A Survey of Citizen Science Gaming Experiences

JOSH AARON MILLER ⓘD

KUTUB GANDHI ⓘD

ANNA GANDER ⓘD

SETH COOPER ⓘD

*Author affiliations can be found in the back matter of this article

⟩u⟨ ubiquity press

ABSTRACT

Citizen science games (CSGs) are a valuable means for motivating citizen science participation. However, many CSGs still suffer from the recruitment and retention issues of traditional citizen science projects, despite much prior literature on what motivates CSG players. In this study, we take a Human-Computer Interaction (HCI) perspective to explore in what ways CSGs are still failing to provide motivating play experiences for players. Using Qualitative Content Analysis, we conducted and analyzed an online survey of 185 players from 9 citizen science games. This survey contributes insights to the current state of CSG experiences and next steps for developers to address these issues. We found that major concerns included scientific communication, instructional design, user interface and controls, task quality, and software issues.

CORRESPONDING AUTHOR:

Josh Aaron Miller

Northeastern University, US

miller.josh@northeastern.edu

INTRODUCTION

Citizen science games (CSGs) are gamified applications, which enable the public to contribute to scientific research by collecting and/or processing scientific data (Cooper 2011; Newman et al. 2012; Wiggins and Crowston 2015) and/or learning and applying a domain skill complementary to the scientists' abilities (Keep 2018). CSGs are an effective means of co-creating knowledge (Schrier 2017), and a valuable way to "provid[e] the public with access to important and challenging problems facing science and society" (Tuite 2014). Being gameful, CSGs draw on the motivational power of games to engage a wider audience (Ponti et al. 2018).

However, simply being gameful is not enough to attract and retain citizen scientists (Miller and Cooper 2022). CSGs suffer from the same widespread issues of retention as traditional citizen science projects (Eveleigh et al. 2014; Iacovides et al. 2013; Jennett et al. 2016). Yet the same body of work unpacks a multitude of motivations for CSG players. If we know what attracts and retains CSG audiences, why are CSGs still struggling to maintain an audience?

This leads to our current research question: What are players experiencing in citizen science games, and how do their experiences differ from what the literature understands to be a motivating experience? On the basis of similar recent work by Miller and Cooper (2022), we hypothesize that the player experience includes significant frustrations that could be addressed by developers.

To investigate this hypothesis, we review the state of CSG player experiences through the lens of Human-Computer Interaction (HCI), a field that focuses on understanding the interactions between users and technology. HCI provides the methods to understand user experiences so developers can act to address present weaknesses. Moreover, project owners and stakeholders can take our findings to encourage developers to address weaknesses and assess success in their own projects. This will, in turn, lead to higher throughput of citizen science gaming and, overall, more effective citizen science games.

Thus, we surveyed the citizen science gaming community about their play experiences. This online survey produced 185 valid responses (after filtering) from 9 different citizen science games, though we note a particular skew toward Foldit due to its popularity and increased advertising for this survey by the Foldit developers. Using qualitative content analysis (QCA), we coded survey responses for commonalities (Guest and MacQueen 2008; MacQueen et al. 1998).

Our main contribution is a series of insights into how CSG players currently perceive the gameplay and surrounding experience of CSGs, as well as associated recommendations for CSG developers to address problems. Among other points, we found that: (1) players are seeking more frequent and clearer scientific communication regarding updates on the projects; (2) players are confused about how to play and need better instructions; (3) user interfaces and controls are often unintuitive; (4) data-focused CSGs suffer from poor task quality, causing player frustration; and (5) CSG software suffers from frequent bugs and crashes that should be addressed.

BACKGROUND

CSGs fall within the subset of serious gaming — gaming for purposes beyond entertainment. Research has been increasingly interested in the playability and player experiences of these games (Rienzo and Cubillos 2020). Although player experience was previously considered as simply player satisfaction, the player experience of serious games has now been understood, generally, as a combination of many factors including immersion, challenge, and emotion, among others (d'Ornellas et al. 2015; Rienzo and Cubillos 2020; Wiemeyer et al. 2016). Player experience has also been explored as an individualized phenomenon; for example, Tasnim and Eishita (2021) measured how a player's Big Five personality traits impact their player experience. Although work has been done to describe the "how-to" of designing for usability, playability, and learnability in serious games (Olsen et al. 2011), this work has not been extended to the design of CSGs specifically. Nor has the player experience of CSGs been examined in great detail, thus motivating the present study.

CSGs have been used in a variety of projects to increase participation and to collect more data in citizen science. For example, Project Discovery (Sullivan et al. 2018) added a citizen science image classification task as a mini-game to the popular Massively Multiplayer Online Roleplaying Game (MMORPG) EVE Online in order to classify fluorescence microscopy images. Similarly, the Borderlands Science project places a citizen science task of aligning RNA gene sequences in the popular shooter game Borderlands 3 as a tile-matching arcade mini-game (Waldispühl et al. 2020). These projects have seen phenomenal success in terms of outreach, but little has been studied on how players experience these games.

Research that has looked at participant experiences in citizen science has primarily focused on learning outcomes, such as the embedded assessment of performance and data quality (Becker-Klein et al. 2016) and the review of citizen science analyses by Aristeidou and Herodotou

(2020), which examined citizen science as a tool for online learning.

Many scholars have also looked at the motivations of volunteers engaging with digital citizen science. In general, volunteers are motivated by altruism, skill improvement, self-development, career benefits, social interactions, and welfare protection (Clary and Snyder 1999). Citizen scientists are initially motivated by stimulation and self-direction, while retaining motivations include achievement, benevolence, and collaboration, among other factors (Palacin et al. 2020; Rotman et al. 2014). Other motivational factors like gamification and novelty have also shown to improve engagement with citizen science (Jackson 2019; Palacin-Silva et al. 2018).

