Increasing player engagement, retention and performance through the inclusion of educational content in a citizen science game

Rogério de Leon Pereira
 Anthony Tan
 Andrea Bunt
 Olivier Tremblay-Savard
 tremblao@cs.umanitoba.ca
 University of Manitoba
 Department of Computer Science
 Winnipeg, Manitoba, Canada

ABSTRACT

Citizen science game developers often choose to transform or simplify the scientific problem of interest to make the game accessible to a wider audience, which is critical for collecting enough data. A potential drawback of this design decision is the disconnect that it creates between the players and the science. We are interested in studying ways to re-establish that connection and give back to the players in the form of educational content embedded in those games. Here, we briefly describe a preliminary study focused on analyzing different formats (text, cartoon or video), then we mainly focus on studying ways to present educational material (full versus gradual access) and evaluating their effects on players in a citizen science game. The results indicate that although there is no clear consensus on which format is more appropriate, providing gradual access to the content as a reward for completing levels can increase the engagement, the retention and the performance of players.

CCS CONCEPTS

• Human-centered computing → User studies; • Theory of computation → Convergence and learning in games; • Applied computing → Computer games.

KEYWORDS

Citizen Science Games, Human Computing Games, Games With A Purpose, Educational Content, Engagement, Retention, Performance.

ACM Reference Format:

Rogério de Leon Pereira, Anthony Tan, Andrea Bunt, and Olivier Tremblay-Savard. 2021. Increasing player engagement, retention and performance through the inclusion of educational content in a citizen science game. In *The 16th International Conference on the Foundations of Digital Games (FDG)*

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 the author(s) 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](https://permissions.acm.org).

FDG'21, August 3–6, 2021, Montreal, QC, Canada

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8422-3/21/08...\$15.00

<https://doi.org/10.1145/3472538.3472554>

2021 (FDG'21), August 3–6, 2021, Montreal, QC, Canada. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3472538.3472554>

1 INTRODUCTION

In recent years, human computing and crowdsourcing have become increasingly popular to tackle problems that are non-trivial for traditional computing approaches, or to complement machine learning approaches when training data is scarce [24]. Those problems that are targeted by human computing approaches are usually tailored for human abilities such as perception, pattern recognition, instinctive knowledge, reasoning and coordination. Problems can be decomposed into tasks that are embedded into Games With A Purpose, or GWAPs. These rely on the engaging aspect of the games for motivating the participants to solve the initial problem of interest¹. One of the first examples of a successful GWAP was the ESP game [71], which was focused on the image labelling problem. In this game, two players were shown the same image, and without communicating to each other, they had to type in words that described the image. The players would win the matches as soon as they had entered the same word, and this word usually turned out to be a great label for the image. Many other GWAPs have been developed since then targeting problems from many different disciplines, such as deep learning [67], astronomy [3, 9], neuroscience [34], health sciences [1] and molecular biology [13, 30, 40].

Most GWAPs based on scientific problems are also called *citizen science games*. Many times, those scientific problems need to be “translated” into puzzles that are more accessible to players who do not have the necessary scientific background for completing the original tasks. Citizen science games vary a lot in their so-called *level of abstraction* (i.e. how transformed/simplified the in-game puzzles are from the initial problem). In some games like Foldit [13], which asks its players to fold proteins in 3D following the complex rules of biochemistry, the game stays very close to the original data and scientific problem. In other games like Phylo [30], where players are asked to match colored blocks, the puzzles are much more abstracted from the original bioinformatics problem of creating a multiple sequence alignment of nucleotide sequences. Both

¹Note that the terms *GWAP* and *serious game* are sometimes used interchangeably. However, a serious game is more often used to describe a game whose primary goal is training or teaching. A GWAP is a game whose main purpose is to complete a task that is hard for computers but easier for humans.

strategies for citizen science games have advantages: games with puzzles that are relatively similar to the problem of interest can better motivate and engage players who are interested in science [15], whereas more abstract games have the advantage of simply being more accessible to a larger audience.

We are interested in developing ways to give back to the citizen science game players and engage them more in science by adding educational content in citizen science games that have a high level of abstraction from the original problem. In this paper, our goal is to evaluate if presenting educational material can enhance player engagement, retention and performance in a citizen science game. More specifically, we first briefly explore in a preliminary study different formats for presenting the scientific content: with text, cartoons or videos. Then, we study two different strategies for giving access to the educational material in a citizen science game already developed in our lab [64]: (1) full access, and (2) gradual access linked to progress in the game. Our results suggest that educational content in a citizen science game can have a positive effect on engagement, retention and performance, but the way it is delivered in the game makes a difference.

2 RELATED WORK

2.1 Learning in citizen science and crowdsourcing

There are crowdsourcing applications that are centered around learning. Duolingo is a good example: it is primarily focused on teaching a new language, and in its first few years, it was inviting participants to translate texts from the web through the *Immersion* tool [70]. Dontcheva *et al.* have explored combining crowdsourcing and learning in their platform *LevelUp for Photoshop*, which embeds a learning framework inside a crowdsourcing platform to edit photos using Adobe Photoshop [19]. The platform uses interactive tutorials to teach volunteers how to retouch photos: users are shown how to complete each step through short texts, and receive points once the “mission” is completed. The authors have shown that in this context, the learning component results in good crowd-sourced work and also enhances worker motivation. Crowdclass introduced the scaffolding presentation of advanced scientific concepts to crowdworkers tasked with galaxy image classification [39]. The idea was to divide the educational material into three learning modules (easy, medium, hard) to avoid overwhelming novice users (after reading all the material of one learning module, the content of the next one is shown). The authors selected to present the learning content in text format, as a way to avoid the “time commitment” of watching a video [39]. Crowdclass was shown to be effective at teaching the scientific concepts within the tasks, although it did not seem to affect the users’ productivity or motivation.

Citizen science applications and crowdsourcing games which are not abstracted from the scientific problem typically offer the players an opportunity to learn while they complete the tasks or communicate with other players through discussion forums. For example, in the crowdsourcing application Galaxy Zoo [44], which asks users to classify images of galaxies, volunteers used the discussion forum to share information about unique-looking galaxies, which was then shared to the Galaxy Zoo investigators [5]. Foldit also has a discussion forum and tools allowing players to share

strategies (recipes) for folding proteins, which allowed players to learn from their peers and even propose improved strategies of their own [32]. In many cases, citizen science applications will indirectly teach the volunteers: the crowdsourcing tasks will often motivate the volunteers to consult external resources to learn more about what they are doing and why [29].

