

Swipe for Science: a single-player game designed for collecting discriminative evidence

Jeffrey Lim , Agathe Balayn , Ujwal Gadiraju , Jie Yang

TU Delft

Abstract

We introduce Swipe for Science, a single-player mobile game designed for collecting discriminative evidence from crowds. When people play this game, we collect data about how certain concepts are associated with different contexts, which is valuable knowledge for machine learning systems. The main gameplay loop consists of swiping concepts up, down, left, or right based on how these is associated with the given contexts. The game is simplistic and intuitive to control, and collects data of overall good quality, but it is not engaging enough to retain its players. Further development and research is required to unlock the full potential of this game.

1 Introduction

Machine learning systems are becoming more important than ever. We rely a lot on systems like facial recognition, image search, and scene recognition, but systems like these are only reliable if these have been trained with large and varied datasets (Gong et al., 2019) consisting of both explicit and tacit knowledge. According to Alexander (2018), explicit knowledge is “knowledge that is easy to articulate, write down, and share” and tacit knowledge is “knowledge gained from personal experience that is more difficult to express”. Because it is difficult to express, it can be difficult to obtain as well.

In order to elicit this knowledge, human contributors can perform certain computation tasks. Usually these tasks are not engaging to the participants, so it is difficult to convince them to perform a lot of those. To solve this problem, the tasks can be gamified so that it becomes a form of entertainment instead of a chore. This way the participants can enjoy playing a game and we collect valuable data at the same time. This type of game is called a “game with a purpose” (GWAP), a term first coined by Von Ahn and Dabbish (2008).

Several well performing GWAPs have been made in the past to collect tacit knowledge. Examples of these are: the ESP game that was used for labeling images (Von Ahn

and Dabbish, 2004), Peekaboom that was used for locating objects in images (Von Ahn, Liu, et al., 2006), Phetch that was used for annotating images with descriptive texts (Von Ahn, Ginosar, et al., 2007), and Verbosity that was used for collecting commonsense facts (Von Ahn, Kedia, et al., 2006).

No GWAP, however, has ever been designed to collect discriminative evidence. The discriminative evidence that we are interested in is how certain concepts are associated with different contexts. In the scope of this project, contexts are rooms in a house and concepts are objects that can belong in those rooms. This can be useful for, for example, identifying rooms in a house based on the objects that are in it. Knowing that a certain room has a stove, it is obvious that the room is a kitchen and not a living room. Or knowing that a room has cutlery, it can only be either a dining room or a kitchen.

1.1 Research question

The research question of this paper is: “How can we elicit discriminative evidence using a single-player game?” We answer that question by answering the following subquestions:

1. How has tacit knowledge been collected before using GWAPs?
2. How can we design a single-player game workflow for collecting large quantities of discriminative evidence?
3. How can we support engagement of the players of the game?
4. How can we evaluate the performance of the game?

We answer the first subquestion by exploring literature about other GWAPs. To answer the second and third subquestions, we create a brand new game ourselves that is designed to both collect large quantities of data and engage players at the same time. The last subquestion is answered by executing an experiment where people play the game for a while and evaluating if the game actually performs as well as intended.

This research paper describes the research process that resulted in Swipe for Science, the GWAP that collects discriminative evidence. Section 2 covers how other GWAPs were designed and what Swipe for Science aims to do differently.

In section 3, we describe the design of the game in depth, how the questions are generated, how the game engages players, and how the collected data is processed to ensure the end results are reliable and usable. Section 4 describes the limited experiment we did to try to evaluate the performance of the game and section 5 reflects on the reproducibility of this research and other ethical concerns about the game. Section 6 suggests what can be done in the future to improve the game and section 7 concludes the paper.

2 Related Work

As stated before, several other successful GWAPs have been made in the past. In this section, we discuss how the GWAPs the ESP game, Peekaboom, Phetch, and Verbosity work and how Swipe for Science differs from those games.

2.1 The ESP Game

The ESP game (Von Ahn and Dabbish, 2004) is a game designed for labeling images. Two players, who were randomly assigned to each other, are presented with an image and they keep guessing labels for a limited time. Within that time, they have to guess the same labels to score points.

2.2 Peekaboom

Peekaboom (Von Ahn, Liu, et al., 2006) is a game designed for locating objects in images. Two randomly assigned players take turns being the Peek and the Boom. Boom has to reveal the parts of an image that contain a specific object and Peek has to guess what object Boom is revealing.

2.3 Verbosity

Verbosity (Von Ahn, Kedia, et al., 2006) is a game designed for collecting common-sense facts. Two players are randomly assigned to each other. One of them is the Narrator and has to give hints about a word. The other is the Guesser and has to guess the word based on the hints from the narrator.

2.4 Phetch

Phetch (Von Ahn, Ginosar, et al., 2007) is a game designed for describing images. In a group of three to five players, one player has to describe a specific image and the others have to use the description to find that image again with an image search engine.