Player motivations for CSGs can be seen as an extension of the motivations of citizen science volunteers. Players are motivated by the scientific topic, their previous interests in science, the specific research topic, curiosity, and a desire to contribute to research (Curtis 2015; Díaz et al. 2020; Iacovides et al. 2013; Jennett et al. 2016). Continued engagement requires recognizing players for their contributions, task enjoyment, proper pacing, teamwork, learning, and intellectual challenge (Curtis 2015; Iacovides et al. 2013; Jennett et al. 2016). Specifically, scientific communication (i.e., scientists communicating findings with the player base) is a key part of participant engagement (de Vries et al. 2019).

Designing for casual contributors (dabblers), however, requires continued rekindling of motivation via scientific communication and accessible design for casual contribution behavior (cf. "snacking" in Alexandrovsky et al. 2019; Eveleigh et al. 2014). Other motivational factors, such as narrative (Prestopnik and Tang 2015) and gamification (Bowser et al. 2013; Eveleigh et al. 2013; Ponti et al. 2018) have been studied but saw mixed results on efficacy.

There has been relatively less work on the player experience of CSGs, with a few notable exceptions. Díaz et al. (2020) asked players directly about their player experience and quality of experience. They found that a player's experience is influenced by the game design, game elements, the player's strategies, the player's involvement with the scientific community, and the opportunity to help science. Factors such as frustration with learning and lack of progression also affected participation and engagement. Díaz et al. describe the issue of player experience as a trade-off between focusing on the scientific data and the game *qua* game — an issue not only for serious games but also in all citizen science projects started by domain experts (Díaz et al. 2020; Kim et al. 2011).

With respect to the tutorial experience, Díaz et al. (2020) found that players experience a steep learning curve and a lack of understanding in the Quantum Moves CSG, suggesting a need for better tutorials and help pages; this need was also identified by a systematic literature review of general citizen science volunteers (Skarlatidou et al. 2019).

To improve future and existing CSGs — especially regarding their game design for citizen engagement — we must first have a clear understanding of what current CSGs are doing well and poorly with respect to the player experience. Continuing the work of Díaz et al. (2020), we attempt to more comprehensively survey CSG player experiences, collecting nuanced input on the players' needs, frustrations, and learning experiences.

METHODS

To determine the current state of player experiences in CSGs, we sent an online questionnaire to CSG players using a combination of in-game advertisements, social media posts, and game website news posts. Thus, announcements were posted both externally and internally with no specific sampling. There were no inclusion or exclusion criteria for games that players could report on; players could submit survey responses for any game (e.g., Sea Hero Quest or Zooniverse). See Supplemental File 4: Appendix A for the full questionnaire.

Methods were approved by the researchers' institutional ethics board and all participants provided informed consent. Data were collected between April 2019 and May 2021. A total of 237 responses were received and then filtered according to the following criteria: age must be 18–98; responses must specify a valid citizen science game, and duplicate responses were removed. After filtering, 185 valid responses remained; a majority of these (140) were from Foldit, while 45 were from games other than Foldit (EteRNA: 14; Stall Catchers: 14; Eyewire: 7; Skill Lab: Science Detective: 4; Phylo: 3; Living Links: 1; Mozak: 1; Questagame: 1). We expect the skew toward Foldit is because: (1) Foldit has a much larger active player base than other CSGs (Miller and Cooper 2022), (2) Foldit recently promoted an Educational mode attracting students and educators (Miller et al. 2020), and (3) Foldit's developers embedded this survey into their tutorial at a point 16 levels into the game (approximately 1–2 hours of gameplay). Participant ages ranged from 18 to 78 (M = 39.5; σ =17.2). The authors' initial familiarity with these games ranged from passing knowledge to deep expertise; researching and playing these games was done on an as-needed basis for analysis. See Supplemental File 5: Appendix B for details on the games studied.

Open-ended responses were coded using a codebook qualitative content analysis (Guest and MacQueen 2008;

MacQueen et al. 1998). Based on recommendations from literature (Forman and Damschroder 2007), one primary coder wrote the codebook based on a preliminary coding with an effort toward mutually-exclusive codes. Thus, codes were created inductively (data-driven, "conventional") rather than deductively (theory-driven, "directed") (Elo and Kyngäs 2008; Hsieh and Shannon 2005). We acknowledge the reflexive nature of qualitative coding, and thus our findings should be considered interpretive, not objective (Schreier 2012).

The codebook was then iterated on through a code-revise-recode process with the other two coders. After five iterations, the codebook stabilized, and the three coders proceeded to code/recode the remaining responses. All three coders are authors on this paper. An intercoder reliability was calculated across all open-ended responses (each question-part treated as a cell and codes measured as present/absent per coder) using Krippendorff's alpha (Krippendorff 2011), resulting in an alpha of 0.734, which is considered acceptable.

RESULTS

We divide results into five sections: (1) descriptive reports on our respondents' relevant backgrounds, (2) update preferences, (3) tutorial experiences, (4) game difficulty, and (5) open-ended game feedback. See Supplemental File 4: Appendix A for details on the questionnaire.

PARTICIPANT BACKGROUNDS
We asked participants when they started playing the game they were reporting on. Their start dates (n = 175) ranged from June 2008 to March 2021 with the mean around January 2018. Participant education and game expertise follow a bell curve, whereas gameplay frequency is a bimodal distribution (see Supplemental File 3: Table 1). The modal participant is a beginner player with novice education (e.g., took a college course on the scientific topic) and plays games daily. Players reported playing puzzle games most (n = 103), followed by citizen science (n = 99), strategy (n = 98), action/adventure (n = 83), casual (n = 77), role-playing (n = 72), and shooter games (n = 49). We further analyzed players who reported playing games daily and playing citizen science games as a preferred genre (n = 44). Of this subset, participants play puzzle games (n = 28), strategy (n = 28), action/adventure (n = 18), role-playing (n = 18), casual (n = 16), and shooter games (n = 12). From this, we conclude that the modal participant enjoys puzzle and strategy games in addition to their citizen science gaming.

