JunkCorp: A Narrative Task Framing Approach to Gamified Citizen Science

Anne Nielsen
an18@student.aau.dk
20183946, Denmark

Marcus O. Hyttel
mhytte18@student.aau.dk
20183838, Denmark

Martin D. Sørensen
mdsa18@student.aau.dk
20183669, Denmark

ABSTRACT

This paper reports our findings regarding the potential of narrative task framing as a motivator in citizen-science projects dealing with image segmentation and annotation tasks.

Research showed there to be an informational gap in direct comparisons between point-based gamification and a narrative approach in the context of citizen science. Therefore, we developed two versions of a singular mobile application to investigate user motivation while completing the task of annotating images of litter. One version with only point-based gamification elements and one with the task diegetically integrated into a storyline.

The application was developed using Unity, with Android as the target platform. The JunkCorp Version of the application allows users to submit annotations while following a story and freely continue contributing once the story has been completed. Features like a leaderboard, progress bars, and group tasks, based on research into gamification and user motivation as well as early user feedback, were implemented into both versions of the application.

The application was evaluated using the User Motivation Inventory, the Gameful Experience Questionnaire (GAMEFULQUEST), and qualitative feedback from interviews with our users.

Our results showed an increase in intrinsic, identified and integrated regulation for the users of the JunkCorp Version. However, our analysis showed no statically significant relation between user motivation and the version of the application.

These results suggest that narrative task framing could help retain motivation in scientific tasks. More research on the area, testing with more participants, and different implementations of narrative task framing are needed to determine its true potential as an intrinsic motivator.

CCS CONCEPTS

• **Human-centred computing → Interaction design theory, concepts and paradigms**.

KEYWORDS

Citizen Science, Image Annotation, Gamification, Intrinsic Motivation, Task Framing, Narrative Gamification, Meaningful Framing

1 INTRODUCTION

Machine learning is increasingly being used for automation within the scientific field. Machine learning may refer to when machines are taught to recognise objects by training on labelled data or image annotations. For a machine learning algorithm to recognise an object, it will need a lot of annotations to train on. The number of annotations may vary but usually, at least, 1000 annotated images are needed [23].

A lot of research has been done on how to collect a large number of annotations effectively. Many datasets have started to accumulate annotations through citizen science. Citizen science is the process of including volunteers in gathering or processing scientific data [17]. When making annotations through citizen science, volunteers need to be motivated to continue to annotate images. A way to keep users motivated is through gamification. This is due to its observed positive effects on intrinsic motivation [40, 53].

Gamification is usually implemented through points, badges, and leaderboards. In recent years, there has been an increase in interest towards utilising narrative elements when making gamified systems. There are citizen science games that have successfully utilised story-based gamification to keep people engaged with citizen science and make it feel more like a game by incorporating the scientific task as a diegetic part of the storyline [35, 36].

Other than the addition of game elements, there is a lack of research on how to intrinsically motivate people to continuously contribute to citizen science projects. While story-based gamification is present in the context of citizen science, due to games like*Forgotten Island* [35], there are yet to be, as far as we are aware, any direct comparisons between a point-based gamified citizen science project and a narratively task-framed version of the same project. Furthermore, we have not been able to find any projects that utilise narrative task framing in image segmentation and annotation tasks.

Therefore, this paper aims to explore the differences between a gamified only and a narratively task-framed version of the same application in terms of user motivation for citizen science projects dealing with image segmentation and annotation.

2 RELATED WORK

In this section, we cover the basics of citizen science and gamification. We look into both point-based and narrative gamification, and how they have been added to citizen science projects and image annotation tasks.

2.1 Citizen Science

Crowdsourcing is a term used to describe researchers asking the public for help in scientific inquiries. The counterpart to crowdsourcing is citizen science. Citizen science is what is performed by the people that accept the researchers' invitation to assist in the collection of data [52]. The definition of citizen science varies between different fields of work. Haklay et al. have made a general definition of citizen science stating that "it includes the generation of scientific data. . . , engages volunteers over a large area. . . , and address a politically relevant issue" [17]. We refer to this definition any time the concept *citizen science* is used in this paper.

Recent research suggests that data gathered and generated through citizen science has become an important resource for scientists and researchers [10]. There are many popular citizen science projects.

For example, the projects *Eyewire* and *Foldit* have large amounts of volunteers [20] and the platform *Zooniverse* [54] lets people create citizen science projects and have them in one place, which allows people to easily find projects they would like to volunteer for.
Despite the growing amount of research on the concept, there are still discussions, within the scientific community, on whether citizen science is a valid method, often pointing towards the quality of the generated data [4]. To counter the doubt of validity, researchers have begun creating guidelines for how to utilise citizen science to its fullest [4, 15, 33].
Citizen science relies on input from volunteers. This means that the method is dependent on maintaining the motivation of users to keep their contributions to the project continuous. A lack of user motivation will inevitably lead people to stop contributing.

2.2 User Motivation

There are a number of ways to motivate users to engage with an application. Motivation can be classified as one of two types. These types are extrinsic motivation and intrinsic motivation. Intrinsic motivation is "not dependent on external incentives or pressure, but rather provide their own satisfactions and joys" [39]. Extrinsic motivation is the exact opposite as it is influenced by external factors. This means that if you are intrinsically motivated you might participate in a project because you think it is interesting or fun. On the other hand, you might be extrinsically motivated to participate in a project because you are being paid to do so.
In research, *self-determination theory* (SDT) is often used to explain and influence user motivation [38]. SDT seeks to enhance intrinsic motivation through psychological needs satisfaction. The three needs that are outlined by SDT are:

- *Competence*, learning and mastering skills. Feeling able to complete a task.
- *Autonomy*, being in control of one's own behaviour and goals.
- *Relatedness*, a sense of belonging or an attachment to other people.

Intrinsic motivation is seen as the preferred type as it contributes to happiness, satisfaction [38], and long-term change [9].

2.2.1 Measuring Motivation. There are many ways to measure the motivation of users after interacting with a technical system. One of the methods is the User Motivation Inventory (UMI) [8]. The UMI builds on Organismic Integration Theory, a sub-category of SDT. SDT mentions intrinsic and extrinsic motivation, Organismic Integration Theory still uses these concepts but with the addition of amotivation, the absence or lack of motivation.
Other than these three types of motivation, the UMI also measures *integrated regulation*, *identified regulation*, and *introjected regulation*. With these measures, the UMI assists in specifying people's reasons for engaging with a system.
The UMI is a well-known method that has been used to: investigate the motives of social virtual reality users [45], look into how beliefs and motivations affect willingness to use new systems [44], and making motivational profiles for players of a specific game [7].
There are also ways to determine user motivation and interest in terms of *gameful experiences*. The term gameful experience refers to "the positive emotional and involving qualities of using a gamified application" [13]. One of the scales made for this purpose was created by Högberg et al. and is known as the Gameful Experience Questionnaire or GAMEFULQUEST [19]. The scale was created based on reviews of other well-known scales such as the Game Experience Questionnaire [21] and designed to assess gameful experience among adults using gamified services [51].
GAMEFULQUEST covers seven subscales. These subscales are; challenge, competition, guided, immersion, playfulness, and social experience [19].
The questionnaire has been used in a number of studies since its creation in 2019. GAMEFULQUEST has been used to measure: the effect of gameful experience on motivation when studying via digital escape rooms [51], the positive effect of gameful experience on online gamers' attitudes [30], and how personalised gamification affects motivation [34].

