims made was not matched by the evidence provided. "Evidence" codes were only 6% of all verbal codes. However, we saw more evidence-based conversations between people who shared the same station. We found that 39% of evidence given for claims came from shared stations. Only 20% of visitors shared stations for up to 20 seconds and above. Evidence was provided to confirm classifications, to resolve disagreements and, to model to younger visitors how to classify galaxies. Giving evidence was an indication of understanding and engaging more deeply with the task of classifying galaxies (see excerpt 6).

excerpt 6:

Daughter: Smooth is 'gradually fades from the center', it doesn't fade...
Father: It doesn't, but it doesn't have any features. (He points into her spot) Features see would have like arms, those thingy thingy.

We often saw parents and children begin their interaction with the exhibit in a single station and eventually separate into individual stations. We also saw parents temporarily leave their stations to help their children to understand the exhibit. This seemed to indicate that visitors could learn to use the exhibit and teach others to use it in a short period of time and that stations were not restrictive. Visitors that took up different stations but were within a group often reached out verbally and physically through pointing to people within their group to collaborate on classifying galaxies. The dynamics of interaction within groups that shared a station included a lot of verbal interaction and pointing as in the excerpt below.

excerpt 7:

Mother: So, this is a real picture of the sky ... and each of these things are pictures that the telescope is taking pictures of. So, we bring it over here to make it go a bit bigger. And then you say, "What does this look like? is it smooth? does it go gradually from the center? does it have features – can you see the things that jump out? or is it not a galaxy?" So, what do you think (Pointing at the table all the time she is explaining) ... ?

Boy: It's not a galaxy! (pointing at the table)

Mother: Um, why do you think that? It looks kind of like one of these to me (pointing to the tabletop). I think it is a galaxy.

Boy: I think it is.

Mother: You think it is now too?

"Point and indicate" codes were 9% of all codes and 91% of all physical codes. U!Scientist implements parallel stations which precluded much interference among visitors. As a result, there were

few occurrences of verbal and physical restraint for U!Scientist, compared to other codes. Physical restraint occurred when parents tried to guide children or direct their attention to other elements on the exhibit or when a sibling tried to restrain another sibling. However, there was a lot of pointing to draw attention to elements on the exhibit. Other research has characterized pointing in museums as a positive physical expression that supports collaboration [19].

We looked for evidence of collaboration in the naturalistic observations in two ways. First, visitor position observations show more than one person at a position for 31.5 of the 197.1 minutes (16%) of the time that positions were recorded. The vast majority of these were 2 people at a position, but we recorded 3 people at a single position for 34 seconds (0.3% of the time). These rates of sharing a position are comparable to, or slightly higher than what was observed in the videotaped observations (7%).

Next, the table allowed visitors to ask others for help in classifying a galaxy – a different kind of evidence of collaboration. Successful requests require having a galaxy placed in the visitor's classifier, and choosing another user at the table who isn't already busy helping others. Over the course of the two days of naturalistic observation, 34 visitors sent 75 requests for help, though only 21 requests from 18 visitors were proper requests (28% of requests from 53% of those who tried). Of these 21, only four were accepted and one was declined (24% responded to), and only one of the four accepted requests was completed. Those asking for help were a mix of those who did *not* classify a galaxy themselves (11 of the 34; 10% of the 114 making no classifications) and those who *did* classify one or more galaxies (23; 34% of those making galaxy classifications). However those who classified a galaxy made successful requests twice as often as those who hadn't classified galaxies (60% of the time vs. 27% of the time). Overall, 18% of visitors tried to use this Ask for Help tool, but only 28% of them succeeded in making a request, and only a quarter of successful requests received responses.

4.3 Gamified Elements

In order to measure the effect of gamified elements in the exhibit, we compared visitors' talk that included an "achievement" code ($n = 108$) to talk that was exclusively about classifying galaxies ("make claim" and "evidence", $n = 237$). The exhibit's gamified elements did not seem to overwhelm the exploratory and learning task of classifying galaxies. Designers of museum exhibits are wary of digital interactions that distract from learning or exploration of the central concept in the design and also want to align with general expectations of families when they visit museums. Thus, we considered more galaxy talk than game talk a success.