UPDATE PREFERENCES
For the remaining closed-ended results (update preferences, tutorial experiences, and game difficulty), because our data is skewed toward Foldit, we first sought to test whether we can combine all data for analysis (i.e., to analyze our data as coming from one population of CSG players, rather than two populations of Foldit and non-Foldit players). To check this, we performed a chi square test of independence on the contingency table of values for the measurements that could be compared (Foldit, n = 140, non-Foldit, n = 45). We corrected for multiple testing using the Holm-Sidak method. We found that most of the tests were non-significant, with the exception of responses to the statements "I feel stuck" and "I try to get hints from within the game" (adjusted $p <$ 0.05). In this case, Foldit players feel more stuck and seek more hints. However, because most other values were non-significant, we combine all data for the purpose of reporting the remaining results.

As shown in Figure 1, players' update preferences are primarily for more scientific news updates. Secondary preferences include more content, new gameplay modes, and developer updates. Bug fixes and quality of life improvements were important to some players but not others. Finally, social and story/gameplay updates were considered least important.

TUTORIAL EXPERIENCES
Because our responses on the tutorials were largely skewed toward Foldit (n = 98), we report only on Foldit's tutorial. As shown in Supplemental File 2: Figure 2, the beginning of the tutorial is extremely easy, while the end of the tutorial is moderately difficult. With respect to the skills needed to play, participants reported that the Foldit tutorial taught: none (n = 1), some (n = 13), about half (n = 17), most (n = 37), and all (n = 27). Participants further reported the tutorial taught these skills: very poorly (n = 0), poorly (n = 8), fairly (n = 38), well (n = 34), and very well (n = 17). From these bell-curve responses, we conclude that players believe the tutorial teaches most of the skills fairly well.

GAME DIFFICULTY
Participant responses across all games indicated that the puzzles were at a reasonable difficulty. A plurality of 39% of players described that most of the puzzles were satisfyingly challenging but doable, and similar percentages of players said that only some of the puzzles were too easy (50%), too hard (54%), or led to the player feeling stuck (48%). This reasonable difficulty translated well to engagement, as a 41% plurality of players said that most of the puzzles felt engaging. When players were stuck however, they were loath to ask for help — 52% of responses indicated that

players didn't ask others for help and 46% of players didn't look up the answers online (for "most of the puzzles"). Players did generally get hints from within the game when stuck though, with a reasonably even spread of answers across the spectrum. A 38% plurality of players found the game "moderately difficult," followed by 23% responding "slightly difficult."

OPEN-ENDED GAME FEEDBACK

Using the codebook qualitative content analysis (QCA) described in the Methods section, we developed a codebook (see <https://osf.io/yd26a/> for the full codebook) which ultimately had 23 codes capturing: educational value, game structure and pace, supporting alternate play modalities, intrinsic game enjoyment (IGE), intellectual challenge, socialization and community, boring or repetitive play, gamification, power user functionality and quality-of-life features, user interface and input controls, software, paratexts such as game wikis and YouTube videos, developer communication, scientist communication, making scientific contributions, understanding the science of the game, game difficulty, knowledge of how to play, game instructions (both positive and negative reviews), unknown, and no answer.

To quantitatively analyze the results of the QCA, we summed the counts of codes across coders, thereby weighting agreements more heavily while still including all assigned codes. We report only on the top 1–5 categories for each result; however, the full quantitative analysis is available at <https://osf.io/yd26a/>. For each of the five response types (see Appendix A), we explored sums of a variety of subsets of games: Foldit, non-Foldit, Foldit-like (includes Foldit, Eterna, and Eyewire), non-Foldit-like, individual games, and all games. We chose these subsets as capturing the diversity of our sample to the extent that we have sufficient data for analysis. However, for this article we report only on findings which showed marked differences between subsets.

We found participants' favorite aspects of the game as: For Foldit (n = 140): IGE (22.7%), educational value (20.2%), and making scientific contributions (17.0%). For non-Foldit (n = 45): making scientific contributions (23.6%) and IGE(17.4%).

We found participants' least favorite aspects of the game as: For Foldit: confusion about how to play (19.1%), unintuitive user interface (UI) and control scheme (15.9%), poor quality or quantity of instructions and examples (13.2%), and software issues such as bugs, freezing, and crashes (12.4%). For non-Foldit: software issues (16.5%), scientific communication (11.6%), and task quality (9.0%). Notably, scientific communication was highest for Eterna (n = 14) — which relies heavily on a scientific feedback

loop — at 25.0%, and the complaints of task quality were primarily driven by players of Stall Catchers (n = 14) and Eyewire (n = 7) — most often regarding data resolution.

We found that participants would like to see the following updates: For all games, they would like power user functionality/quality-of-life features (19.2%); for Foldit, they would like UI and control scheme (13.6%), better instructions with more examples and other learning assistance (10.9%); and for non-Foldit, they would like scientific communication (16.7%) and software updates (10.3%).

Because the majority of our responses came from Foldit (n = 84; 5 non-Foldit) and Wilcoxon rank sum tests indicated significant differences on the closed-ended questions (*p* < 0.0001), we focus our analysis only on Foldit's tutorial and note this limitation of generality. Their favorite and least favorite aspects were identical: instructions (53.1% favorite; 25.2% least favorite) and pacing and structure (20.9% favorite; 16.0% least favorite).

OVERALL

Overall, Foldit players commented on its instructions (both positively and negatively, 18.1%), their understanding (or lack thereof) the science of the game (10.0%), and their intrinsic game enjoyment (9.3%). Non-Foldit players were more interested in science communication (10.1%), making scientific contributions (9.3%), and gamification (9.3%). For non-Foldit-like games (n = 24), participants focused on the gamification (15.6%), software issues (12.4%), and task quality (9.7%).

DISCUSSION

This work sought to gauge the CSG players' experience through the lens of HCI so that developers can improve their games and collect more and better scientific data. We hope that these findings can also inform project leads, educational specialists, researchers, and other stakeholders of CSGs to critically evaluate the player experiences they are providing and encourage developers to make improvements.

In this section, we discuss the results, followed by takeaways, limitations, and future work.