2.2.2 User Types. It is important to note that not everyone is motivated by the same activities and contrivances. Even players who put a lot of time and effort into games have different tastes. This is why there are different genres of games, which attracts different kinds of players. This means that in order to design a motivating gamified system, you need to know which types of users there are, which users you are trying to attract, and what motivates those user types.

The three concepts from SDT; competence, autonomy, and relatedness, are also the motivations for the intrinsic user types on the *Gamification User Type Hexad Scale* [11, 48]. The Scale helps determine which gamification elements to use for motivating users from its six pre-determined user types. The six types are; *Achievers*, *Philanthropists*, *Socializers*, *Free Spirits*, *Players* and *Disruptors*. The four first-mentioned user types are the ones referred to as intrinsic user types. Achievers are motivated by competence, Philanthropists by purpose and meaning, socializers by relatedness, and free spirits by autonomy.
It is improbable that people fit into a singular user type. Typically, users are a mix of two or more user types and are attracted by gamification elements that motivate each of the user types they have higher scores in.

2.3 Gamification

Both dictionaries [12, 29] and researchers like Galetta [14] define gamification as: "the use of game elements, in non-gaming contexts, in order for a task to seem more game-like and increase individual engagement and motivation".
Gamification uses a combination of rewards, either extrinsic or intrinsic, to motivate people to continuously engage with a system [24].

Blohm and Leimeister [3] expand on the idea of adding game mechanics to non-gaming contexts. Gamification is described as using game elements to support a core offer. This is what differentiates gamification from an actual game. They outline how different game mechanics translate into dynamics and which motives each one relates to.
These connections give an overview of which game elements can be used to cover the different psychological needs mentioned in SDT or to attract a specific type of user as outlined by Marczewski's

hexad scale.

Mazarakis and Bräuer [25] investigate the potentials of different game elements through experimenting and prior research into the individual elements' effect on users' motivation in a non-gaming context. They investigate the four elements; badges, feedback, progress bars, and narrative.
Their results showed no significant increase in motivation with a progress bar alone or combined with badges. However, they still highlight the effectiveness of progress bars as visual representations of progress to give users an easy overview of how far they are with a task or activity.
They state that narrative, in the form of a story, is essential for gamification as it has the ability to add meaning to a task and other gamification elements as well as enhance the sense of an emotional experience [16, 22, 31].

2.4 Narrative Gamification

In recent years, narrative gamification has become a big research topic [1, 16, 35, 36, 41]. Conventionally, gamification has relied on points, achievements, and other well-known game mechanics. Narrative gamification borrows from narratology research and adds a story element to the mix.

Prestopnik and Tang [36] study the difference between point-based and narrative gamification concerning player engagement. The difference is studied by evaluating the citizen science games *Happy Match* and *Forgotten Island*.
Happy Match makes the scientific task obvious to users and uses points to reward the players' performances in a classification task. Forgotten Island uses a storyline as its primary focus. The classification task is integrated diegetically, which does not necessarily mean that the scientific task is completely hidden, however, it adds another layer of meaning to the completion of the task. The diegetic integration of the scientific task is presented as a way to reward player advancements, which are typically rewarded with points or badges, by further expanding on the story.
This could add to users' sense of immersion. A higher level of immersion may help to engage people that usually only make a few contributions to citizen science projects. This combats the fact that a small core of people makes the majority of contributions to citizen science projects [35, 36].
They mentions diegesis, story, fantasy, and characters as key factors to keep players immersed in games with a purpose. While these elements were originally identified in a different context, it is stated that they would, likely, be successful in a scientific context as well.

Seo et al. compare narrative gamification to other types of gamification. Their research focuses on gamification implemented in the workplace. They review prior research that implemented performance- or productivity-based gamification. This type of gamification led to the *Hawthorne effect*. The Hawthorne effect refers to the participants changing their behaviour solely to score better instead of it being natural behavioural changes based on enjoyment or intrinsic motivation [43].

Halan et al. combine conventional point-based gamification and narrative gamification. Through the use of leaderboards, narrative, and deadlines, they investigate whether this combination of gamification elements can heighten user participation on a long-term basis. They state that the elements had a positive effect on user participation. They also highlight the importance that leaderboard scores should align with the area of deployment in order to avoid just being meaningless points [18].

O'Donnell attempted to isolate the use of narrative from conventional point-based gamification. He focused on the effect of narrative gamification on motivation, willingness to act, and user experience. His study implemented narrative gamification in three different levels or "doses". After gathering data during a time period of multiple months, O'Donnell concludes that narrative, in medium doses, positively affects users' feelings of meaningfulness [32]. He states that more research is needed to determine the impact of higher doses [32].

When implementing narrative gamification, researchers tend to only use embedded narrative; an explicitly told storyline [26]. Birk et al. study the use of enacted narrative; the development of in-game characters. The idea behind an enacted narrative is that it uses *identification* as a motivator in order to have players feel a sense of connection to the characters and meaningfulness when playing the game [26]. Birk et al. conclude that greater identification leads to more investment in tasks and has a positive effect on motivated behaviour [2].

Narrative gamification and framing have been added to citizen science contexts a couple of times. Pretopnik and Tang [35, 36] evaluate Forgotten Island through an hour of interaction, which was not enough time to complete the full story. The use of narrative gamification by O'Donnell [32] is implemented as different doses of information. He does not include a story or story world but focuses solely on the amount of text users are presented with when reporting an issue in their town. Furthermore, his study does not make any definitive suggestions towards a "correct" dosage of narrative. This means that it is still unknown what motivational impact the length of a narrative has, especially longer ones.

This research highlights the importance of intrinsic motivation in gamified systems as users will change their behaviour for what they think is wanted if only extrinsically motivated by external pressure [43].
It is also suggested that a storyline with elements to evoke a feeling of identification with the in-game characters will have a positive effect on intrinsic user motivation.