Citizen science and crowdsourcing games with a higher level of abstraction typically do not contain much educational material embedded inside them. Tutorials in many citizen science games focus mostly on explaining the basic rules of the game (e.g. Phylo [68], Eterna [6], EyeWire [21]). Some games present a certain amount of educational information but only on the game’s webpage (e.g. EyeWire [22], ColonyB [69]) or Youtube channel (e.g. Eterna [7]). Phylo [68] recently added a story mode to their game and included some short bits of “knowledge” that appear after completing certain chapters, but those are mostly very short definitions (one or two sentences).

However, to the best of our knowledge, the effect of integrating educational content in citizen science games on player engagement, retention and performance has not been studied previously.

2.2 Engagement and motivation in citizen science games

We explored the literature to find what aspects of citizen science projects makes participants engaged and motivated. In the context of citizen science games, engagement mostly refers to *behavioral engagement* (being focused on a task, paying attention to what is presented and the overall involvement within the game), and motivation is normally referring to *intrinsic motivation* (participating in the project for reasons that are inherent to it, and not because of external pressures or rewards).

Bowser *et al.* [4] used a gamified citizen science app to investigate the different nuances of motivation on two self-identifying user groups: citizen science volunteers and gamers. They have found that the citizen scientists preferred having more autonomy and had a perception of the inherent value of the app, as a way to contribute to science. On the other hand, the gamers desired more guidance and valued the entertainment provided by the interaction with the app. Raddick *et al.* [54] explored the different motivations of citizen science volunteers in the context of Galaxy Zoo. They have observed that the majority of participants had multiple motivations for participating in this citizen science project (e.g. interest in astronomy, desire to contribute, beauty and vastness of space, fun), and many were driven by a desire to learn and have fun at the same time.

A study was conducted on the Foldit game to better understand the players’ motivations [15]. Once again, the results showed that among the most important motivations for participation were the interest in science and the desire to contribute to scientific research. Moreover, the intellectual challenge provided by the game and community interactions were also key for the continuous engagement of the players. In another study on Foldit [49], the authors also found out that prior domain knowledge or prior gaming experience were important elements of engagement. On the one hand, the participants classified as gamers completed more levels, making more game progress, but on the other hand, the participants

with more domain knowledge played more sessions, demonstrating more interest in returning to the game.

Another way to motivate participation is through curiosity. Malone *et al.* [45] in their *Framework for a Theory of Intrinsically Motivating Instruction* describe several intrinsic motivation theories, among them two related to curiosity: (1) sensory curiosity; and (2) cognitive curiosity. The first involves sensory stimuli in an environment (in our case, in the game): it could be a sound, an object's movement, changes in brightness, etc. The second is evoked by the expectation of changing higher-level cognitive structures. It could be thought of as the ambition to improve one's knowledge. The authors point out that there is an ideal level of informational complexity to evoke curiosity. With respect to the user's existing knowledge, the environment should not be too simple nor too complicated. Curiosity has also been studied in the specific context of crowdsourcing. Law *et al.* [38] found out that curiosity interventions can be used to increase workers' retention in a paid crowdsourcing environment. Their strategy was making use of the information gap theory: the smaller the knowledge gap, the higher is the effect of the curiosity stimulus. First, they asked a question relevant to the task, to induce curiosity and create an information gap. After, they asked participants to guess if they did not know the answer, to make them aware of the knowledge gap. Finally, more information was revealed as they progressed in the task to help them to close the gap. Dai *et al.* [16] have used curiosity-inducing mechanisms to increase crowdsourcing workers' output. They interspersed a sequence of tasks with task-irrelevant activities to produce a micro-diversion. In other words, a micro-diversion is a small entertainment intervention between boring or stressful tasks. One of the tested micro-diversions was a comic that was shown one page at a time after each completed task. They showed that micro-diversions can improve the worker retention rate while retaining the work quality.

This related work on engagement and motivation in citizen science projects suggests the importance of designing a fun game experience, but not at the cost of the scientific knowledge dissemination, which is important for a majority of participants.

3 DESCRIPTION OF THE CITIZEN SCIENCE GAME

The citizen science game used for the experiments was developed previously in our lab [64] to solve a problem in bioinformatics called the *genome sorting problem*. Let S_1 and S_2 be two strings of characters on an alphabet A . Each character $c \in A$ represents a gene, and each string represents an ordered sequence of genes (also called a *gene order*). The solution to the genome sorting problem is the shortest sequence of events that can transform S_1 into S_2 , using a predefined set of possible operations, which represent biological events that can occur during the evolution of genomes. This problem is NP-hard in the presence of duplicate genes [10], and the main objective of this citizen science project is to devise new heuristics from the players' solutions.

In the game, we use colored shapes to represent the different genes (instead of using characters). In each level, two rows of colored shapes represent the two gene orders being compared. The top row is the *target* row, while the bottom row is the *mutable* row, i.e. the one that has to be modified by the player to transform it into

the target. Players can modify the mutable row by selecting one or multiple genes (colored shapes), and applying one of the three possible operations: duplication, deletion or inversion. In the case of a duplication, the player must also choose where the duplicated copies should go in the mutable sequence. A deletion will make the selected genes disappear from the mutable sequence. Finally, an inversion event can only be applied to selections of genes that contain the *terminus of replication* (represented by a black dot near the middle of the row). This is simply because inversions are known to occur mostly around this *terminus* in the gene orders considered in this game. A screenshot of the game is presented in Figure 1.

The game has several levels and the players play them in sequence. A *top score*, representing an achievable number of moves required to solve the puzzle, is associated with each level, since it was shown to lead players to produce better solutions [46]. Each event performed by the player increases their score by 1 (the score starts at 0), and the goal is to complete the puzzle with the lowest possible score. At the end of the level, the player receives points according to their performance following these rules:

- 50 points only, if the score surpasses the top score (by any amount);
- 100 points, if the score matches exactly the top score;
- $100 + (\text{topScore} - \text{score}) * 50$, if the player beats the top score.

3.1 Educational material and CODEX

Since the citizen science game is aiming to solve a problem in bioinformatics, the educational material included in it is also about topics in bioinformatics. We chose 6 different topics related to the game: (1) genomes, chromosomes, DNA, genes and nucleotides, (2) RNAs, mRNAs and tRNAs, (3) comparative genomics, (4) gene orientations and gene orders, (5) genome evolution, mutations, gene duplications, gene losses, and inversions, and (6) how to calculate the distance between genomes and what is the genome sorting problem. Those six topics were selected because they are all complementary and when read in succession, they explain the scientific problem behind the game.