PARTICIPANT BACKGROUNDS

The most salient findings regarding our participants were that they are novices to the game and its topic, play games frequently, and enjoy puzzle and strategy games alongside their citizen science gaming. These results suggest that citizen science games benefit from having well-designed tutorials, reasons to log in daily, and puzzle and strategy

elements. Good tutorials are a goal of every game, and most citizen science games already have puzzle or strategy elements. However, little has been done to explore daily login incentives, such as daily quests or bonuses (Legner, Eghtebas, and Klinker 2019); this may be an interesting avenue to explore for further development.

UPDATE PREFERENCES

As described in Figure 1, the modal first request from players was more news updates from scientists. This agrees with prior literature that the motivation of contributing to science is one of the most, if not the most, important motivator (Curtis 2015; de Vries et al. 2019; Díaz et al. 2020; Eveleigh et al. 2014, 2013; Iacovides et al. 2013). Along with scientific updates, new content — such as more puzzles or datasets — was ranked highly among the most participants. This finding suggests that, like many long-standing commercial games, the CSGs we studied follow the "games as a service" model, which relies on continuous content updates to maintain engagement and participation (Clark 2014; Delgado and Bazán 2019).

Bug fixes, quality of life improvements, and new ways to play (e.g., new tools, new game modes) spanned a wide range of rankings. However, a closer look at these responses grouped by player sub-populations (experts, new players, dabblers, etc.) would be necessary to better understand which sub-populations are requesting which updates (cf. citizen science profiling, e.g., Aristeidou et al. 2017). Lastly, updates to social features, story updates, and news from developers were least preferred. The first two of these may be an artifact due to the fact that the CSGs studied lack significant story and meaningful social features (besides basic groups and chat functions), or it could speak to a latent trend among CSG players that they are more focused on the task and game mechanics than on the surrounding community and narrative framing. The fact that players care little for developer updates may speak to the motivation of CSG players to be more interested in the science of the game than the game itself. Alternatively (or in addition), improvements to the software may be seen as less exciting than scientific advances or new gameplay features.

TUTORIAL EXPERIENCES

In reporting on the tutorial experiences of CSGs, we are unfortunately limited to describing only Foldit's tutorial. However, we believe this contribution is of value for further consideration of tutorial development in CSGs because several of the themes discussed are agnostic to Foldit's content and mechanics.

As illustrated in Supplemental File 2: Figure 2, Foldit's tutorial begins trivially and ends with moderate to extreme difficulty, our participants report. This demonstrates the steep learning curve participants experience in moving from simple controls to the science challenges presented by the game. Participants also note that the tutorial teaches most of the skills needed to play fairly well, though this still leaves room for improvement — and, conversely, room for confusion. Extending the work of Díaz et al. (2020), these findings show that both of the CSG tutorials studied in-depth (of Foldit in our study and Quantum Moves in theirs) had issues with a steep learning curve.

In open-ended responses, participants praised the tutorial for its gradual progression and clear steps, but felt frustrated when the few instructions were insufficient for solving their problem. They suggested that the tutorial could be improved with more examples, more connection to the science topic, and more and better feedback on their performance. Similar to prior work, both CSG tutorials studied have lacked a strong connection to the scientific subject matter, which caused players to feel lost or confused at how their play was meaningful (Díaz et al. 2020). We further found that Foldit's tutorials violated a playability heuristic by taking away the player's hard-won possessions — in this case, the tools they unlocked by completing previous tutorial levels (Korhonen and Koivisto 2006). Other playability heuristics might also be considered violated upon closer inspection, such as having clear goals, balanced challenge, consistent gameplay, and intuitive controls (Desurvire and Wiberg 2009; Korhonen and Koivisto 2006).

GAME DIFFICULTY

With respect to the game's overall difficulty level, we find that the puzzles are mostly engaging though leaning toward moderate difficulty. However, participants were hesitant to look up help, as the plurality of responses indicated that players rarely looked answers up online or asked others for help. This is concerning since there was evidence that some skills were not adequately taught in the tutorial. If players are hesitant to look up help and those skills are not found in the tutorial, then this can lead to those skills never being taught and players consequently feeling stuck.

Our results agree with previous findings of the difficulty of CSGs (Díaz et al. 2020; Keep 2018). Yet, we take this opportunity to ask whether this is where CSGs would ultimately like to be positioned in the space of gaming. This level of difficulty can lead to disengagement or low performance (Lomas et al. 2017, 2013). Moreover, difficulty is a cognitive barrier, much like the logistical barriers of participation that already muddy citizen science participation (Keep 2018; Spiers et al. 2019). These barriers bias participation and dictate who gets to participate in scientific knowledge production and, ultimately, who benefits from it (Curtis 2018; Keep 2018; Spiers et al. 2019).

However, how much can feasibly be done to make these games easier? The value of some CSGs is employing human cognition and creativity to solve extremely difficult problems; is it the CSG creators' fault for the difficulty of gameplay? We argue yes. Yes, CSG scientists and developers are responsible for lowering barriers to participation of all kinds, especially cognitive ones. As science bears the burden of communicating truth, we must do what we can to make that truth accessible and understandable, enabling participants to engage science and its society-facing problems (Tuite 2014). In doing so, CSGs must aim to improve their instructional design and scientific communication to make even difficult problems accessible to all peoples.

OPEN-ENDED GAME FEEDBACK

Favorite aspects of the game

According to open-ended feedback, one of the primary values of these games is making scientific contributions. This agrees with prior literature on the motivations of CSG players (Curtis 2015; Díaz et al. 2020; Iacovides et al. 2013; Jennett et al. 2016). Moreover, like prior literature, we found that players appreciate the game for having real applications, contributing to scientific knowledge, helping scientists, and feeling like their gameplay matters. Yet, Foldit players often described IGE more so than making scientific contributions. IGE was coded as the value of the game *qua* game (i.e., the gamefulness of the experience). Participants enjoyed the games because they found them relaxing, with aesthetically pleasing color schemes, and they enjoyed simply improving their play and enjoying success with a gameful experience. Foldit players described, for example, the enjoyment of making a stable protein or an interesting [protein] design, and appreciating the coloring and the game's soundscape. It is perhaps because of Foldit's more pronounced gameful and gamification aspects that IGE was the dominant code compared to other games.