2.5 Point-based Gamification in Citizen Science

There are multiple examples of point-based gamification being used as a motivational tool in the scientific field of citizen science. Studies have been done on gamification and its potential to keep user participation high due to its reported positive effect on intrinsic motivation [41, 49, 50]

Bowser et al. [5] investigated the possibility of engaging younger

people, so-called "digital natives" in citizen science projects. They evaluated a gamified citizen science application focused on "floracaches" which was inspired by the world of geocaching. A smaller percentage of the participants were interested in the presented project, they conclude that this may be due to the use of unfamiliar terminology and a general lack of interest in the area. The participants responded positively to the use of badges and social aspects in terms of community involvement. Their participants stated that if the focus was more on the game part of the application they would be more likely to use it. The motivations of these participants were compared with the citizen science volunteers. The comparison showed that: *Fun*, *Personal Interest*, *Community Involvement* and *General Socialisation* were motivators for both groups [5].
Tinati et al. investigated user participation in the gamified citizen science project *Eyewire*. Eyewire incorporates point-based gamification in the form of individual rankings on a leaderboard but also has scheduled events such as marathons and team versus challenges that cover relatedness from self-determination theory and the "socializer" user type from the gamification user type hexad scale.
Their focus was on the why and how of the participants' engagement with Eyewire. They identified four main themes of intrinsic and extrinsic motivations that kept the players engaged with Eyewire. The four identified themes were *contribution and science*, *gaming and entertainment*, *community and learning*, and *personal interest* [47]. These themes, to some degree, line up with the themes identified to motivate both groups presented by Bowser et al. [5].

Iacovides et al. [20] looked at how game elements affect the acquiring of new citizen science volunteers and sustaining engagement over time. The paper compares the online citizen science projects; *Foldit* and Eyewire.
Foldit lets users contribute to science by aiding in folding selected proteins with game-provided tools. The project uses a combination of gamification elements to keep its users motivated. These include points, leaderboards, and chat forums. Foldit has a clear focus on making the project a group task, but it uses gamification elements that cater to both relatedness and competence.
They conclude that game elements are unnecessary when it comes to acquiring new volunteers. However, a positive effect on engagement over time was observed [20].

Bowser et al. [5], much like Prestopnik and Tang [36], suggest that the scientific task should not be the focus when trying to engage and motivate new citizen science volunteers, especially if the target group involves younger people.
Implementing gamification in a citizen science application may not be necessary for attracting volunteers, but it does not seem to have a negative impact and can help in retaining user engagement over time [20]. Having gamification elements that focus on satisfying the need for relatedness is used in several citizen science projects and has been a main motivator for volunteers for a long time [47]. This means that having a social aspect in citizen science projects should be a priority.

2.6 Gamified Image Annotation

Mekler et al. investigated the effect of points, levels, or leaderboards on intrinsic motivation and needs satisfaction but found no significant difference when applying the individual game elements. This observation was stated to be due to a lack of meaningful and informational feedback, which would give users the ability to follow their own progress as well as the quality of their performances.
Even though they did not observe any significant differences between the gamification and control groups, more images were annotated by the group exposed to gamification [28].

Mekler et al. experimented with different combinations of points and meaningful framing. Each variable was tested individually as well as combined together. They tested each condition in terms of the quality of annotations, the number of annotations produced, and intrinsic motivation. They report that the combination of points and meaningful framing attained the best results in all of the categories [27].

Mekler et al. [28] did not test their findings in a complete system that incorporates multiple gamification elements. Their experiments were carried out systematically in a lab setting and were not measured with use over a period of time.
Similar experiments may benefit from being tested through natural interactions over a longer period of time.

Mekler et al. [27] suggest that if a citizen science project was to use points as a reward, adding meaning to the points, that tie into the field of the project, will provide a higher level of intrinsic.
This coincides with how Forgotten Island [35, 36] uses narrative task framing to add a sense of meaning to the completion of the task, which makes it more intrinsically motivating.

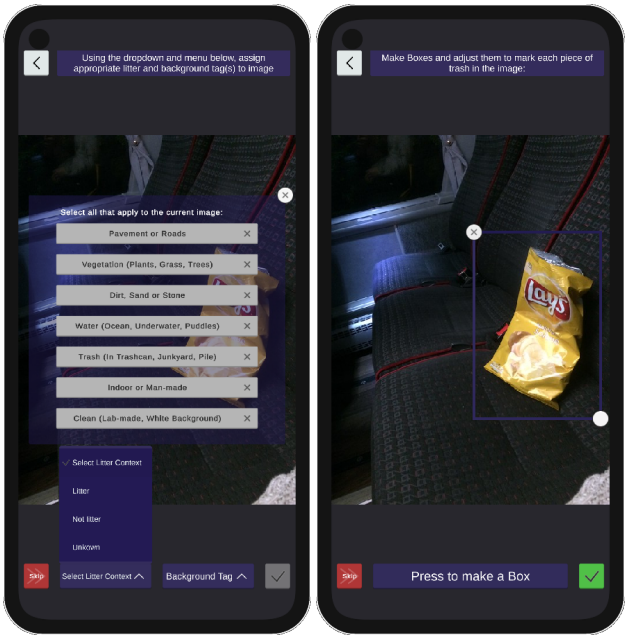
This prior work shows that there, as far as we are aware, has been no research on narrative task framing or narrative-based gamification being applied to a citizen science project that works with image segmentation and annotation.

3 DESIGN

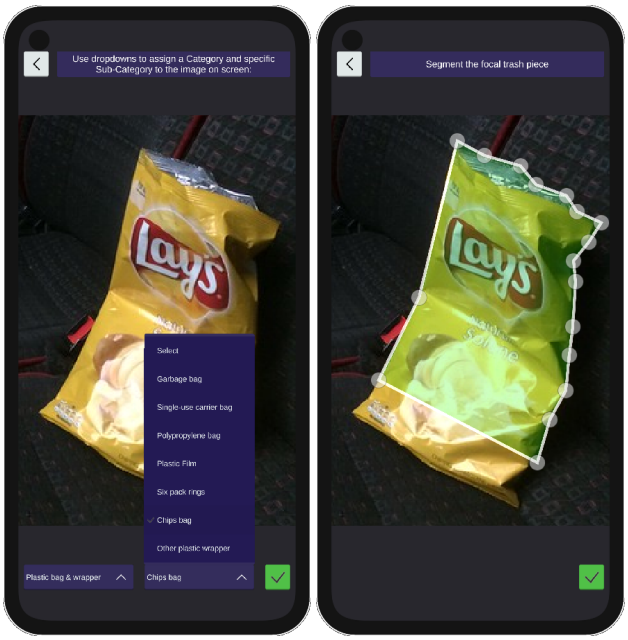
This section outlines the goal of the application, and the narrative task framing used for the application and relates it to the prior research mentioned in the related work section.