This scientific content was produced by an expert in bioinformatics to assure its quality and accuracy. All the educational material was displayed in the CODEX, which is what we called the interface for this embedded in-game encyclopedia. The CODEX was accessible by clicking on a small white book icon on the bottom right side of the screen (see bottom right of Figure 1). All the scientific content used in this work is included in the supplementary material.

4 PRELIMINARY STUDY ON THE EDUCATIONAL MATERIAL FORMATS

Before studying the effects of educational content in a citizen science game, establishing which type of format to use for presenting it seemed like a logical and necessary first step. So far as we know, the effectiveness of different formats of educational material has not been studied in the context of citizen science and crowdsourcing games. That motivated us to conduct a preliminary study to investigate what educational content format would fit better in a citizen science game, in terms of potential knowledge gain and user interest. Here we prepared three versions for each of the topics 1, 2 and 4 (see section 3.1): one in text format with figures, one in

Figure 1: Screenshot of the human computing game. The small white book on the bottom right is the button to access the educational content.

cartoon format and one in video format. Only three topics out of the six were selected for this study to keep the experiment time reasonably short, and topic 3 was skipped in favour of topic 4 which contained more advanced topics and was more suited to test knowledge gain (see section 4.3). The specific research questions that were guiding this preliminary study are listed below.

4.1 Research questions

- (1) Which educational material format (text, cartoons, videos) has the best learning outcome in a citizen science game?
- (2) Which educational material format (text, cartoons, videos) is preferred by players in a citizen science game?

4.2 Description and justification of the three formats

The text format is the simplest one: for all three topics, it consists of approximately one page of text with some figures that were added to help demonstrate some of the concepts (Figure 2a). Testing the text format was an obvious choice, since it has played and still plays a major role in education and learning [33, 36]. One of the main advantages of text is its stability, which helps the reader in constructing meaning from it [37]. The stability of the text allows the readers to slow down in certain areas that are more complex, and regress in the text to review some previous concepts [2, 62]. Moreover, many studies have shown that textual information is remembered better when accompanied with figures [8, 42, 43, 58]. Mental imagination,

i.e. the creation of mental images of the elements described in a scientific text, was also proven to be a very efficient strategy to increase knowledge acquisition and retention [41].

The cartoons were created using *ComiPo!* [14], a software for creating manga-style short stories. In the cartoons, which were about 2 pages long, characters are introduced and they present the scientific material through their interactions (Figure 2b). Educational cartoons like these are often called *concept cartoons*, and they constitute another approach for introducing scientific concepts in the classroom. They are similar to traditional cartoons (comic strips), but the goal is to present different ideas or points of views through the characters' discussions, which help the students explore and integrate the concepts [31, 35, 61]. Concept cartoons were successfully used to evaluate student preferences in terms of learning environments and approaches [35, 61], to assist with problem based learning [27], as alternative assessment methods [11, 12, 28] and as efficient tools to identify and overcome student misconceptions [20, 51, 66]. Cartoons have the advantage of catching the eye, engaging the reader, and they are able to present explanations one step at a time by dividing them into separate panels [63].

The videos were created using the *VideoScribe* [65] software tool, which creates hand-drawn animated videos, using a sequence of figures provided by the user. During the produced videos, a narrator reads the same text that was presented in the text format (and cartoon, but without the characters and story which are only present in the cartoon format). The narration was synchronized

Figure 2: Example of the educational content in the three different formats.

with the hand-drawn figures also using *VideoScribe*. The videos for all topics were on average 2:45 minutes long (Figure 2c). Videos were the third format selected because several studies have shown the benefits of the modality effect: the educational practice of showing visual information and related textual information auditorily, through narration for example [23, 47, 55, 59]. Indeed, multimedia formats utilizing visually presented animations combined with narration were shown in these articles to be more effective than just images and text. A possible explanation is that using two different modalities “minimizes working-memory load” [47], because “auditory texts make use of an additional cognitive resource” [55]. On the other hand, some studies on animated demonstrations, in the context of software tutorials, have shown that while they can help with learning the steps faster, knowledge retention seems to be worse than with text-only tutorials [52, 53].

4.3 Experiment and results

In the preliminary study, the independent variable was the format of the educational material (text with figures, cartoon or video). The dependent variables were the knowledge gain (quantitative), and user interest and engagement (qualitative). In order to measure the knowledge gain, we used a 15-questions pre-test (presented at the very beginning of the experiment), and three 5-questions post-tests which were taken after the presentation of each topic. The post-tests contained different questions than the ones of the pre-test, but they were covering the same material. Recruited participants were asked to play the game for five minutes in between each educational material presentation and the corresponding post-test, to evaluate if the participants were able to retain the information they learned after completing a different task. This was a within-subject study, as each topic was presented in a different format (and the order of formats was randomized across participants). User interest and engagement data was collected using semi-structured interviews.

We recruited 18 participants (5 female and 13 male) in a university for convenience. The average experiment duration time was 45 minutes and each participant received a \$10 gift card as compensation for their time commitment.

The results of the preliminary study showed that although the average knowledge gain was the highest for the text format (44.22% increase in correct answers for the text format, compared to 40.98%

and 40.09% for the cartoon and video formats respectively), a one-way ANOVA test has shown no statistically significant difference between the three educational material formats ($f(2,49) = 0.05$, $p = 0.95$). Moreover, the effect size was very low (partial eta squared = 0.0019). In terms of the favorite format during the experiment, 55% of participants voted for the video, whereas 28% and 17% of participants voted for the cartoon and text formats respectively. When asked about which format the participants would like to use to learn something new, cartoons and videos both obtained 39% of the votes and the text obtained 22%. More details about the preliminary study are available in this work-in-progress paper [17].

In summary, this preliminary study did not result in a clear consensus about which format was the best for presenting educational content in a citizen science game. Videos seemed to be more popular, but learning outcomes were slightly higher for the text format, although not significantly. For these reasons, and because text was a bit easier to integrate into the game, we decided to use only the text format for the main study presented below.

5 MAIN STUDY: EFFECT OF INCLUDING EDUCATIONAL CONTENT IN A CITIZEN SCIENCE GAME

After completing the preliminary study to guide our choice of educational material format, we moved on to the main study, which focused on evaluating the effects of presenting educational content (either gradually or all at once vs a control condition with no access) on players in the same citizen science game. Our research questions for this main study are presented below.

5.1 Research questions

- (1) Does presenting educational content gradually or all at once affect player retention in the game?
- (2) Does presenting educational content gradually or all at once has an effect on the player engagement with the CODEX and the game in general?
- (3) Does presenting educational content gradually or all at once affect player performance in the game?