Foldit players also commented often on its educational value, which was seen primarily as an "interactive way to see science in action," contrasting static texts and classroom lectures. This is likely due in part to the recent addition of Education mode (Miller et al. 2020); however, even before this mode was introduced, Foldit has been used by many teachers for its real-time interactivity in teaching biochemistry (e.g., Farley 2013). To date, more than 65 teachers and researchers have contacted or collaborated with the Foldit team regarding educational applications (Foldit team, personal communication, 2021).

Least favorite aspects of the game

The least favorite aspects of these games were more diverse. Players described confusion, software issues,

scientific communication, interface and control issues, and task quality as barriers to their enjoyment, engagement, and productive contribution. For example, participants noted slow feedback on puzzle results and a lack of updates on the research being done based on the game, including publications and progress reports.

These least-favorite results can be seen as a takeaway for what CSGs should focus their efforts on improving. Namely, CSG developers can try to: (1) communicate more clearly and quickly regarding what scientific progress is being made and how players are contributing to it, (2) better teach players how to play, (3) listen to player feedback on interface and controls and collaborate with professional UI/UX designers to effect changes, (4) improve task quality, and (5) fix bugs and crashes (cf. Miller and Cooper 2022). Although some aspects will look different for each CSG, such as improving task quality, this refinement starts first and foremost with listening to player feedback.

Updates they would like to see

Curiously, the open-ended responses to update preferences did not align with the closed-ended responses. When given the space to elaborate, participants tended to request power user functionality and quality-of-life features. Several times, new players commented that they had no suggestions because they were too unfamiliar with the game to make good recommendations, resulting in expert players dominating the space with their long-lived frustrations and idiosyncratic desires. Thus, "power user functionality/quality-of-life features" was the highest category for Foldit and non-Foldit games alike, and included, for example, features to improve convenience, new interfaces, more access to the internal game functions, new tools, and features that would improve only some advanced workflows.

This finding is similar to the case study of game company Jagex (developers of the MMORPG RuneScape), who found that crowdsourcing suggestions from players is limited by which players engage with the crowdsourcing, the shape of ideas they generate, and the aspects of design and development that they value (Osborne 2016). In our study, not only were most requests limited to features for veteran users, but the remaining requests tended to reflect the participant's least favorite qualities of the game: the UI and controls, the instructions, scientific communication, or bugs and other software issues.

Favorite and least favorite aspects of the tutorial

Participants were foremost concerned with the instructional design of the tutorial and secondly with the pacing and structure. For example, participants commented positively that the learning progression was gradual, there were

multiple ways to solve the puzzles, and the instructions were easy to follow. However, the instructions and feedback were sometimes not thorough enough, the tutorial doesn't connect to the real science, and the levels often prevented the use of tools previously given to the player which violates standard playability heuristics (Korhonen and Koivisto 2006). Taken together, these findings suggest that tutorials could be improved by additional just-in-time guidance (Gee 2005; Shannon et al. 2013), as well as a clearer link to the science of the game and a better adherence to standard playability heuristics (de Vries et al. 2019; Miller and Cooper 2022).

Overall

Across all open-ended participant feedback, the most common codes for Foldit were instructions, understanding (or lack thereof) the science of the game, and IGE, while for all other games the most common codes were science communication, making scientific contributions, and gamification. The interest of science communication and making scientific contributions is best seen in Eterna, as noted earlier regarding Eterna's close connection with scientific feedback and real lab results. When also excluding Eterna and Eyewire — the two most similar games to Foldit — the remaining 24 participants placed gamification as their top concern, followed by software and task quality. These results are notably driven by participants from Stall Catchers who requested better gamification, software improvements, and higher video resolution. Together, the overall feedback suggests three core — and equally important — recommendations for improving the CSG player experience: make it about the science, make it understandable, and make it fun.

TAKEAWAYS

Throughout all participant feedback, their responses highlighted flaws with the current game instruction, both because participants were confused about how to play and because they didn't understand the science of the game, despite wanting to. This agrees with our initial hypothesis that the player experience is one of frustration, and indicates a need for better teaching of the big picture and the science-game loop, or contribution model (Miller et al. 2021). This was identified especially in Foldit's tutorial, whose instructions were not thorough enough, not connected to real science, and violated standard playability heuristics — such as taking away tools the player had earned, inconsistent gameplay, and unintuitive controls (Desurvire and Wiberg 2009; Korhonen and Koivisto 2006) — all of which can create further confusion by not meeting standards.

For some games like Stall Catchers, gamification was their top concern. CSG teams might consider collaborating with professional game designers to satisfy player interest in gameful or gamified experiences with the task. As reported in the Results, participants like puzzles and strategy games, so tailoring the task design to those preferences is likely to better attract and retain players.

Overall, these results provide confirmation with previous literature that making scientific contributions remains one of the most, if not the most, important motivating factors for CSG participants (Curtis 2015; Díaz et al. 2020; Eveleigh et al. 2014; Iacovides et al. 2013; Jennett et al. 2016). Further, our analysis of participant responses contributes a clearer direction for CSG developers to improve their games, specifically with respect to scientific communication, instructional design, interface and controls, task quality, and software issues. It is important to teach the core gameplay loop and scientific contribution model early (cf. Miller et al. 2021) and iteratively refine your instructions and communication, especially if the project evolves over several years (Keep 2018). Scientific communication is critical since it feeds into the satisfaction of making scientific contributions and can also teach and inform players. In this way, communication is the linchpin of CSG success. To this, we suggest quicker, clearer, more frequent, and more regular scientific communication as the single most important aspect CSG developers could focus on. For more details on implementation of these practices, we refer to recent citizen science literature on communication and accessibility (Paleco et al. 2021; Rüfenacht et al. 2021).

LIMITATIONS

The most notable limitation of this work is a data skew toward Foldit and similar games. However, because our findings are in line with prior work (e.g., Díaz et al. 2020; Miller and Cooper 2022; Tinati et al. 2016), we believe that the contributions of this article remain generalizable to other CSGs. Moreover, our statistical comparisons between Foldit and non-Foldit responses showed non-significant differences for update preferences and game difficulty, suggesting that these aspects may be consistent across CSGs.