3.1 Goal of the Design

Our goal was to evaluate whether narrative task framing, surrounding or framing a task with a narrative, affects user motivation in citizen science projects.
To determine this, we created a system that uses narrative task framing and a system that solely relies on conventional point-based gamification in a citizen science task, image segmentation, and annotation in our case. The systems would be used to compare if the way a task is framed is a factor in participants' motivation and decisions to continue their use.
Besides the narrative and gamification elements, we focused on the aspect of meaningful framing as part of the task framing. Meaningful framing refers to either visuals, text, or a combination of the two, that gives a sense of meaning to completing a task. Meaningful



(a) Context Annotation (b) Border Box



(c) Category Annotation (d) Segmentation

Figure 1: The four tasks that users can unlock and complete within the application (a) Task 1: *CoA*, with the context dropdown and menu with selectable background tags. (b) Task 2: *BB*, with a box placed around the object of the image. (c) Task 3: *CA*, where users choose a super category and category to represent the litter in the image. (d) Task 4: *Seg*, where the users, through instance segmentation, outlines the litter in the image.

Task	Avg. SUS Score	SD
Context Annotation	90.50	5.42
Border Box	89.00	8.22
Category Annotation	75.50	21.46
Segmentation	86.50	16.55

Table 1: A table showing the average SUS scores and standard deviation for each of the application's four tasks.

framing makes sure that users can relate the tasks they are completing with solving real-world problems which gives a sense of importance to completing the tasks and leaves users with a sense of completion [27, 46], both through personal performance and contributing to science and the public good.

3.2 Tasks

To judge the effect of narrative task-framing, our application follows the process of completing image segmentations and annotations as a citizen science volunteer.
For the application, we needed a large number of images for our participants to annotate and segment. The images used were taken from the TACO dataset [37]. TACO is a dataset focusing on segmentations and annotations of trash in context.

To make sure our citizen science application produces proper data, we used the steps and categories from the workflow that was used to generate the TACO dataset [37]. Our annotation task is split into four sub-tasks. The first step is to annotate the context and one or more background tag(s) to the image, this task is referred to as *Context Annotation* (CoA), see Figure 1a. The second step requires the user to add a box around the border of each piece of litter in the image to make sub-images of each of the pieces, this task is called *Border Box* (BB), see Figure 1b. For the third step, the user is given a sub-image to categorise using TACO's super categories and categories, this task is called *Category Annotation* (CA), see Figure 1c. The final step is to segment a classified sub-image via instance segmentation, this task is referred to as *Segmentation* (Seg), see Figure 1d.
When users initially open the application they only have access to the first task, CoA. Completing a task is rewarded with points. To access the next task, the users are required to level up. The user levels up at 25, 50, 100, and 150 points. The final level at 150 points does not give access to a task but rather thanks the user for completing this many tasks and lets them know that no more tasks or features will be unlocked. The user is also made aware of the fact that their continued use will still be appreciated and contributes to the dataset. This is to add meaningful framing to using the application even after unlocking all the different tasks. Meaningful framing with the combination of points has a positive effect on users' sense of purpose and intrinsic motivation [27].

3.2.1 Task Usability Testing. to finalise the functionality and design of the four tasks, we carried out a usability test with five, male, participants. The test had the participants interact with each of the four sub-tasks, CoA, BB, CA, and Seg, for 2 minutes. After interacting with each sub-task, a System Usability Scale (SUS) [42] was

(a) JunkCorp Home Screen **(b) Gamified-Only Home Screen**

Figure 2: Both of the home screens with the tasks and goals presented to the users. (a) The home screen of the JCV is made to resemble an office space. (b) The home screen of the GOV application is flatter and minimalistic.

filled out. After completing all the tasks, participants were asked to rank the tasks based on difficulty.

Our average SUS score was 85.38 (SD = 14.54). We also calculated the average SUS score for each task as seen in table 1.
The scores show that the only task that, on average, was ranked below the threshold for excellent (80.3) [42] was the third task (CA). The high standard deviation indicates that users had very different experiences with this step. SUS scores for the task varied from 40 to 92.5. Participants were asked to focus on the task itself regardless of the quality of images or knowledge needed. However, the scores were still affected by the images presented and how easy or difficult it was to identify the correct categories from the images. As the tasks were completed the same way as when interacting with the finished application, the images were different from participant to participant as they were randomly chosen from the dataset. This may have resulted in varying experiences with each task for different participants. This was further confirmed through the oral feedback as well as the difficulty rankings.
Two participants ranked the tasks in order (CoA, BB, CA and Seg). Two participants ranked the tasks as BB, Seg, Coa and CA. These participants reported ranking CA as the most difficult due to the number of categories as well as the low quality of the images making it difficult to determine the type of litter in the image. The last participants ranked the difficulty of the tasks as Seg, BB, CoA and CA. The reasons for ranking CA as the most difficult task were similar as mentioned by the other two participants.

3.3 Narrative Task Framing

Our narrative task framing focuses on diegetically integrating the scientific task into a story and adding a sense of meaning without fully disguising the scientific task. This transparency ensures that the meaningfulness of citizen science is still present. This way of integrating the narrative elements is similar to the way it was done with the citizen science game Forgotten Island [35, 46]. Forgotten Island was compared to Happy Match [46], while the task of classification was the same it was two completely different games with different controls. This means that no direct comparison between Forgotten Island and a non-narrative version of it was ever made.

The base of the narrative was chosen through an evaluation with 17 participants (12 males and 5 females). The participants were presented with three different story scenarios in the form of storyboards. First, the participants read through all three storyboards and then we moved on to their feedback for each one. They gave feedback based on their initial reaction to the storyboard and were asked open-ended questions to gain an understanding of what was liked and disliked about each narrative. After the feedback, the participants were asked to rank the storyboards based on which one they liked more.
The overall, highest-ranking storyboard was used as the foundation and the feedback from this evaluation was used to change and shape the narrative into the final and implemented narrative.

The narrative version, known as JunkCorp Version (JCV), of the application, uses both story and appropriate visuals to further immerse users in the narrative surrounding the citizen science task.

The visuals represent a workstation inside the fictional corporation, JunkCorp, where the player's in-game character is working. The workstation includes Post-it notes, old '80s-inspired screens, and a keyboard. These visuals can be seen on the home screen of the JCV in Figure 2a.
The story takes place at JunkCorp: A giant trash-gathering and recycling corporation, where the user has recently been hired. The user has face-to-face discussions with their boss. Subtitles of these discussions are shown on one of the computer screens. Each character is associated with a different-coloured text and name to make it easier to distinguish who is currently talking. An example of a face-to-face discussion and the colour-coordinated text and names can be seen in Figure 3.
After being introduced to their job, the playable character quickly realises that it is very repetitive and could easily be automated. With this in mind, the character installs an AI to make their job easier. However, the AI is untrained for the task and unable to do a satisfactory job. It generates tasks for the user, the annotation task, which will train it to do better. After several rounds of reprimands from their boss, the user uses the feedback to properly train the AI with the goal of their job being fully automated.
Outside of the in-app narrative framing, daily notifications were sent to the users. The notifications functioned as part of the narrative as they would remind users that their daily quotas have been updated and are ready to track their progress again. This meant

(a) Talking to the AI (b) Talking to Jeremiah

Figure 3: Instances of dialogue captured from the JCV application. (a) A conversation with the AI on how it can improve itself with the user's help. (b) Jeremiah, the user's manager at JunkCorp, complaining about poor performance.

that the notifications fit into the office setting. This made the notifications diegetic which is an important part of keeping users motivated and interested in the story [36].