However, groups that noticed the gamified element seemed motivated to continue to classify galaxies until they levelled up. Excerpt 8 includes some interaction within a group when they discover the levels feature and subsequent interactions that suggest that a competitive element has been introduced. This group spent more time at the exhibit than any other group in the video study. In addition, of 54 visitors who classified five or more galaxies in the video study, 33% of them ended their interaction with the exhibit at the completion of a level. In the naturalistic study, we also found some evidence that "levelling up" was motivating to the 19 visitors who

reached the first "level" by classifying at least 5 galaxies – visitors stopped more often after classifying multiples of 5 galaxies than at any other time ($\chi^2 = 13.9, df = 4, p = .008$).

Visitors found other ways to gamify their interaction with the exhibit beyond the explicit gamified features. For instance, some parents cheered their children when they classified a galaxy consistent with the majority classification and some visitors started to look for particular types of galaxies.

excerpt 8:

Boy 1: Why are you guys not on level 5? Maybe I'm getting it right?

...

Boy 1: I'm almost at level 6! Yes! What level are you on?

Boy 2: Yes! 3 more and I get to level up!

5 DISCUSSION

In this section we discuss our strategies that made U!Scientist suitable for the museum space and to engage new audiences. We also discuss how the museum space can support different ways of collaborating on people-powered research.

5.1 Making People-Powered Research Suitable for Museum Spaces

We wanted our design to be engaging as well as to communicate to museum visitors that they were contributing to real science. However, museum exhibits often have short dwell times as there are many exhibits vying for the attention of visitors. Our preliminary studies led us to simplify the classification task, include messaging that communicated to users that they were engaging in real science, and make the contributions of other visitors visible.

The exhibit was able to hold people's attention. Median dwell time in the naturalistic study was 47 seconds, but that doubled to 96 seconds for the 41% who classified at least one galaxy. Longer dwell times were observed for video recorded consented participants as compared to naturalistic observations and this is consistent with findings in other HCI studies (see [5, 16]). Others have hypothesized that the process of learning about the exhibit and giving consent to enter a cordoned off space for videotaping increases visitors' sense of commitment compared with those who "wander by" and decide in the moment whether to stay longer. The simplified task was quick and easy to do, taking just 4.7 seconds (median) once a galaxy example was chosen. There is some evidence that modest gamification elements seem to have motivated the small group of visitors who experienced "leveling up", without overly distracting others.

Analysis of survey data also indicate that the exhibit was easy to use and quite fun, and suggests that visitors' value for science, science identities, and attitudes towards science were positively impacted. We also found that visitors recognized that they were contributing to real science. Communicating to visitors the value of their contribution was important for getting them to engage with the exhibit. We tried to help visitors understand that they were engaging in authentic research work and not just exploring a science exhibit or playing a game. This information was important because it had the potential to foster the feeling of co-ownership and co-creation of knowledge, motivate volunteers to continue to

contribute to science, and build their science identities. In order to communicate the authenticity of the task and the relevance of each volunteer's contribution, we made the process of reaching consensus transparent to volunteers. These findings indicate that promoting understanding and communicating the values of citizen science to volunteers is important for designing people-powered research in museums.

5.2 Engaging New Audiences

Although Galaxy Zoo has a rich base of contributors, engaging new audiences in citizen science projects is hard [35]. This work explores the museum as a potential site for engaging new audiences. Our video data shows that connections can be made to enthusiasts and amateur scientists in the museum space. In addition, our findings suggest that younger audiences who are typically not targeted by citizen science projects are able to contribute to citizen science with some simple design changes, for example, communicating project instructions in simpler and exciting language. Although the instructional language in this project was not particularly designed for children below the age of 12, some were able to engage intelligently and fruitfully with the exhibit with a little help from older siblings or adults accompanying them.

Children who were unaccompanied by their parents had a harder time engaging with U!Scientist. They had difficulty understanding the classification task. They also had some difficulty interpreting the summary of classification screens. We captured this behaviour in our video data analysis as 'confusion'. We created this code when we observed that some young children were expressing their frustration through non-specific verbal questions about how to engage with the exhibit. We also imagine that a request for help on U!Scientist prior to classifying a galaxy was an expression of this confusion and possibly our design did not clearly communicate to users the difference between instructional help and help to classify galaxies. In the future, we may be able to support younger audiences to use U!Scientist more independently – just like parents observed in this study – by exploring simplified and more playful instructional material.