Secondly, we note that qualitative coding is a tradeoff of subjective bias and lack of statistical analysis in exchange for depth and nuance in analysis. Future work would benefit from examining player experiences from a quantitative perspective as well. This has not been performed to date because embedding the same gameplay data logging technology (telemetry hooks) in all of these games is currently infeasible, and adding the same telemetry hooks in only one or several games runs a greater risk of skew than in the present study.

CONCLUSIONS

In this article, we surveyed 185 players on their experiences with CSGs to understand the differences between real player experiences and theoretical motivations. Participants responded on 9 different citizen science games, which we analyzed using qualitative content analysis. We found that major concerns included scientific communication, instructional design, user interface and controls, task quality, and software issues.

The next step in this line of research is to make iterative improvements to these CSGs based on the current findings, followed by another survey of the field. CSGs, like other design-centered research, benefit greatly from iteration (Prestopnik 2010). Further, CSG developers would benefit from more communication as a community in order to share ideas and solutions, rather than having separate isolated issues and solving similar problems repeatedly.

DATA ACCESSIBILITY STATEMENT

The anonymized codebook analysis is available at <https://osf.io/yd26a/>. The remaining anonymized data and qualitative analysis are available on request. Please contact the first author for access.

SUPPLEMENTARY FILES

The supplementary files for this article can be found as follows:

- **Supplemental File 1.** Figure 1 Rankings of update preferences. DOI: https://doi.org/10.5334/cstp.500.s1
- **Supplemental File 2.** Figure 2 Summary of Foldit's tutorial difficulty. DOI: https://doi.org/10.5334/cstp.500.s2
- **Supplemental File 3.** Table 1 Participant reports of education level, game expertise, and game-playing frequency. DOI: https://doi.org/10.5334/cstp.500.s3
- **Supplemental File 4.** Appendix A Full questionnaire. DOI: https://doi.org/10.5334/cstp.500.s4
- **Supplemental File 5.** Appendix B Descriptions of games reported on. DOI: https://doi.org/10.5334/cstp.500.s5

ETHICS AND CONSENT

All protocols were approved by the Northeastern University Institutional Review Board (#17-10-07). All participants provided informed consent prior to participation.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Conceptualization, J.A.M.; Data curation, J.A.M., K.G.; Formal analysis, J.A.M., K.G., A.G.; Investigation, J.A.M.; Methodology, J.A.M.; Project administration J.A.M., S.C.; Software, K.G.; Supervision S.C.; Visualization, K.G.; Writing – original draft, J.A.M., K.G.; Writing – review & editing, J.A.M., K.G., A.G., S.C.

AUTHOR AFFILIATIONS

Josh Aaron Miller orcid.org/0000-0001-6783-7567
Northeastern University, US
Kutub Gandhi orcid.org/0000-0002-6004-8925
Northeastern University, US
Anna Gander orcid.org/0000-0001-6170-2653
University West, SE
Seth Cooper orcid.org/0000-0003-4504-0877
Northeastern University, US

REFERENCES

Alexandrovsky, D, Friehs, MA, Birk, MV, Yates, RK and **Mandryk, RL.** 2019. Game Dynamics that Support Snacking, not Feasting. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Barcelona, Spain: ACM, 573–588. DOI: https://doi.org/10.1145/3311350.3347151

Aristeidou, M and **Herodotou, C.** 2020. Online citizen science: A systematic review of effects on learning and scientific literacy. *Citizen Science: Theory and Practice*, 5: 1–12. DOI: https://doi.org/10.5334/cstp.224

Aristeidou, M, Scanlon, E and **Sharples, M.** 2017. Profiles of engagement in online communities of citizen science participation. *Computers in Human Behavior*, 74: 246–256. DOI: https://doi.org/10.1016/j.chb.2017.04.044

Becker-Klein, R, Peterman, K and **Stylinski, C.** 2016. Embedded Assessment as an Essential Method for Understanding Public Engagement in Citizen Science. *CSTP*, 1: 8. DOI: https://doi.org/10.5334/cstp.15

Bowser, A, Hansen, D, He, Y, Boston, C, Reid, M, Gunnell, L and **Preece, J.** 2013. Using gamification to inspire new citizen science volunteers. In: *Proceedings of the First International Conference on Gameful Design, Research, and Applications.* Toronto, Ontario, Canada: ACM, 18–25. DOI: https://doi.org/10.1145/2583008.2583011

Clark, O. 2014. Games As A Service: How Free to Play Design Can Make Better Games, 1st ed. Burlington, MA, USA: Focal Press.

Clary, EG and **Snyder, M.** 1999. The Motivations to Volunteer: Theoretical and Practical Considerations. *Curr Dir Psychol Sci*, 8: 156–159. DOI: https://doi.org/10.1111/1467-8721.00037

Cooper, S. 2011. A framework for scientific discovery through video games (Doctoral Dissertation). University of Washington.

Curtis, V. 2018. Who Takes Part in Online Citizen Science?. In: *Online Citizen Science and the Widening of Academia. Springer International Publishing*. Cham, 45–68. DOI: https://doi.org/10.1007/978-3-319-77664-4_3

Curtis, V. 2015. Motivation to Participate in an Online Citizen Science Game: A Study of Foldit. *Science Communication*, 37: 723–746. DOI: https://doi.org/10.1177/1075547015609322

d'Ornellas, MC, Cargnin, DJ and **Prado, ALC.** 2015. Evaluating the Impact of Player Experience in the Design of a Serious Game for Upper Extremity Stroke Rehabilitation. In: *Proceedings of the 15th World Congress on Health and Biomedical Informatics*. IOS Press, 363–367.

de Vries, M, Land-Zandstra, A and **Smeets, I.** 2019. Citizen scientists' preferences for communication of scientific output: a literature review. *Citizen Science: Theory and Practice*, 4: 2. DOI: https://doi.org/10.5334/cstp.136