3.4 Point-based Gamification Elements

The point-based gamified version of the application, known as Gamified Only Version (GOV), focuses solely on the tasks with no story elements added to it. A more modern and simple design was used so as to not distract from the tasks. The design consists, mostly, of variations of blue with no major details.

The concept of daily tasks is implemented to encourage the users to complete certain tasks in exchange for additional points that may up their ranks on the leaderboard. The daily tasks may also serve as a minimal to-do list for users to feel that they have sufficiently contributed to the underlying scientific cause or be seen as a challenge to be completed. This means that daily tasks can function as an intrinsic motivator for people motivated by purpose and challenge and an extrinsic motivator for people motivated by obtaining points and climbing the leaderboard [11]. The progress of each of the daily tasks was visually tracked via progress bars due to their ability to give a quick visual representation of how far along the user is in a task [25]. The home screen showing the daily tasks of GOV is displayed in Figure 2b.

Like JCV, GOV also sent out notifications to the users. While the notifications of JCV added to the story, the notifications sent out for GOV only served as reminders for users to interact with the application by stating that the daily tasks have been reset.

We conducted a test focusing on gamification elements. Five participants (3 males and 2 females) were given a comprehensive explanation of the goal of the application and were asked how they would like different elements to be displayed or if they would want them to be there at all. For each gamification element, visual examples were given of different ways they could be implemented to give the participants an idea of what could be done. The design test focused on leaderboards and community goal progress. The verbal feedback from the participants was analysed via thematic analysis [6]. Thematic analysis is used to find themes within qualitative data by highlighting similarities between participants' responses. The findings from the test led to decisions on how the elements ended up being displayed within the application.

The leaderboard functions as both a competitive tool to compare yourself to other users and a visual of your own progress in cases where competition is not a motivator [11, 18]. The leaderboard uses a username, picked by the user at their first interaction with the application, and a rank to showcase the user's placement on the leaderboard.
The points of the leaderboard are assigned after a task has been completed. Each task or step is assigned a point value based on its difficulty. Tasks 1 and 2 are worth 1 point and tasks 3 and 4 are worth 2 points.
The leaderboard is absolute and scrollable, meaning all players are visible on it and can be navigated by scrolling up and down. The two different versions of the leaderboard can be seen in Figure 4.
We incorporated a weekly community event which lets all users work together to reach a common goal. This serves as a motivator for purpose and relatedness [11, 20, 38]. For the display of progress during a community event, it was decided that both the overall group progress as well as the users' own contribution should be showcased. This is done through a standard progress bar that tracks the overall progress and a number on the side of the bar that tracks the user's individual contributions. The tracking of the community goals can be seen in Figure 2b.

4 EVALUATION

This section describes our evaluation procedure, what results we gathered, and a discussion of what the results mean for our investigation of narrative task framing as a factor for user motivation.

4.1 Participants

We conducted our evaluation with 18 participants (14 male and 4 female) with an average age of 24.28.
The participants were assigned to either the Gamified Only Version (GOV) or the JunkCorp Version (JCV). When the application was downloaded, the server assigned the participant to the group that currently had the least amount users. This was done to ensure an equal number of participants in each group while still keeping it random.

(a) Gamified-Only Leaderboard **(b) JunkCorp Leaderboard**

Figure 4: The leaderboards as they are visualised within the two different versions of the application. (a) The GOV application uses more direct labels while (b) the JCV uses terms that root it in the narrative.

4.2 Procedure of Evaluation

Our evaluation followed the procedure:

(1) Obtain informed consent from participants.
(2) Have participants fill out gamification user type hexad scale.
(3) Let participants interact with the application for 6 days.
(4) Observe the data as it is added to the database.
(5) Set up times for post-test interviews.
(6) Let participants fill out the UMI and GAMEFULQUEST.
(7) Conduct a semi-structured interview and obtain feedback.

We needed informed consent to store the names of our participants to associate with their chosen usernames, which were used to track their individual activity through our database. Consent was also needed to obtain a way to contact the participants, e.g. an e-mail address.

The gamification user type hexad scale was used to gain an understanding of the preferences of the participants to be used when discussing each participant's reaction and experience with the application. The scale was filled out before interacting with the application in order to avoid bias of participants answering solely based on what they liked or disliked about our application.

The UMI and GAMEFULQUEST were chosen due to their ability to produce quantifiable data about the motivation and gameful experience of each participant. This quantitative data served as a basis for comparing with qualitative data from the semi-structured interviews and feedback given at the end of the evaluation.

4.3 Results

4.3.1 Gamification User Type Hexad Scale. The results from the hexad scale showed that all but one participant had one of the four intrinsic user types as their main type. We assigned the main types based on the user type the participant had the highest score in. Eighth of the participants were philanthropists, four were free spirits, three were socializers, one was a mix of philanthropist and achiever, one was a mix of philanthropist and socializer, and the final participant was player.

While these were the assigned main user types, each participant had varied scores throughout each user type. This is also seen in the closeness of the average score for each user type, which is shown in Table 2. The individual participants' scores were, in some cases, very similar for multiple user types. This means that some participants may be motivated by gamification aimed towards three or four different user types. An example of this is Participant 13, who had scores of 23, 23, 23, 30, 30, and 31. This meant that the participant was assigned the user type with the score of 31 as their main user type, but they would still be motivated by game elements suggested for the user types where they scored 30.

User Type	Avg. Score	SD
Socializer	25.72	3.49
Achiever	26.61	3.22
Philanthropist	29.06	2.88
Free Spirit	26.44	4.05
Player	23.67	3.51
Disruptor	19.28	5.28

Table 2: The average scores for each of the different user types in the Gamification User Type Hexad Scale.

4.3.2 User Motivation Inventory. The UMI was filled out by all participants to get an overview of any differences in motivation between GOV and JCV.

The scores from GOV, see Table 3, showed amotivation as the highest scoring motivation (3.89). Intrinsic was the second highest with external and introjected regulation following fairly close behind.