These findings extend our understanding about age groups of people who can contribute to citizen science projects and introduces potential younger audience-contributors to citizen science.

5.3 Supporting New Ways of Collaborating on People-Powered Research in Museums

Online people-powered research is collaborative by definition, but this collaboration is often distributed, asynchronous, and anonymous. Some citizen science tasks have benefited from co-located groups or groups in which members are known to each other to produce even better solutions than when volunteers work individually [7]. Our experience with U!Scientist shows that there are new opportunities for side-by-side, real-time collaboration in people-powered research in museum settings. We implemented U!Scientist as a multi-touch table device with six stations running independently of each other. Individuals within a group could work on classifying galaxies independently or collectively. Evidence of collaboration while using U!Scientist was manifold: Visitors worked

side-by-side at the same station 16% of the time in naturalistic observations (7% in video observations); 18% of them used the Ask for Help tool (though only a small fraction were successful); and there was substantial direct video evidence of collaboration and communication among visitors. In addition, shared U!Scientist stations supported evidence-based interaction between users.

Our findings suggest that U!Scientist tasks have suitable complexity for volunteers to work independently and in parallel or in collaboration with others. The museum setting and multi-touch table supported the task of classifying galaxies as a collaborative one, and this collaboration allowed users – even those younger than our target group – to enjoy and find value in classifying galaxies.

Co-located volunteers and/or scientists might be worth exploring for Galaxy Zoo and other Zooniverse projects in the future.

ACKNOWLEDGMENTS

This research was supported by grant #1713425 from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. The Galaxy Zoo project that U!Scientist is based on is part of the Zooniverse.org platform, development of which is funded by generous support from the NSF, NASA, a Google Global Impact Award, the Helmsley Foundation, and the Alfred P. Sloan Foundation, among others.

We would like to thank all of the museum visitors who participated in this study and all the online volunteers contributing to research through Zooniverse. We would also like to thank Adler Teen Interns David Balderas and Claire Finley for their invaluable assistance testing early prototypes of the touch table interface with museum guests. Shoshi Shapiro supervised and synthesized their findings, providing insight and recommendations throughout the design process. We gratefully acknowledge Will Granger's efforts as the lead Zooniverse software engineer for the U!Scientist experience.