Figure 3: Visual representation of the three conditions. Participants of the Control Group had no access to the educational content. The participants of the Knowledge Group had access to all 6 topics since the beginning of the experiment. The participants of the Curiosity Group unlocked one new topic every 3 levels, and when that happened the CODEX button started flashing to indicate that new content was available.

5.2 Experimental design

In this study, the independent variable is the presence of the educational material. It was tested using three different groups (see Figure 3 for a visual representation of the conditions):

- (1) Control Group: participants in this group played the game with no educational material.
- (2) Knowledge Group: all the educational material (six topics described in section 3.1) was available to the participants beginning at the first level of the game.
- (3) Curiosity Group: participants were gradually getting access to the educational material, *i.e.* the participants had access to more educational content the more levels they played. The first educational topic was unlocked at level 3, and every 3 levels after that until level 18 (six topics in total). A visual cue (flashing button) was displayed to indicate to the players that new educational content was available, to stimulate their curiosity.

The dependent variables are performance (quantitative), retention (quantitative) and engagement (quantitative / qualitative). To measure performance, we compared the points obtained during the game sessions of the different groups. In the context of this game, recall that the points are obtained according to the number of moves used to solve a puzzle (score) and the top score provided for the respective puzzle (as described in section 3). To verify the retention of participants we checked if they played more levels than they were asked for. The participants' engagement with the CODEX was measured using the amount of time spent reading it. The players' engagement with the game was evaluated using the surveys described below. This was a between-subject study, with the participants being allocated to one of the three different test conditions through a round-robin format.

This study was conducted entirely online. We used two well-known social media platforms [25, 26] and a citizen science network [60] to recruit participants through a recruitment web page. The experiment was designed to be completed in 30 minutes, however,

the time used to complete the study varied because the participants could complete the requirements of the study in one or multiple game sessions (the progress was saved on the server). There was no monetary compensation for participating in this study.

After reading the consent form, the participants followed a link on the recruitment page to load the game in their browser. Before playing it, the participants needed to register in the game by creating a username and password. When hitting the play button for the first time, the game showed a message asking the participants to agree with the consent form (the participants were not able to play the game without accepting it). Then the participants learned the rules of the game by playing a 6-level tutorial.

After finishing the tutorial, the game showed a message asking the participants to answer a first survey consisting mainly of demographic questions. Note that there was an option to skip the survey. After playing the first 11 required levels, the game prompted the players to answer a second survey, which was also optional, about their experience with the game and the educational material. We used Google Forms as the surveying platform for both questionnaires. It was possible to access both surveys in the game's main menu, in case the participants unintentionally skipped them. The first survey was the same for everyone, but we had two different versions of the second survey, one for the control group and another for both the knowledge and the curiosity groups. This was necessary because only the groups 2 and 3 had access to the educational content. All the surveys are presented in the supplementary material.

After playing the first 11 levels, the participants could keep playing the game if they wanted to, since 30 levels in total were available in the game (see section 5.3).

5.3 Game levels and data set

For the purpose of this study, we decided to use a synthetic data set to have better control over the gradual increase in difficulty and the achievable score for each puzzle. We slowly increased the

difficulty of the levels to make sure that levels were accessible at the beginning of the game and to keep the players challenged as they progress (i.e. keep them in between boredom and anxiety, in the so-called flow channel [57]).

To create each puzzle we used a script to generate a random synthetic genome (the mutable row). The script then applies a number of events to it, to generate a second genome (the target row). The number of events used to transform the first genome into the second is used in the puzzle as the top score. The script uses 3 parameters: (1) the genome size, (2) the number of events, and (3) the average size of the events. The first two parameters varied according to the game level to gradually increase the difficulty, while the third parameter was the same for all the puzzles. Even though we required the participants to play only 11 levels, we created 30 levels, allowing them to keep playing if they wanted.

6 RESULTS

6.1 Participants demographics

We recruited 96 participants (32 in each group). Participation was considered completed when (1) the consent form was accepted, (2) a username and password were created, and (3) the first 11 levels were played. 29.17% of participants dropped out of the experiment before having accomplished those 3 tasks (28 in total, 9 in the control group, 10 in the knowledge group, and 9 in the curiosity group) and were excluded from the results. From the remaining 68 participants, 80.88% answered both surveys (55 in total). The average age of the participants who answered the surveys was 29 (SD = 10) and the average time spent playing video games was 15 hours/week (SD = 17). 47.3% of the participants had a post-secondary degree at the undergraduate level. 58.2% had taken biology classes in high school. All the quantitative results were based on the data generated by the 68 participants and the qualitative results are based on the 55 who answered both surveys.

6.2 Quantitative results

6.2.1 Analysis of retention. The average number of levels played in this study by all participants was 14.75. Recall that the participants were asked to play the first 11 levels only, but they could keep playing the game after completing the study. We analyzed the total number of levels played to verify if any condition was able to keep the players more engaged and retain them for longer. Participants of the *curiosity* group played more levels than any other group, with 16.78 levels played on average by this group. We ran a one-way ANOVA test (Figure 4) and we found a main effect of the group on the number of levels played ($f(2, 65) = 4.73, p = 0.01$). Significant pairwise differences were found between the curiosity group and the control group, as well as the curiosity group and the knowledge group ($p = 0.05$ and $p = 0.02$ respectively).

We also ran a linear regression test to analyze the correlation between the time spent in the CODEX and the average number of levels played for players in the knowledge and curiosity groups. The results indicate that the time spent reading the educational content can reliably predict the number of levels played ($p < 0.001$) and 56% of the variance in the average number of levels played can be predicted from the time spent reading ($R^2 = 0.5589$).

Figure 4: A one-way ANOVA test of the number of levels played by group. $f(2, 65) = 4.73, p = 0.01$. The error bar represents the standard error.

Figure 5: A one-way ANOVA test of the average obtained points by readers and non-readers. $f(1, 66) = 9.75, p = 0.002$. The error bar represents the standard error.

6.2.2 Analysis of engagement with the CODEX. The average time spent by participants in the knowledge group interacting with educational content was 1.22 minutes, while the average time spent in the CODEX by participants in the curiosity group was 4.47 minutes. We performed a one-way ANOVA test and found that the participants of the curiosity group spent significantly more time in the CODEX than the ones of the knowledge group ($f(1, 43) = 4.97, p = 0.03$).