Swipe for Science differs in many ways from those games. Firstly, it collects an entirely different type of data. As stated before, this would be the first GWAP to collect discriminative evidence. More specifically, it collects data about how concepts are associated with contexts through pairwise comparison.

Secondly, this game aims for sustainable long-term engagement rather than short-term engagement. It attempts to make it a habit for users to play every day for a few minutes, which is also something no previous GWAP has done. This should result in more collected data in the long run.

Lastly, the threshold for playing the game should be much lower. The game is designed for mobile platforms, so it can

be played anywhere. It is a single-player game, so there is no need to wait for another player. The game can be played offline as well. The game is simplistic, easy to understand and intuitive to control with one hand.

3 Design of Swipe for Science

Swipe for Science is a mobile application that we created for this research. It is a single-player game that is designed to collect discriminative evidence from crowds. It takes some inspiration from GWAPs made in the past, but it has mostly been designed by creative processes, and trial and error.

3.1 Game Flow

In Swipe for Science, players are presented with two different contexts, one on the left side of the screen and one on the right. The players are then presented with a concept that they can swipe up, down, left, or right, depending on how they associate the concept with the contexts. Swiping to the left or right means that the concept is exclusively associated with the left or right context respectively, swiping up means that it is associated with both and swiping down means that it is associated with neither. Figure 1 shows a concrete example of this. Asking questions in this way only requires a single swipe to answer, making it simple, easy and intuitive. After they have swiped, a new concept will be presented. After they have swiped a batch of twenty concepts, the contexts will be switched and they continue swiping with a new batch.

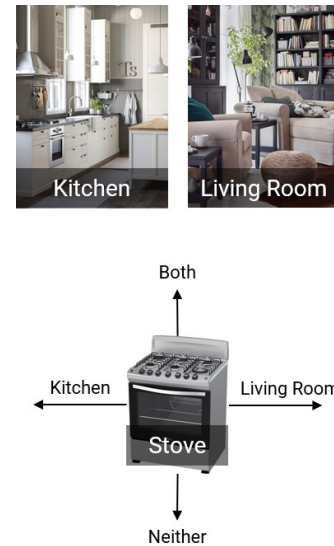


Figure 1: Example swipe instance

3.2 Dataset of contexts and concepts

The entire dataset of contexts and concepts that are combined to create questions consists of labelled images that we assembled ourselves. In total, there are 11 contexts and 80 concepts. When we selected which contexts would appear in the game, we took into consideration what different rooms there are in a house, what their distinct function is, and how globally well known we estimated these rooms to be. With the selection

of concepts, we mostly chose objects that can be commonly found in a house, but do not clearly belong in one specific room. This ensures that a few concepts have, however, been specifically chosen because they belong in one specific room. Those concepts function as ground truths, which we discuss further in section 3.5.

The full data set of labelled images used in this game can be found in Appendix D.

3.3 Question Generation

To ensure that we actually collect the data that we want, we need to start at asking the right questions to the player. In this section, we will discuss how Swipe for Science generates the questions that will be presented to the users.

Priority questions

The first step in generating a batch of questions is looking at the priority questions. Priority questions are specific questions that researchers can manually set in the database in order to collect specific associations more quickly. Each set of priority questions takes the form of a partial batch, so it has two contexts and one or more concepts. The question generator connects with the database and takes the first set of priority questions that the user has not completed yet and uses this as a base to create a full batch.

If there are no more priority questions that the user has not completed yet, then the question generator selects two random contexts and continues.

Ground truths

The second step is adding ground truths to the partial batch. The question generator retrieves the ground truths from the database that can be remotely updated, but there is also a local copy to fall back on. For each context, there are a few concepts that clearly belong in that context only. The question generator takes the ground truths for both contexts and adds them to the partial batch. The ground truths are important for detecting malicious input, which is discussed in section 3.5.

An overview of all the ground truths can be found in Appendix C.

Priority concepts

In the next step, the question generator adds priority concepts to the partial batch. Similarly to priority questions, priority concepts are specific concepts that researchers can manually set in the database in order to collect specific associations more quickly. The question generator repeatedly takes a random priority concept that the player has not encountered before with the specific contexts until the batch has 20 concepts or until there are no more priority concepts.

In the last step, the question generator will add random concepts to the partial batch until it has 20 concepts. After that, the order of the concepts get shuffled. Now the batch is finished and ready to be presented to the player.

3.4 Player Engagement Systems

Although the novelty and zen nature of the game might keep players engaged at first, additional systems are needed to keep them engaged in the long term. There are several

engagement systems implemented in Swipe for Science that will be discussed in this section.

Daily Streak

The first engagement system is the daily streak. For every consecutive day the player completes five batches, the streak will increment. If the player does not play for an entire day, the streak will be reset. This ensures that the player plays every day, but for only a limited time per day.