REFERENCES

- [1] Sue Allen and Joshua Gutwill. 2004. Designing With Multiple Interactives: Five Common Pitfalls. *Curator: The Museum Journal* 47, 2 (2004), 199–212. <https://doi.org/10.1111/j.2151-6952.2004.tb00117.x>
- [2] Alissa N. Antle, Allen Bevans, Josh Tanenbaum, Katie Seaborn, and Sijie Wang. 2010. Futura: Design for Collaborative Learning and Game Play on a Multi-Touch Digital Tabletop. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (Funchal, Portugal) (TEI '11). Association for Computing Machinery, New York, NY, USA, 93–100. <https://doi.org/10.1145/1935701.1935721>
- [3] Elham Beheshti, David Kim, Gabrielle Ecanow, and Michael S. Horn. 2017. Looking Inside the Wires: Understanding Museum Visitor Learning with an Augmented Circuit Exhibit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 1583–1594. <https://doi.org/10.1145/3025453.3025479>
- [4] Elham Beheshti, Mmachi Obiorah, and Michael S. Horn. 2015. "Let's Dive into It!": Learning Electricity with Multiple Representations. In *Proceedings of the 14th International Conference on Interaction Design and Children* (Boston, Massachusetts) (IDC '15). Association for Computing Machinery, New York, NY, USA, 263–266. <https://doi.org/10.1145/2771839.2771892>
- [5] Florian Block, James Hammerman, Michael Horn, Amy Spiegel, Jonathan Christiansen, Brenda Phillips, Judy Diamond, E. Margaret Evans, and Chia Shen. 2015. Fluid Grouping: Quantifying Group Engagement around Interactive Tabletop Exhibits in the Wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 867–876. <https://doi.org/10.1145/2702123.2702231>
- [6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [7] Francesco Cappa, Jeffrey Laut, Oded Nov, Luca Giustiniano, and Maurizio Porfiri. 2016. Activating social strategies: Face-to-face interaction in technology-mediated citizen science. *Journal of Environmental Management* 182 (2016), 374 – 384. <https://doi.org/10.1016/j.jenvman.2016.07.092>
- [8] Mark Christiansen. 2013. *Adobe After Effects CC Visual Effects and Compositing Studio Techniques*. Adobe Press, San Francisco, CA.
- [9] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756–760. <https://doi.org/10.1038/nature09304> Number: 7307 Publisher: Nature Publishing Group.
- [10] Jeremy Dietmeier, Ben J. Miller, Ben Devane, Kristen Missall, and Salloni Nanda. 2017. Shredding with Mom and Dad: Intergenerational Physics Gaming in a Children's Museum. In *Proceedings of the 12th International Conference on the Foundations of Digital Games* (Hyannis, Massachusetts) (FDG '17). Association for Computing Machinery, New York, NY, USA, Article 58, 4 pages. <https://doi.org/10.1145/3102071.3106365>
- [11] Alexandra Eveleigh, Charlene Jennett, Stuart Lynn, and Anna L. Cox. 2013. "I Want to Be a Captain! I Want to Be a Captain!" Gamification in the Old Weather Citizen Science Project. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Toronto, Ontario, Canada) (Gamification '13). Association for Computing Machinery, New York, NY, USA, 79–82. <https://doi.org/10.1145/2583008.2583019>
- [12] Betty Farrell and Maria Medvedeva. 2010. Demographic Transformation and the Future of Museums. <https://doi.org/10.6082/uchicago.1272>
- [13] Chiara Franzoni and Henry Sauermann. 2014. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy* 43, 1 (2014), 1–20. <https://doi.org/10.1016/j.respol.2013.07.005>
- [14] Trisha Gura. 2013. Citizen science: Amateur experts. *Nature* 496, 7444 (2013), 259–261. <https://doi.org/10.1038/nj7444-259a> Number: 7444 Publisher: Nature Publishing Group.
- [15] Mahboobeh Harandi, Corey Brian Jackson, Carsten Osterlund, and Kevin Crowston. 2018. Talking the Talk in Citizen Science. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Jersey City, NJ, USA) (CSCW '18). Association for Computing Machinery, New York, NY, USA, 309–312. <https://doi.org/10.1145/3272973.3274084>
- [16] Michael Horn, Zeina Atrash Leong, Florian Block, Judy Diamond, E. Margaret Evans, Brenda Phillips, and Chia Shen. 2012. Of BATs and APEs: An Interactive Tabletop Game for Natural History Museums. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). Association for Computing Machinery, New York, NY, USA, 2059–2068. <https://doi.org/10.1145/2207676.2208355>
- [17] Michael S. Horn, Brenda C. Phillips, Evelyn Margaret Evans, Florian Block, Judy Diamond, and Chia Shen. 2016. Visualizing biological data in museums: Visitor learning with an interactive tree of life exhibit. *Journal of Research in Science Teaching* 53, 6 (2016), 895–918. <https://doi.org/10.1002/tea.21318>
- [18] Eva Hornecker. 2008. "I don't understand it either, but it is cool" - visitor interactions with a multi-touch table in a museum. In *2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems* (2008–10). IEEE, New York, 113–120. <https://doi.org/10.1109/TABLETOP.2008.4660193>
- [19] Thomas. Humphrey and Joshua P. Gutwill. 2017. *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits*. Taylor & Francis, Philadelphia, PA. <https://books.google.com/books?id=40ZnDwAAQBAJ>
- [20] Filemaker Inc. 2019. FileMaker Pro 18 Advanced (Version 18.0.3.317): Filemaker, Inc.
- [21] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences* 4, 1 (1995), 39–103. https://doi.org/10.1207/s15327809jls0401_2
- [22] Elizabeth Kunz Kollmann, Christine Reich, Larry Bell, and Juli Goss. 2013. Tackling Tough Topics: Using Socio-Scientific Issues to Help Museum Visitors Participate in Democratic Dialogue and Increase Their Understandings of Current Science and Technology. *Journal of Museum Education* 38, 2 (2013), 174–186. <https://doi.org/10.1080/10598650.2013.11510768>
- [23] Margaret Kosmala, Andrea Wiggins, Alexandra Swanson, and Brooke Simmons. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14, 10 (2016), 551–560. <https://doi.org/10.1002/fee.1436>
- [24] Chris J. Lintott, Kevin Schawinski, Anze Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389, 3 (2008), 1179–1189. <https://doi.org/10.1111/j.1365-2966.2008.13689.x> Publisher: Oxford Academic.
- [25] Marti Louye and Kevin Crowley. 2013. New Ways of Looking and Learning in Natural History Museums: The Use of Gigapixel Imaging to Bring Science and Publics Together. *Curator: The Museum Journal* 56, 1 (2013), 87–104. <https://doi.org/10.1111/cura.12009>