6.2.3 Analysis of performance. All participants received points for each completed level. The points varied according to the players' performance (recall the points system described in section 3). The average points per level, on the first 11 required levels, was 106.35 for all participants. The mean points per level for the control, knowledge and curiosity groups were 106.72 (SD = 5.79), 108.68 (SD = 6.71), and 103.75 (SD = 14.22) respectively. We ran a one-way ANOVA test on the average number of points per level and we found no significant differences between the three groups ($f(2, 65) = 1.47, p = 0.24$).

Table 1: Distribution of players as readers and non-readers and the average number of levels played. The value inside parentheses next to the number of non-reader players represents the number of those non-readers who did not open the CODEX at all.

Groups	# of Players			AVG levels played (std error)		
	Control	Knowledge	Curiosity	Control	Knowledge	Curiosity
non-readers	23	17 (4)	11 (3)	14 (0.76)	13.41 (0.71)	14.18 (1.21)
readers	0	5	12	0	13.40 (0.93)	19.17 (1.40)

The previous result showed no statistical difference between the three different groups with regards to the performance in the first 11 required levels. However, having access to the educational content does not mean that players were actually spending time reading it. This is why we decided to re-categorize all the participants from all groups into: (1) *readers*: a group of participants who spent at least 60 seconds reading the educational content and (2) *non-readers*: a group of participants who had no access to the educational content (control group) or who spent less than 60 seconds interacting with it. The 60 seconds threshold was chosen because it would take at least one minute to read one CODEX entry.

We ran a one-way ANOVA test on the average number of points obtained per level (for the 11 required levels) and we found a significant difference between the two groups of readers and non-readers (Figure 5). It turns out that participants who interacted more with the CODEX (the readers) obtained more points per level on average, which means that they were performing better in the game.

We also ran a linear regression test to analyze the correlation between the time spent in the CODEX and the average number of points obtained per level for the first 11 levels. The results indicate that the time spent reading the educational content can reliably predict the average points per level ($p < 0.001$) and 44% of the variance in average points per level can be predicted from the time spent reading ($R^2 = 0.4358$).

6.2.4 Additional exploratory analysis on readers/non-readers. The number of players categorized as readers and non-readers (using the 60 seconds threshold) and the average number of levels played by each subgroup are shown in Table 1. As the control group had no access to the educational material, they are formed only by non-readers. The knowledge group has more non-readers than readers. The participants in the curiosity group are almost evenly distributed as readers and non-readers. The number of players who never opened the CODEX at all was almost the same in both the knowledge (4) and curiosity (3) groups.

When looking at the average number of levels played by readers and non-readers in each group, surprisingly we can see that all the subgroups have a similar average except the readers of the curiosity group. In other words, the fact that the curiosity factor encourages players to spend more time reading the CODEX, coupled with the fact that readers of the curiosity group play a lot more levels, created the significant difference in the average number of levels completed by the curiosity group (as shown in section 6.2.1).

6.3 Overview of survey answers

Two of the authors went through the survey answers looking for quotes that could provide more context to explain the participants'

interactions with the game and the scientific content. Our second goal was to collect information on how to better design the educational material.

Some participants from the control group actually mentioned that they would have liked to get more explanations about the science behind the game, and how the actions in the game correlate to biological events.

"Is there a correlation between the shapes and proteins? I didn't see the explanation there." - P43 (Control)

"Some of the controls are quite strange, they don't seem familiar and some things u think u can do, but u cant [...] but if you make a video showing in the gene itself with what those controls are, it would be nice." - P44 (Control)

We believe that P44 was referring to the rules for the different operations (what events are allowed to happen, which is based on biology). These rules were explained in the tutorial but not having access to the educational material in the control group meant that participants like P44 could not see the biological context behind them.

On the other hand, some participants of the knowledge and curiosity groups appreciated the idea of having the educational content presented inside the game, allowing them to learn more about key terms related to biology and bioinformatics.

"The puzzle game was fun, also the side notes with explanations about biology." - P10 (Curiosity)

"Great job! I hope to learn more about this project and what you learn from it." - P16 (Knowledge)

Moreover, the strategy of gradually giving players access to more educational content as a reward every time they completed three levels kept players from the curiosity group engaged:

"I personally appreciated how it wasn't a heavy explanation all at once, teaching things as we unlock new levels was nice." - P4 (Curiosity)

"The level 12 was very hard. I am at level 14 right now, but I will keep playing to see how much information I can add in my white-orange book." - P21 (Curiosity)

The CODEX material in this study was designed as text with figures, presented in a pop-up window and not directly within the game scene. Moreover, the figures presented in the text did not reuse directly the assets (shapes) presented in the game to explain the biological concepts. As a result of this, some volunteers pointed out that it was not always clear how each component of the game and the puzzles were related to the bioinformatics concepts presented in the CODEX.

"My knowledge in Biology is close to nothing. It was fun to read about it, but I am not sure how it is connected to the game though." - P34 (Curiosity)

"I wasn't sure how I was contributing towards some project. I understood that my actions reflected events found in nature, but I felt a disconnect between the game and nature." - P26 (Knowledge)

"It could help if the game explained the basics of science within the game and then link to more information." - P27 (Knowledge)

"I think if you tried [...] to make the figures we have to match the actual things we were learning about instead of the squares etc it would help fixating what i just learned over the levels." - P4 (Curiosity)

This feedback could be very useful for citizen science game designers who wish to integrate scientific content in their games. These comments demonstrate a potential benefit of integrating the educational content as much as possible within the gameplay and the puzzles themselves, perhaps in the form of a story mode.

Another important feedback received through the volunteers' responses is that despite our best efforts to create simple and accessible texts, the quantity and depth of the content presented could be challenging for players.

"Even though i know you tried to make the text as succinct as possible (since biology can't really be shortened all that much without losing quality in the explanation) it was still somewhat heavy." - P4 (Curiosity)

"It should have more figures and less text (I guess?)." - P20 (Knowledge)

"IMHO the texts are a little bit too difficult." - P34 (Curiosity)

These comments highlight the importance of dividing the educational material as much as possible into bite-sized topics. In this study, the texts, which were about one page long, could have been shortened even more to make them more accessible. This could also explain why the participants in the knowledge group didn't engage as much with the educational content. Having access to everything all at once was probably too much, and players from that group gave up more quickly on reading all the information.

7 DISCUSSION

The results indicate that educational content in a citizen science game can have a positive effect on retention, engagement and performance. However, it is also clear that it is not just a matter of including scientific content in the game, but also finding the right approach to present it.