For JCV, see Table 4, Intrinsic motivation was the highest scoring (4.30) with identified and integrated regulation being the ones that followed after.

Motivation	Avg. UMI Score	SD
Amotivation	3.89	0.96
External	3.22	1.53
Introjected	3.19	1.75
Identified	2.96	1.39
Integrated	2.70	1.69
Intrinsic	3.56	1.00

Table 3: A table showing the average UMI scores and standard deviation for each of the motivations measured through the inventory for the participants from GOV.

Motivation	Avg. UMI Score	SD
Amotivation	3.33	1.57
External	2.52	1.32
Introjected	2.41	0.94
Identified	3.93	1.58
Integrated	3.41	1.82
Intrinsic	4.30	2.40

Table 4: A table showing the average UMI scores and standard deviation for each of the motivations measured through the inventory for the participants from JCV.

4.3.3 GAMEFULQUEST. GAMEFULQUEST uses its 56 items to gather information about the participants' gameful experiences over seven factors: *Accomplishment*,*Challenge*, *Competition*, *Guided*, *Immersion*, *Playfulness*, and *Social Experience*.
For GOV, competition scored highest (4.24) with accomplishment (3.67) and guided (3.49) following behind it. See Table 5 for all scores and standard deviation
With JCV, accomplishment scored highest (4.85). Immersion (3.91) and competition (3.87) were the next ones with the other factors not being far behind. All the scores and standard deviations can be seen in Table 6.

Factor	Avg. Score	SD
Accomplishment	3.67	1.52
Challenge	2.81	1.08
Competition	4.24	1.49
Guided	3.49	1.28
Immersion	2.73	0.98
Playfullness	2.63	1.12
Social Experience	3.26	1.07

Table 5: A table showing the average scores and standard deviation for each of the factors measured through GAMEFULQUEST for GOV.

Factor	Avg. Score	SD
Accomplishment	4.85	1.62
Challenge	3.33	1.69
Competition	3.87	1.65
Guided	3.51	1.74
Immersion	3.91	1.61
Playfullness	3.67	1.78
Social Experience	3.78	1.86

Table 6: A table showing the average scores and standard deviation for each of the factors measured through GAMEFULQUEST for JCV.

4.3.4 Qualitative Data. The questions for the post-test interviews were the same for participants of JCV and GOV. The only difference was the addition of a question regarding thoughts about the story and asking for a summary of the story from the participants of JCV.

Thematic analysis [6] was used to find themes and similarities in the answers given by the participants throughout the interviews. When asked about their motivations for interacting with the application participants of GOV reported a combination of helping with the project and the leaderboards being the main motivators. For JCV, the participants were more varied in their answers, mentioning the story, leaderboard, daily tasks, curiosity, and the application being fun as reasons for their use.

We asked the participants to rank the tasks they had interacted with based on difficulty. A vast majority of all participants agreed that task 3 (CA) was the most difficult task due to the overwhelming amount of categories.
A singular participant of JCV only gained access to the first task due to a busy schedule resulting in them interacting with the application, for the first time, on the second to last day. Three participants from GOV only had access to the first task. This was due to them losing interest in continuously completing the first task.

4.4 Result Analysis

The data from GAMEFULQUEST was normally distributed and had homogeneity of variance, which means the data were parametric. The subscales of introjected, integrated and intrinsic regulation were deemed non-parametric. This was concluded using Shapiro-Wilk and Levene's tests. No statistically significant differences were found between the two groups in both the UMI and GAMEFULQUEST subscales. This was observed via the use of Welch's t-test, for parametric data, and Mann-Whitney U-test, also known as the Wilcoxon rank sum test, for the non-parametric data.

In the UMI scores the GOV group scored higher than JCV in three subscales: amotivation, external regulation and introjected regulation. This would indicate that users from this group were, on average, less motivated to interact with the application and felt more pressure from external factors like being part of a project. The external factors can also be a part of introjected regulation as it may refer to internal pressure such as ego enhancement and pride or to avoid punishment e.g., guilt or shame.
The higher introjected regulation score from GOV was likely influenced both by external pressure, as six participants mentioned helping us with the project as part of their motivation, as well as by internal pressure. The internal pressure of ego enhancement and pride are also seen in the fact that competition (4.24) was the highest scoring subscale of the GAMEFULQUEST.

The JCV group scored higher in intrinsic, integrated, and identified regulation. Much like intrinsic regulation, integrated refers to when behaviour is integrated into the user's personal values and beliefs. Identified regulation is when behaviour is explicitly recognised and valued. While JCV scored higher than GOV in these subscales, only intrinsic regulation is considered as a higher score (4.30). This would mean that the main motivation of the participants was intrinsic motivation, such as the fact that they found the game fun to play or because they felt like they contributed to something

bigger than themselves.
The higher score in intrinsic regulation is presumably most influenced by the latter of the two. While some participants mentioned the application being fun, six out of the nine participants had philanthropist as, one of, their main user types. This was further supported by the participants' interview answers as they mentioned helping us and contributing to science as motivations for interacting with the application. Due to the philanthropic nature of the participants, helping us with the project comes across as an intrinsic motivator due to the motivation of purpose rather than an extrinsic motivation in the form of external pressure.

The interviews and questionnaires indicated that there were people in the GOV group that were not motivated by the leaderboard or even got demotivated seeing people far ahead in points. Similarly, there were people from the JCV group that indicated that a story is not motivating to them as they were not fans of reading in general. This shows that experiences with different gamification elements and narratives are subjective, and it would take more research and studies with more participants to narrow down towards a recipe close to making a universally enjoyable gamified system.

5 DISCUSSION

This section describes areas of the project that we have reflected on after completing the evaluation and goes over limitations, changes, and future work that could improve the application as well as serve as ideas for a framework of a similar application.

5.1 Application Design

We used progress bars to visually represent the progress towards completing daily tasks but did not use them to show progress towards a new level.
The participants from GOV who only unlocked the first task mentioned being uncertain about whether there was more to unlock or when it would be unlocked to be the main reasons for them not progressing further. The uncertainty of when something new would be unlocked could potentially be avoided if a progress bar showed how far into a level they were. This would show the users a visual representation and they might be motivated to continue if they see they are not far from a new level.

5.2 Task Design

While our tasks were modelled after the task from the TACO dataset, we made a small difference with the first task, Context Annotation, which could affect the understanding of the task.
When annotating via TACO, users are first asked to segment an object of the image and then classify the context of that object. Afterwards, they are allowed to segment other objects in the same image.
In our applications, the user annotates the context of the image before specifying the area with the border box or segmentation task. This could have caused some confusion when encountering an image with multiple objects, if a user determined that the objects belong to different litter contexts. This could be avoided by switching the orders of the tasks or incorporating the context annotation into one of the other tasks.