[26] Paul Marshall, Richard Morris, Yvonne Rogers, Stefan Kreitmayer, and Matt Davies. 2011. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada, 2011-05-07) (CHI '11). Association for Computing Machinery, New York, NY, USA, 3033–3042. <https://doi.org/10.1145/1978942.1979392>

[27] Ellen McCallie, Larry Bell, Tiffany Lohwater, John H. Falk, Jane L. Lehr, Bruce V. Lewenstein, Cynthia Needham, and Ben Wiehe. 2009. *Many Experts, Many Audiences: Public Engagement with Science and Informal Science Education. A CAISE Inquiry Group Report. Executive Summary*. Center for Advancement of Informal Science Education, Washington, DC. Publication Title: Center for Advancement of Informal Science Education.

[28] John Prpic and Prashant Shukla. 2016. Crowd Science: Measurements, Models, and Methods. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, New York, 4365–4374. <https://doi.org/10.1109/HICSS.2016.542>

[29] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. 2013. Galaxy Zoo: Motivations of Citizen Scientists. arXiv:1303.6886 [physics.ed-ph]

[30] Jessica Roberts, Amartya Banerjee, Annette Hong, Steven McGee, Michael Horn, and Matt Matcuk. 2018. Digital Exhibit Labels in Museums: Promoting Visitor Engagement with Cultural Artifacts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, Article 623, 12 pages. <https://doi.org/10.1145/3173574.3174197>

[31] Frauke Rohden, Christopher Kullenberg, Niclas Hagen, and Dick Kasperowski. 2019. Tagging, Pinging and Linking – User Roles in Virtual Citizen Science Forums. *Citizen Science: Theory and Practice* 4, 1 (2019), 19. <https://doi.org/10.5334/cstp.181> Number: 1 Publisher: Ubiquity Press.

[32] Avi Segal, Ya'akov (Kobi) Gal, Robert J. Simpson, Victoria Victoria Homsy, Mark Hartwood, Kevin R. Page, and Marina Jirotna. 2015. Improving Productivity in Citizen Science through Controlled Intervention. In *Proceedings of the 24th International Conference on World Wide Web (Florence, Italy) (WWW '15 Companion)*. Association for Computing Machinery, New York, NY, USA, 331–337. <https://doi.org/10.1145/2740908.2743051>

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

[34] Katie Todd, Gretchen Haupt, Elizabeth Kunz Kollmann, and Sarah Pfeifle. 2018. Fostering Conversation about Synthetic Biology Between Publics and Scientists: A Comparison of Approaches and Outcomes. *Journal of Microbiology & Biology Education* 19, 1 (2018), 19.1.32. <https://doi.org/10.1128/jmbe.v19i1.1434>

[35] John Zimmerman, Anthony Tomasic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafaee Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field Trial of Tiramisu: Crowd-Sourcing Bus Arrival Times to Spur Co-Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 1677–1686. <https://doi.org/10.1145/1978942.1979187>

[36] Zooniverse. 2020. Zooniverse. <https://www.zooniverse.org/> Library Catalog: www.zooniverse.org.