Retention, as measured by the total number of levels played, was higher for participants in the curiosity group only, which was getting progressive access to the educational content. We did not see any significant difference in terms of the average number of levels played between the control group and the knowledge group, which had access to the entire material since the beginning of the game. Most of the change in retention was coming from the fact that the curiosity approach was engaging more participants to spend

more time in the CODEX, which was supported by some survey answers.

Many factors could play a role in this increased engagement with the educational content and the game with the curiosity approach. Firstly, there is obviously the curiosity stimulus, which in this case was the book icon that started flashing whenever new content was available. However, the stimulus was only encouraging players to open the CODEX; there was no obligation to read the new unlocked entry in order to keep playing. Secondly, the idea of parcelling the educational content might have reproduced the effects of the curiosity strategies explored in existing literature [16, 38]. Unlocking a new entry in the CODEX for every 3 completed levels potentially created a micro-diversion, as studied in [16], which was beneficial for keeping players interested in the game. Also, as the educational content was presented in a logical order and each entry had information related and complementary to the previous one, this could have enabled the participants to feel that their knowledge on those subjects was increasing proportionally to their progress in the game, effectively shortening the knowledge gap as was seen in [38]. Thirdly, the fact that entries in the CODEX were unlocked as players were progressing in the game could have had the effect of a reward, which is an important gamification component that has been shown to positively affect engagement in the human computing literature [18, 48, 50, 56]. These observations pave the way for further exploration of how these different elements of motivation (visual stimulus, curiosity and reward) interact together and contribute to increased engagement with the game and the educational content in future work.

The last aspect that we were interested in was how scientific content in our citizen science game could affect player performance, as measured by the average number of points obtained per level. Note that the content that was included in the CODEX was directly related to the problem tackled by the game, but there was no specific game tips or strategies presented in it. At first, we did not observe any significant difference in terms of average points per level between the participants of the three studied groups. However, having access to the educational content does not mean that players are interacting with it. That insight opened a new avenue of analysis, leading us to group all the participants as readers and non-readers using a threshold of at least 60 seconds spent in the CODEX. The results suggest that interacting with the educational material for a minimum amount of time can help in improving players' performance, as the average number of points per level was significantly higher for readers. Moreover, there was a significant correlation between the amount of time spent in the CODEX and the average number of points obtained per level. Although part of this correlation could be resulting from more conscientious participants who are both thoroughly reading the educational material and trying to obtain good scores on the puzzles, interacting more with the CODEX appears to be enhancing players' performance in the game.

Based on our results, some lessons for the design of citizen science games and the presentation of educational content can be drawn. First of all, even if the game itself is abstracted from the original scientific problem, it can be beneficial to include educational material to better explain the science behind it. Second, it seems clear that progressive access to the content, *i.e.* presenting

it in small doses, results in more engagement with the content than full access. Making sure that each unlocked topic is very short and accessible should also be a priority. Third, we believe that the stronger the integration of the educational content is with the game, the more it will have positive effects on engagement, retention and performance. In our study, the best form of integration was made by unlocking a new educational entry for every three levels completed. However, the CODEX was still presented in a pop-up window and not directly within the puzzle interface. Finding ways to fully integrate the educational content in the main game interface and establishing clear connections between the game elements and the scientific problem could further increase the observed benefits.

8 LIMITATIONS

Some of the results of the preliminary study on different educational material formats relied on the evaluation of the knowledge gain using pre/post quizzes. The questions were designed by one of the authors who has 5+ years of experience teaching bioinformatics courses at the university level. Long or compound questions were avoided to make sure the questions were very specific, which facilitated the creation of equivalent questions. However, we acknowledge that despite our best efforts, the two quizzes might not have the exact same difficulty level.

We also acknowledge that the two crowds of participants who might be interested in a game like this, *i.e.* "gamers" and "citizen scientists", could potentially interact with the game and the scientific content differently. Although we have collected some information on how much time participants were usually playing videos games every week, it is not so simple to categorize them. Further studies should be devoted to getting a better understanding of how these different types of players interact with the educational content.

Furthermore, some limitations come from the online setting. Participants could play all the required levels in a single session or in multiple sessions over several days. Our analysis of the game logs indicate that more than 70% of participants completed the experiment in one session. However, it is not clear if completing the experiment using multiple sessions could have influenced the performance in the game or the interaction time with the educational content in some ways. Similarly, some participants could fill out the second survey immediately after completing the 11 required levels or keep playing and come back to it later, which might have affected the players' responses. Another consequence of the online setting was that we had to use surveys to get feedback from the participants. The answers obtained from surveys are usually less detailed and clear than the ones that can be obtained from a semi-structured interview.

Finally, these results are based on a study of one specific citizen science game in bioinformatics. It would be interesting to see in the future how well these findings would translate to other types of games, with different kinds of puzzles, levels of abstraction, gameplay loops and scientific topics or areas.

9 CONCLUSION

This study focused on testing two different methods of presenting educational content in a citizen science game, and evaluating the effect on player retention, engagement and performance compared

with a control group. We observed that there are more efficient ways than others to present scientific content in the game. The only approach that had an effect on retention was the one called *curiosity*, in which players were given gradual access (with a visual cue) to the content as they progressed in the game. Similarly, players were more engaged with the educational material and the game in the curiosity group. In terms of performance in the game, we observed that players who interacted for more than 60 seconds with the scientific content obtained more points per level and there was a significant correlation between time spent reading and points obtained.

These results highlight the importance of explaining the science behind the citizen science games, even in the context of a game that is more abstracted from the original scientific problem. Including educational content in these games can be beneficial for all the parties involved. On the one hand, it can help citizen science game developers by retaining players longer and improving solution quality. On the other hand, the users can feel more engaged in the game, learn about science and better appreciate the importance of their contribution. In citizen science games, very often the focus is mainly on the collection of data, and not as much on the learning experience for players. We believe that our results are useful for citizen science game developers, as it would encourage them to put in the extra effort to create and include good teaching material inside their games.

Future work should be devoted to finding other potentially more efficient approaches to integrate educational content into citizen science games. Deeper analyses of different types of educational material formats and how they can be connected to the main gameplay loop should be conducted. Investigating how exactly the gradual release of scientific information, such as in the curiosity condition, can act as a guide to help players progress in the game would also be interesting. We argue that the inclusion and integration of this kind of teaching material should be part of the main design decisions during the development of citizen science games, and not just an afterthought. Consequently, creating tools to guide and help citizen science game developers with the integration of educational content should also be investigated in future work.