Our participants agreed that the third task (Category Annotation) was the most difficult task. They stated that this ranking was due to the many categories being overwhelming and making the task very time-consuming.
In TACO's annotation tool, the super-categories and categories are presented in the same dropdown menu. doing this would avoid the frustration of choosing a super category only to discover that none of the categories matched the litter, as one participant mentioned. We theorise that utilising the national trash-sorting categories of Denmark as the super categories and then showing the relevant super category/category pairs would further simplify the task, as it would add some sense of familiarity to the task and make it less overwhelming with the number of displayed categories.

5.3 Technical Difficulties

Through the post-test interviews, we found knowledge about any trouble that the users encountered or that we had suspicions about and wanted to confirm.
One of these issues was regarding the notifications that were supposed to be sent from the application. In the interview, participants were asked if they received any notification and if yes, how often. Most participants reported never receiving any notifications, two participants reported getting a singular notification after downloading the app, but then never again, one participant got notifications every other day, and one participant received daily notifications as intended.
This happened due to an internal storage problem, resulting in the notification manager assuming a notification was already scheduled, perpetually. This meant a new notification would never be scheduled.

5.4 Evaluation Reflections

During the evaluation, we were monitoring the data as it came in. We realised on the second day that we had no way of knowing how many of each task was completed, daily, by GOV and JCV, respectively. Because of this, we started noting down the number of each task that was completed by the different groups, at the same time each day, the remaining days. Due to noticing this too late, this data was not recorded for the first day.

As it was mentioned in the design, users were made aware that they would not unlock new content after reaching 150 points. This could potentially result in users completely stopping their use of the application. However, the leaderboard showed that, for the majority of our users, this was not the case.
One participant reached 151 points, which means they stopped shortly after receiving the message. All other participants that reached 150 points went on for at least 10 points, with the highest-ranked user ending with 663 points.
For future iterations or similar future work, it would be worthwhile to consider whether to let users know that no more new content is available as directly as we did or at all.

6 CONCLUSION

The goal of this study was to explore the potential of narrative task framing as a factor for user motivation in citizen science image segmentations and annotation tasks. The 18 participants' use of our applications was monitored to determine how many tasks were completed by each group. These observations were paired with questionnaires and interviews to gain insight into what motivated people to use the application. The application was developed using Unity, built for Android, and distributed directly to our participants.

Early usability testing was carried out using the System Usability Scale (SUS). This gave an indication of the overall usability of the tasks incorporated into the applications. The main study incorporated the User Motivation Inventory (UMI) and the Gameful Experience Questionnaire (GAMEFULQUEST). The UMI and GAMEFULQUEST were used to quantify the users' motivation and motives to segment and annotate images along with the narrative framing versus with no narrative elements.

The results showed that the users' use of the JunkCorp Version had higher scores in the UMI subscales intrinsic, identified, and integrated regulation. However, our analysis showed no statistically significant difference in the relationship between user motivation and which version of the application was used.

Our results still suggest that narrative framing could have the potential to motivate users to volunteer in citizen science projects.

To fully address the motivational potential of narrative task framing, further studies with more participants and various implementations of narrative task framing would have to be conducted.

REFERENCES

[1] Michael B Armstrong and Richard N Landers. 2017. An evaluation of gamified training: Using narrative to improve reactions and learning. *Simulation & Gaming* 48, 4 (2017), 513–538.

[2] Max V Birk, Cheralyn Atkins, Jason T Bowey, and Regan L Mandryk. 2016. Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2982–2995.

[3] Ivo Blohm and Jan Marco Leimeister. 2013. Gamification. *Business & information systems engineering* 5, 4 (2013), 275–278.

[4] Rick Bonney, Jennifer L. Shirk, Tina B. Phillips, Andrea Wiggins, Heidi L. Ballard, Abraham J. Miller-Rushing, and Julia K. Parrish. 2014. Next Steps for Citizen Science. *Science* 343, 6178 (2014), 1436–1437. https://doi.org/10.1126/science.1251554 arXiv:https://www.science.org/doi/pdf/10.1126/science.1251554

[5] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. 2013. Using Gamification to Inspire New Citizen Science Volunteers. In *Gamification '13* (Toronto, Ontario, Canada). Association for Computing Machinery, New York, NY, USA, 18–25. https://doi.org/10.1145/2583008.2583011

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa

[7] Florian Brühlmann, Philipp Baumgartner, Günter Wallner, Simone Kriglstein, and Elisa D Mekler. 2020. Motivational profiling of league of legends players. *Frontiers in Psychology* 11 (2020), 1307.

[8] Florian Brühlmann, Beat Vollenwyder, Klaus Opwis, and Elisa D. Mekler. 2018. Measuring the "Why" of Interaction: Development and Validation of the User Motivation Inventory (UMI). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173680

[9] Elaine Clanton Harpine. 2015. *Is Intrinsic Motivation Better Than Extrinsic Motivation?* Springer International Publishing, Cham, 87–107. https://doi.org/10.1007/978-3-319-19102-7_6

[10] Alex de Sherbinin, Anne Bowser, Tyng-Ruey Chuang, Caren Cooper, Finn Danielsen, Rorie Edmunds, Peter Elias, Elaine Faustman, Carolynne Hultquist, Rosy Mondardini, et al. 2021. The critical importance of citizen science data. *Frontiers in Climate* (2021), 20.

[11] Lisa Diamond, Gustavo Tondello, Andrzej Marczewski, Lennart Nacke, and Manfred Tscheligi. 2015. The HEXAD Gamification User Types Questionnaire: Background and Development Process. In *Workshop on Personalization in Serious and Persuasive Games and Gamified Interactions*. London, UK, 6.

[12] Cambridge Dictionary. [n. d.]. Gamification. https://dictionary.cambridge.org/dictionary/english/gamification

[13] René Eppmann, Magdalena Bekk, and Kristina Klein. 2018. Gameful Experience in Gamification: Construction and Validation of a Gameful Experience Scale [GAMEX]. *Journal of Interactive Marketing* 43 (2018), 98–115. https://doi.org/10.1016/j.intmar.2018.03.002

[14] GIUSEPPE Galetta. 2013. The gamification: Applications and developments for creativity and education. In *Creative Personality. Collection of Scientific Papers*, Vol. 11. Creativity and Innovation in Education Conference, Riga, Latvia, 1–10.

[15] Francisco Sanz García, Maite Pelacho, Tim Woods, Dilek Fraisl, Linda See, Mordechai Muki Haklay, and Rosa Arias. 2021. Finding what you need: A guide to citizen science guidelines. *The science of citizen science* (2021), 419.

[16] Jaroslaw Grobelny, Joanna Smierzchalska, and Krzysztof Czapkowski. 2018. Narrative gamification as a method of increasing sales performance: a field experimental study. (2018).