Delgado, JCS and **Bazán, P.** 2019. Educational Serious Games as a Service: Challenges and Solutions. *JC&ST*, 19: e07. DOI: https://doi.org/10.24215/16666038.19.e07

Desurvire, H and **Wiberg, C.** 2009. Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration. In: Ozok, AA and Zaphiris, P. (Eds.), Online Communities and Social Computing, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 557–566. DOI: https://doi.org/10.1007/978-3-642-02774-1_60

Díaz, C, Ponti, M, Haikka, P, Basaiawmoit, R and **Sherson, J.** 2020. More than data gatherers: exploring player experience in a citizen science game. *Qual User Exp*, 5: 1. DOI: https://doi.org/10.1007/s41233-019-0030-8

Elo, S and **Kyngäs, H,** 2008. The qualitative content analysis process. *J Adv Nurs*, 62: 107–115. DOI: https://doi.org/10.1111/j.1365-2648.2007.04569.x

Eveleigh, A, Jennett, C, Blandford, A, Brohan, P and **Cox, AL.** 2014. Designing for dabblers and deterring drop-outs in citizen science. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, Ontario, Canada: ACM, 2985–2994. DOI: https://doi.org/10.1145/2556288.2557262

Eveleigh, A, Jennett, C, Lynn, S and **Cox, AL.** 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: ACM, 79–82. DOI: https://doi.org/10.1145/2583008.2583019

Farley, PC. 2013. Using the Computer Game "FoldIt" to Entice Students to Explore External Representations of Protein Structure in a Biochemistry Course for Nonmajors. *Biochem. Mol. Biol. Educ*, 41: 56–57. DOI: https://doi.org/10.1002/bmb.20655

Forman, J and **Damschroder, L,** 2007. Qualitative Content Analysis. In: *Advances in Bioethics*. Elsevier, 39–62. DOI: https://doi.org/10.1016/S1479-3709(07)11003-7

Gee, JP. 2005. Learning by Design: Good Video Games as Learning Machines. *E-Learning and Digital Media*, 2: 5–16. DOI: https://doi.org/10.2304/elea.2005.2.1.5

Guest, G and **MacQueen, KM.** (Eds.), 2008. Handbook for team-based qualitative research. Altamira, Lanham.

Hsieh, H-F and **Shannon, SE.** 2005. Three Approaches to Qualitative Content Analysis. *Qual Health Res*, 15: 1277–1288. DOI: https://doi.org/10.1177/1049732305276687

Iacovides, I, Jennett, C, Cornish-Trestrail, C and **Cox, AL.** 2013. Do games attract or sustain engagement in citizen science?: a study of volunteer motivations. In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1101–1106. DOI: https://doi.org/10.1145/2468356.2468553

Jackson, C. 2019. Characterizing Novelty as a Motivator in Online Citizen Science (PhD Thesis). Syracuse University.

Jennett, C, Kloetzer, L, Schneider, D, Iacovides, I, Cox, A, Gold, M, Fuchs, B, Eveleigh, A, Mathieu, K, Ajani, Z and **Talsi, Y.** 2016. Motivations, learning and creativity in online citizen science. *JCOM*, 15: A05. DOI: https://doi.org/10.22323/2.15030205

Keep, BE. 2018. Becoming Expert Problem Solvers: A Case Study in what Develops and how. Stanford University.

Kim, S, Robson, C, Zimmerman, T, Pierce, J and **Haber, EM.** 2011. Creek watch: pairing usefulness and usability for successful citizen science. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Vancouver, BC, Canada: ACM, 2125–2134. DOI: https://doi.org/10.1145/1978942.1979251

Korhonen, H and **Koivisto, EMI.** 2006. Playability heuristics for mobile games. In: *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services – MobileHCI '06*. Helsinki, Finland: ACM Press, 9. DOI: https://doi.org/10.1145/1152215.1152218

Krippendorff, K. 2011. Computing Krippendorff's Alpha-Reliability. Retrieved from https://repository.upenn.edu/asc_papers/43.

Legner, L, Eghtebas, C and **Klinker, G.** 2019. Persuasive Mobile Game Mechanics For User Retention. In: *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. Presented at the CHI PLAY '19: The Annual Symposium on Computer-Human Interaction in Play*, ACM, Barcelona Spain, 493–500. DOI: https://doi.org/10.1145/3341215.3356261

Lomas, JD, Koedinger, K, Patel, N, Shodhan, S, Poonwala, N and **Forlizzi, JL.** 2017. Is Difficulty Overrated?: The Effects of Choice, Novelty and Suspense on Intrinsic Motivation in Educational Games. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* Denver, Colorado, USA: ACM, 1028–1039. DOI: https://doi.org/10.1145/3025453.3025638

Lomas, JD, Patel, K, Forlizzi, JL and **Koedinger, KR.** 2013. Optimizing challenge in an educational game using large-scale design experiments. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris, France: ACM, 89–98. DOI: https://doi.org/10.1145/2470654.2470668

MacQueen, KM, McLellan, E, Kay, K and **Milstein, B.** 1998. Codebook Development for Team-Based Qualitative Analysis. *CAM Journal*, 10: 31–36. DOI: https://doi.org/10.1177/1525822X980100020301

Miller, JA and **Cooper, S.** 2022. Barriers to Expertise in Citizen Science Games. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, p. in press. DOI: https://doi.org/10.1145/3491102.3517541

Miller, JA, Horn, B, Guthrie, M, Romano, J, Geva, G, David, C, Sterling, AR and **Cooper, S.** 2021. How do Players and Developers of Citizen Science Games Conceptualize Skill Chains? In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. DOI: https://doi.org/10.1145/3474671

Miller, JA, Khatib, F, Hammond, H, Cooper, S and **Horowitz, S.** 2020. Introducing Foldit Education Mode. *Nat Struct Mol Biol*, 27, 769–770. DOI: https://doi.org/10.1038/s41594-020-0485-6

Newman, G, Wiggins, A, Crall, A, Graham, E, Newman, S and **Crowston, K.** 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10: 298–304. DOI: https://doi.org/10.1890/110294

Olsen, T, Procci, K and **Bowers, C.** 2011. Serious games usability testing: How to ensure proper usability, playability, and effectiveness. In: *International Conference of Design, User Experience, and Usability*. Springer, 625–634. DOI: https://doi.org/10.1007/978-3-642-21708-1_70

Osborne, D. 2016. Tales From a Crowdsourced Game Design.