ACKNOWLEDGMENTS

The authors would like to thank *Brackeys* game development and the Phylo community for helping us to advertise this study on Twitter. The authors would also like to thank all the participants from the preliminary and the main study who made this work possible. This work was supported by a Genome Canada / CIHR grant (Bioinformatics and Computational Biology Competition).

REFERENCES

- [1] Graham Alvare and Richard Gordon. 2015. CT brush and CancerZap!: two video games for computed tomography dose minimization. *Theoretical Biology and Medical Modelling* 12, 1 (2015), 7.
- [2] Evalyn Bayle. 1942. The nature and causes of regressive movements in reading. *The Journal of Experimental Education* 11, 1 (1942), 16–36.
- [3] François Bouchy, Maxime Marmier, and Oliver Turner. 2018. Light-Curve Classifications Scrutinized by UNIGE Astronomers. <http://mmos.ch/news/2018/01/26/light-curve-classifications-scrutinized-by-unige-astronomers.html>
- [4] Anne Bowser, Derek Hansen, Jennifer Preece, Yurong He, Carol Boston, and Jen Hammock. 2014. Gamifying citizen science: a study of two user groups. In *Proceedings of the companion publication of the 17th ACM conference on Computer*

- supported cooperative work & social computing. 137–140.
- [5] Carolin Cardamone, Kevin Schawinski, Marc Sarzi, Steven P Bamford, Nicola Bennert, CM Urry, Chris Lintott, William C Keel, John Parejko, Robert C Nichol, et al. 2009. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society* 399, 3 (2009), 1191–1205.
- [6] Stanford University Carnegie Mellon University. 2010. eterna. <https://eternagame.org/> Accessed: 2020-09-08.
- [7] Stanford University Carnegie Mellon University. 2018. eterna. <https://www.youtube.com/watch?v=V2qtNyJ3KQg> Accessed: 2020-09-08.
- [8] Russell N Carney and Joel R Levin. 2002. Pictorial illustrations still improve students' learning from text. *Educational psychology review* 14, 1 (2002), 5–26.
- [9] CCP. 2017. EVE Online: Project Discovery. <https://www.eveonline.com/discovery/>
- [10] Xin Chen, Jie Zheng, Zheng Fu, Peng Nan, Yang Zhong, Stefano Lonardi, and Tao Jiang. 2005. Assignment of orthologous genes via genome rearrangement. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2, 4 (2005), 302–315.
- [11] Christine Chin and Lay-Yen Teou. 2009. Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education* 31, 10 (2009), 1307–1332.
- [12] Christine Chin and Lay-Yen Teou. 2010. Formative assessment: Using concept cartoon, pupils' drawings, and group discussions to tackle children's ideas about biological inheritance. *Journal of Biological Education* 44, 3 (2010), 108–115.
- [13] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, et al. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756.
- [14] Web Technology Corp. 2014. ComiPo! <https://www.comipo.com/en/index.html>
- [15] Vickie Curtis. 2015. Motivation to participate in an online citizen science game: A study of Foldit. *Science Communication* 37, 6 (2015), 723–746.
- [16] Peng Dai, Jeffrey M. Rzeszutski, Praveen Paritosh, and Ed H. Chi. 2015. And Now for Something Completely Different: Improving Crowdsourcing Workflows with Micro-Diversions. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) (CSCW '15). Association for Computing Machinery, New York, NY, USA, 628–638. <https://doi.org/10.1145/2675133.2675260>
- [17] Rogerio de Leon Pereira, Andrea Bunt, and Olivier Tremblay-Savard. 2019. Promoting Learning and Engagement in Human Computing Games: a Study of Educational Material Formats. (2019).
- [18] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. 2011. Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems*. 2425–2428.
- [19] Mira Dontcheva, Robert R Morris, Joel R Brandt, and Elizabeth M Gerber. 2014. Combining crowdsourcing and learning to improve engagement and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3379–3388.
- [20] Fatma Ekici, Erhan Ekici, and Fatih Aydin. 2007. Utility of Concept Cartoons in Diagnosing and Overcoming Misconceptions Related to Photosynthesis. *International Journal of Environmental and Science Education* 2, 4 (2007), 111–124.
- [21] Eyewire. 2019. Eyewire. <https://eyewire.org/> Accessed: 2020-09-08.
- [22] Eyewire. 2019. Into the Brain: The Science of Brain Mapping. <https://science.eyewire.org/> Accessed: 2020-09-08.
- [23] Paul Ginns. 2005. Meta-analysis of the modality effect. *Learning and instruction* 15, 4 (2005), 313–331.
- [24] Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 3, 2 (2016), 119–131.
- [25] Facebook Inc. 2004. Facebook. <http://www.facebook.com> Accessed: 2020-06-01.
- [26] Twitter Inc. 2006. Twitter. <http://twitter.com> Accessed: 2020-06-01.
- [27] Didem Inel and Ali Günay Balım. 2013. Concept cartoons assisted problem based learning method in science and technology teaching and students' views. *Procedia-Social and Behavioral Sciences* 93 (2013), 376–380.
- [28] Sebnem Kandil Incec. 2008. Use of Concept Cartoons as an Assessment Tool in Physics Education. *Online Submission* 5, 11 (2008), 47–54.
- [29] Charlene Jennett, Laure Kloetzer, Daniel Schneider, Ioanna Iacovides, Anna Cox, Margaret Gold, Brian Fuchs, Alexandra Eveleigh, Kathleen Methieu, Zoya Ajani, et al. 2016. Motivations, learning and creativity in online citizen science. *Journal of Science Communication* 15, 3 (2016).
- [30] Alexander Kawrykow, Gary Roumanis, Alfred Kam, Daniel Kwak, Clarence Leung, Chu Wu, Eleyine Zarour, Luis Sarmiento, Mathieu Blanchette, Jérôme Waldspühl, et al. 2012. Phylo: a citizen science approach for improving multiple sequence alignment. *PLoS one* 7, 3 (2012), e31362.
- [31] Brenda Keogh and Stuart Naylor. 1999. Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education* 21, 4 (1999), 431–446.
- [32] Firas Khatib, Seth Cooper, Michael D Tyka, Kefan Xu, Ilya Makedon, Zoran Popović, and David Baker. 2011. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* 108, 47 (2011), 18949–18953.
- [33] M Khine. 2013. Critical analysis of science textbooks. *Evaluating instructional effectiveness*. Netherlands: Springer (2013).
- [34] Jinseop S Kim, Matthew J Greene, Aleksandar Zlateski, Kisuk Lee, Mark Richardson, Srinivas C Turaga, Michael Purcaro, Matthew Balkam, Amy Robinson, Baradia F Behabadi, et al. 2014. Space-time wiring specificity supports direction selectivity in the retina. *Nature* 509, 7500 (2014), 331.
- [35] Ian Kinchin. 2004. Investigating students' beliefs about their preferred role as learners. *Educational Research* 46, 3 (2004), 301–312.
- [36] Mary Koppal and Ann Caldwell. 2004. Meeting the challenge of science literacy: Project 2061 efforts to improve science education. *Cell biology education* 3, 1 (2004), 28–30.
- [37] Robert B Kozma. 1991. Learning with media. *Review of educational research* 61, 2 (1991), 179–211.
- [38] Edith Law, Ming Yin, Joslin Goh, Kevin Chen, Michael A. Terry, and Krzysztof Z. Gajos. 2016. Curiosity Killed the Cat, but Makes Crowdwork Better. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 4098–4110. <https://doi.org/10.1145/2858036.2858144>
- [39] Doris Jung-Lin Lee, Joanne Lo, Moonhyok Kim, and Eric Paulos. 2016. Crowdclass: Designing classification-based citizen science learning modules. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.
- [40] Jeehyung Lee, Wipapat Kladwang, Minjae Lee, Daniel Cantu, Martin Azizyan, Hanjoo Kim, Alex Limpacher, Snehal Gaikwad, Sungroh Yoon, Adrien Treuille, et al. 2014. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences* 111, 6 (2014), 2122–2127.
- [41] Claudia Leopold and Richard E Mayer. 2015. An imagination effect in learning from scientific text. *Journal of Educational Psychology* 107, 1 (2015), 47.
- [42] W Howard Levie and Richard Lentz. 1982. Effects of text illustrations: A review of research. *ECTJ* 30, 4 (1982), 195–232.
- [43] Joel R Levin. 1987. On empirically validating functions of pictures in prose. *The psychology of illustration* 1 (1987), 51–86.
- [44] Chris J Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M Jordan Raddick, Robert C Nichol, Alex Szalay, Dan Andreescu, et al. 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.
- [45] Thomas W Malone, Robert Calfee, Bill Clancey, Allan Collins, Jamey Friend, George Furnas, and Laura Gould. 1981. Toward a theory of intrinsically instruction* Motivating. (1981).
- [46] Anjum Matin, Mardel Maduro, Rogerio de Leon Pereira, and Olivier Tremblay-Savard. 2020. Effect of Timer, Top Score and Leaderboard on Performance and Motivation in a Human Computing Game. In *International Conference on the Foundations of Digital Games*. 1–10.
- [47] Richard E Mayer and Roxana Moreno. 1998. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology* 90, 2 (1998), 312.
- [48] Elisa D Mekler, Florian Brühlmann, Alexandre N Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71 (2017), 525–534.
- [49] Josh Aaron Miller, Uttkarsh Narayan, Matthew Hantsbarger, Seth Cooper, and Magy Seif El-Nasr. 2019. Expertise and engagement: re-designing citizen science games with players' minds in mind. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–11.
- [50] Lennart E Nacke and Christoph Sebastian Deterding. 2017. The maturing of gamification research. *Computers in Human Behaviour* (2017), 450–454.
- [51] Stuart Naylor and Brenda Keogh. 2013. Concept Cartoons: what have we learnt? *Journal of Turkish Science Education* 10, 1 (2013).
- [52] Susan Palmiter and Jay Elkerton. 1993. Animated demonstrations for learning procedural computer-based tasks. *Human-Computer Interaction* 8, 3 (1993), 193–216.
- [53] Susan Palmiter, Jay Elkerton, and Patricia Baggett. 1991. Animated demonstrations vs written instructions for learning procedural tasks: a preliminary investigation. (1991).
- [54] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2009. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925* (2009).
- [55] Ralf Rummel, Judith Schweppe, Anne Fürstenberg, Tina Seufert, and Roland Brünken. 2010. Working memory interference during processing texts and pictures: Implications for the explanation of the modality effect. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 24, 2 (2010), 164–176.
- [56] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. 2017. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human*