[17] Mordechai Muki Haklay, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, and Katrin Vohland. 2021. *What is citizen science? The challenges of definition.* Springer Nature, : Gewerbestrasse 11, 6330 Cham, Switzerland. 13 pages.

[18] Shivashankar Halan, Brent Rossen, Juan Cendan, and Benjamin Lok. 2010. High Score! - Motivation Strategies for User Participation in Virtual Human Development. In *Intelligent Virtual Agents*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 482–488.

[19] Johan Högberg, Juho Hamari, and Erik Wästlund. 2019. Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction* 29, 3 (2019), 619–660.

[20] Ioanna Iacovides, Charlene Jennett, Cassandra Cornish-Trestrail, and Anna L. Cox. 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* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 1101–1106. https://doi.org/10.1145/2468356.2468553

[21] W.A. IJsselsteijn, Y.A.W. de Kort, and K. Poels. 2013. *The Game Experience Questionnaire.* Technische Universiteit Eindhoven.

[22] Karl M Kapp. 2012. *The gamification of learning and instruction: game-based methods and strategies for training and education.* John Wiley & Sons.

[23] Sage Lazzaro. 2021. Machine Learning's rise, applications, and challenges. https://venturebeat.com/2021/06/21/machine-learnings-rise-applications-and-challenges/

[24] Andrzej Marczewski. 2019. Hexad: A player type framework for Gamification Design. https://www.gamified.uk/user-types/

[25] Athanasios Mazarakis and Paula Bräuer. 2023. Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human–Computer Interaction* 39, 3 (2023), 612–627. https://doi.org/10.1080/10447318.2022.2041909 arXiv:https://doi.org/10.1080/10447318.2022.2041909

[26] Elle McFadzean. 2019. 4 types of narrative in games. https://ellemcfadzean.com/4-types-of-narrative-in-games/

[27] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 1137–1142. https://doi.org/10.1145/2468356.2468559

[28] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Toronto, Ontario, Canada) *(Gamification '13)*. Association for Computing Machinery, New York, NY, USA, 66–73. https://doi.org/10.1145/2583008.2583017

[29] Merriam-Webster. [n. d.]. Gamification definition & meaning. https://www.merriam-webster.com/dictionary/gamification

[30] Sita Mishra and Gunjan Malhotra. 2021. The gamification of in-game advertising: Examining the role of psychological ownership and advertisement intrusiveness. *International Journal of Information Management* 61 (2021), 102245.

[31] Fiona Fui-Hoon Nah, Venkata Rajasekhar Telaprolu, Shashank Rallapalli, and Pavani Rallapalli Venkata. 2013. Gamification of education using computer games. In *Human Interface and the Management of Information. Information and Interaction for Learning, Culture, Collaboration and Business, 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part III 15*. Springer, 99–107.

[32] Nicholas O'Donnell. 2021. *Narrative gamification of online civic interactions.* Ph. D. Dissertation. Queensland University of Technology.

[33] OECD. 2022. *OECD Guidelines for Citizen Participation Processes.* 94 pages. https://doi.org/https://doi.org/10.1787/f765caf6-en

[34] Wilk Oliveira, Juho Hamari, Sivaldo Joaquim, Armando M Toda, Paula T Palomino, Julita Vassileva, and Seiji Isotani. 2022. The effects of personalized gamification on students' flow experience, motivation, and enjoyment. *Smart Learning Environments* 9, 1 (2022), 16.

[35] Nathan Prestopnik and Dania Souid. 2013. Forgotten Island: A Story-Driven Citizen Science Adventure. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 2643–2646. https://doi.org/10.1145/2468356.2479484

[36] Nathan Prestopnik and Jian Tang. 2015. Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior* 52 (2015), 492–506.

[37] Pedro F. Proença and Pedro Simões. 2020. TACO: Trash Annotations in Context for Litter Detection. *arXiv preprint arXiv:2003.06975* (2020), 6.

[38] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic motivation inventory (IMI). http://selfdeterminationtheory.org/intrinsic-motivation-inventory/

[39] Richard M. Ryan and Edward L. Deci. 2020. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology* 61 (2020), 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

[40] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion* 30, 4 (2006), 344–360.

[41] 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. https://doi.org/10.1016/j.chb.2016.12.033

[42] Jeff Sauro. 2011. Measuring usability with the system usability scale (SUS). https://measuringu.com/sus/

[43] Kyoungwon Seo, Sidney Fels, Mujin Kang, Changho Jung, and Hokyoung Ryu. 2021. Goldilocks conditions for workplace gamification: How narrative persuasion helps manufacturing workers create self-directed behaviors. *Human–Computer Interaction* 36, 5-6 (2021), 473–510.

[44] Dina Stiegemeier, Johannes Kraus, Sabrina Bringeland, and Martin Baumann. 2022. Motivated to use: Beliefs and motivation influencing the acceptance and use of assistance and navigation systems. *International Journal of Human–Computer Interaction* (2022), 1–16.

[45] Philipp Sykownik, Linda Graf, Christoph Zils, and Maic Masuch. 2021. The most social platform ever? A survey about activities & motives of social VR users. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 546–554.

[46] Jian Tang and Nathan Prestopnik. 2019. Exploring the impact of game framing and task framing on user participation in citizen science projects. *Aslib Journal of Information Management* (2019).

[47] Ramine Tinati, Markus Luczak-Roesch, Elena Simperl, and Wendy Hall. 2017. An Investigation of Player Motivations in Eyewire, a Gamified Citizen Science Project. *Comput. Hum. Behav.* 73, C (aug 2017), 527–540. https://doi.org/10.1016/j.chb.2016.12.074

[48] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 229–243. https://doi.org/10.1145/2967934.2968082

[49] Horst Treiblmaier and Lisa-Maria Putz. 2020. Gamification as a moderator for the impact of intrinsic motivation: Findings from a multigroup field experiment. *Learning and Motivation* 71 (2020), 101655. https://doi.org/10.1016/j.lmot.2020.101655

[50] Áron Tóth and Sarolta Tóvölgyi. 2016. The introduction of gamification: A review paper about the applied gamification in the smartphone applications. 000213–000218. https://doi.org/10.1109/CogInfoCom.2016.7804551

[51] Hava E Vidergor. 2021. Effects of digital escape room on gameful experience, collaboration, and motivation of elementary school students. *Computers & Education* 166 (2021), 104156.

[52] Ashley Wichman. 2015. Challenges, crowdsourcing, citizen science: What's the DIF? https://digital.gov/2015/12/16/challenges-crowdsourcing-citizen-science-whats-the-dif/

[53] Nannan Xi and Juho Hamari. 2019. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management* 46 (2019), 210–221.

[54] Galaxy Zoo. [n. d.]. Zooniverse. https://www.zooniverse.org/