Palacin, V, Gilbert, S, Orchard, S, Eaton, A, Ferrario, MA and **Happonen, A.** 2020. Drivers of participation in digital citizen science: Case Studies on Järviwiki and safecast. *Citizen Science: Theory and Practice*, 5. DOI: https://doi.org/10.22323/2.15030205

Palacin-Silva, MV, Knutas, A, Ferrario, MA, Porras, J, Ikonen, J and **Chea, C.** 2018. The Role of Gamification in Participatory Environmental Sensing: A Study In the Wild. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal, QC, Canada: ACM, 1–13. DOI: https://doi.org/10.1145/3173574.3173795

Paleco, C, García Peter, S, Salas Seoane, N, Kaufmann, J and **Argyri, P.** 2021. Inclusiveness and Diversity in Citizen Science. In: Vohland, K, Land-Zandstra, A, Ceccaroni, L, Lemmens, R, Perelló, J, Ponti, M, Samson, R and Wagenknecht, K. (Eds.), *The Science of Citizen Science*. Cham: Springer International Publishing, 261–281. DOI: https://doi.org/10.1007/978-3-030-58278-4_14

Ponti, M, Hillman, T, Kullenberg, C and **Kasperowski, D.** 2018. Getting it Right or Being Top Rank: Games in Citizen Science. *CSTP*, 3: 1. DOI: https://doi.org/10.5334/cstp.101

Prestopnik, N. 2010. Theory, Design and Evaluation – (Don't Just) Pick any Two. *THCI*, 2: 167–177. DOI: https://doi.org/10.17705/1thci.00021

Prestopnik, NR and **Tang, J.** 2015. Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior*, 52: 492–506. DOI: https://doi.org/10.1016/j.chb.2015.05.051

Rienzo, A and **Cubillos, C.** 2020. Playability and player experience in digital games for elderly: A systematic literature review. *Sensors*, 20: 3958. DOI: https://doi.org/10.3390/s20143958

Rotman, D, Hammock, J, Jenny, P, Hansen, D, Boston, C, Bowser, A and **He, Y.** 2014. Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries. In: *IConference 2014 Proceedings*. iSchools. DOI: https://doi.org/10.9776/14054

Rüfenacht, S, Woods, T, Agnello, G, Gold, M, Hummer, P, Land-Zandstra, A and **Sieber, A.** 2021. Communication and Dissemination in Citizen Science. In: Vohland, K, Land-Zandstra, A, Ceccaroni, L, Lemmens, R, Perelló, J, Ponti, M, Samson, R and Wagenknecht, K. (Eds.), *The Science of Citizen Science*. Cham: Springer International Publishing, 475–494. DOI: https://doi.org/10.1007/978-3-030-58278-4_24

Schreier, M. 2012. Qualitative content analysis in practice. Los Angeles: SAGE.

Schrier, K. 2017. Designing Learning with Citizen Science and Games 4, 9.

Shannon, A, Boyce, A, Gadwal, C and **Barnes, DT.** 2013. Effective Practices in Game Tutorial Systems. In: *Proceedings of the 8th International Conference on the Foundations of Digital Games*. ACM, 8.

Skarlatidou, A, Hamilton, A, Vitos, M and **Haklay, M.** 2019. What do volunteers want from citizen science technologies? A systematic literature review and best practice guidelines. *JCOM*, 18: A02. DOI: https://doi.org/10.22323/2.18010202

Spiers, H, Swanson, A, Fortson, L, Simmons, B, Trouille, L, Blickhan, S and **Lintott, C.** 2019. Everyone counts? Design considerations in online citizen science. *JCOM*, 18: A04. DOI: https://doi.org/10.22323/2.18010204

Sullivan, DP, Winsnes, CF, Åkesson, L, Hjelmare, M, Wiking, M, Schutten, R, Campbell, L, Leifsson, H, Rhodes, S, Nordgren, A, Smith, K, Revaz, B, Finnbogason, B, Szantner, A and **Lundberg, E.** 2018. Deep learning is combined with

massive-scale citizen science to improve large-scale image classification. *Nat Biotechnol*, 36: 820–828. DOI: https://doi.org/10.1038/nbt.4225

Tasnim, RA and **Eishita, FZ.** 2021. Analyzing the Distinctive Impact of Personality Traits on Serious Gameplay Experience. In: *2021 IEEE 9th International Conference on Serious Games and Applications for Health (SeGAH).* IEEE, 1–8. DOI: https://doi.org/10.1109/SEGAH52098.2021.9551856

Tinati, R, Luczak-Rösch, M, Simperl, E and **Hall, W.** 2016. Because science is awesome: studying participation in a citizen science game. In: *Proceedings of the 8th ACM Conference on Web Science*. Hannover, Germany: ACM, 45–54. DOI: https://doi.org/10.1145/2908131.2908151

Tuite, K. 2014. GWAPs: Games with a Problem. In: *Proceedings of the 9th International Conference on the Foundations of Digital Games*. Ft. Lauderdale, FL, USA: ACM, 7.

Waldispühl, J, Szantner, A, Knight, R, Caisse, S and **Pitchford, R.** 2020. Leveling up citizen science. *Nat Biotechnol*, 38, 1124–1126. DOI: https://doi.org/10.1038/s41587-020-0694-x

Wiemeyer, J, Nacke, L, Moser, C and **'Floyd' Mueller, F,** 2016. Player Experience. In: Dörner, R, Göbel, S, Effelsberg, W and Wiemeyer, J. (Eds.), *Serious Games*. Cham: Springer International Publishing, 243–271. DOI: https://doi.org/10.1007/978-3-319-40612-1_9

Wiggins, A and **Crowston, K.** 2015. Surveying the citizen science landscape. FM 75. DOI: https://doi.org/10.5210/fm.v20i1.5520