- Behavior* 69 (2017), 371–380.
- [57] Jesse Schell. 2019. *The Art of Game Design: A book of lenses*. AK Peters/CRC Press.
 - [58] Wolfgang Schnotz. 2002. Commentary: Towards an integrated view of learning from text and visual displays. *Educational psychology review* 14, 1 (2002), 101–120.
 - [59] Wolfgang Schnotz. 2011. Colorful bouquets in multimedia research: A closer look at the modality effect. *Zeitschrift für Pädagogische Psychologie* (2011).
 - [60] SciStarter. 2018. SciStarter. <http://scistarter.org> Accessed: 2020-06-01.
 - [61] Matthew Sexton. 2010. Using Concept Cartoons to Access Student Beliefs about Preferred Approaches to Mathematics Learning and Teaching. *Mathematics Education Research Group of Australasia* (2010).
 - [62] Wayne L Shebilske and L Starling Reid. 1979. Reading eye movements, macro-structure and comprehension processes. In *Processing of visible language*. Springer, 97–110.
 - [63] Joel Shurkin. 2015. Science and Culture: Cartoons to better communicate science. *Proceedings of the National Academy of Sciences* 112, 38 (2015), 11741–11742.
 - [64] Akash Singh, Chris Drogaris, Elena Nazarova, Mathieu Blanchette, Jérôme Waldispühl, Anjum Ibna Matin, and Olivier Tremblay-Savard. 2017. A human-computation platform for multi-scale genome analysis. *Association for the Advancement of Artificial Intelligence (www.aaai.org)*. Retrieved from https://www.humancomputation.com/2017/papers/100_human-computation-platform.pdf (2017).
 - [65] Sparkol. 2018. VideoScribe. <https://www.videoscribe.co/en/>
 - [66] Philip Stephenson and Paul Warwick. 2002. Using concept cartoons to support progression in students' understanding of light. *Physics Education* 37, 2 (2002), 135.
 - [67] Devin P Sullivan, Casper F Winsnes, Lovisa Åkesson, Martin Hjelmare, Mikaela Wiking, Rutger Schutten, Linzi Campbell, Hjalti Leifsson, Scott Rhodes, Andie Nordgren, et al. 2018. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nature biotechnology* (2018).
 - [68] McGill University. 2010. Phylo. <https://phylo.cs.mcgill.ca/> Accessed: 2020-09-08.
 - [69] McGill University. 2016. ColonyB. <http://csb.cs.mcgill.ca/colonyb/> Accessed: 2020-09-08.
 - [70] Luis Von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 1–2.
 - [71] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 319–326.