XP System

The second engagement system is the XP system. The players earn XP by playing the game. They earn 20 XP for every batch they complete and 100 XP for every time they complete a daily streak. They can also at any time view the total amount of XP they have accumulated. A similar score system was also present in the ESP game and was proven to be successful (Von Ahn and Dabbish, 2008).

Leagues

The third engagement system is the leagues system. There are six leagues: Wood, Stone, Copper, Iron, Gold, and Diamond. All players start in the Wooden league and need to earn enough XP before the end of the week to advance to the Stone league. In the next leagues there are two XP goals for every week; the first one is the amount of XP needed to stay in the same league and the next one is the amount of XP needed to advance to the next league. Since the Diamond league is the last league, there is only one XP goal for staying in the same league.

The league system is an improvement over the rank systems in the ESP game and Peekaboom. The rank systems has been proven to be effective because the player has a goal to work towards (Von Ahn and Dabbish, 2008). One flaw with that system that the league system aims to solve is that when the player has reached the final rank, there is no more goal to work towards. With the league system, there is also a final league, but the player has to keep playing to stay in that league.

3.5 Data Evaluation

To obtain usable machine learning data from the answers of the players of the game, the data needs to be processed in a certain way. We will discuss how this process works in Swipe for Science in this section.

Accounting for Malicious Inputs

A problem with all GWAPs is that there needs to be some method to account for malicious input. The way Swipe for Science handles this is having some questions in a batch for which the answers are already known. These so called “ground truths” can help determine if the player is answering the questions truthfully. If the player answers the ground truth questions wrong, that could mean that the player is not playing truthfully, but that is not always the case. Players could make mistakes or they might simply associate certain concepts differently than we do. To account for this, the batch will only be discarded when the player answers half of the ground truth questions wrong. This should ensure that there

will be no malicious input in the database. Additionally, the player will not receive any XP and the batch will also not count towards the daily streak, which discourages the player to play untruthfully.

Aggregating the answers per question

The next step is aggregating the data. For every question the answers of all players will be aggregated into a ranking. This process yields decimals of how the players associate the specific concept with the contexts. An example computation with random data is shown in figure 2.

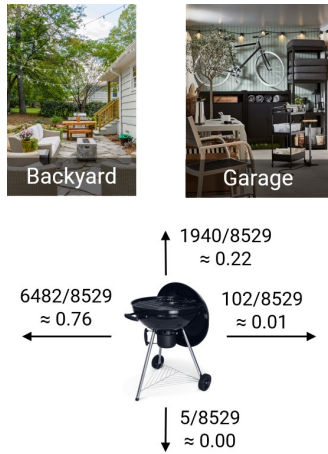


Figure 2: Example computation with random data

The final result is a dataset with all possible questions with aggregated answers. It is ready to be used for machine learning systems.

4 Evaluation

To get an indication of how well Swipe for Science actually engages players, how well the malicious input detection system works, and how good the quality of the collected data is, we performed a small experiment. The setup of the experiment is described in section 4.1 and the results are described in section 4.3.

4.1 Experimental Setup

The experiment was set up by distributing the application over the four participants. We told them to play the game for however long they felt engaged with it. We also asked a few of them to try to break the malicious input detection system. The game collected data about the participants, such as their current and highest daily streak, total number of batches completed, and, of course, the answers of the questions they were presented.

We would distribute an engagement review form after five days, but because we noticed that no one made it past a two-day streak, we decided to distribute the engagement form on the third day already.

The questions on the form were separated in three sections. The first section contained questions about the

game in general. These questions were based on the User Engagement Scale (UES) (O'Brien et al., 2018), but none of the questions of the UES forms were directly used for multiple reasons. The first reason is that the focused-attention (FA) factor does not apply since this game was not designed to absorb the player in the interaction. The second reason was that we wanted to make the questions more specific to this game and its goals.

The second section contained questions about the engagement systems. This section is the most extensive because we wanted to find out why the players were not engaged with the game. This section contains questions about how well these systems are balanced. It also contains questions specifically about why the player stopped playing and what they would change to make the game more engaging.

The third section contained a few open questions about the game. These questions are for collecting more qualitative data and these can be useful for additional feedback. The full form can be found in appendix E.

4.2 Evaluation Metrics

Engagement is evaluated in two ways. On one hand, we evaluate the engagement based on what the players do. Since the game has been designed for long-term engagement, so we look at whether players actually keep playing the game. We measure this by looking at the daily streak of each participant, which is saved in the database every time players complete a batch. On the other hand, we evaluate the engagement based on how the players think and feel about the game. We measure this by letting them fill in the aforementioned form and analyzing the results.

The quality of the data is evaluated in multiple ways. We analyze the quantity of the data that we have collected from the participants. We also ask the players that try to input malicious data if they succeeded and if they have found out how the system works. Additionally, we take a look at how accurate the results are for some priority questions, which are the following:

Contexts	Concepts
Backyard, Garage	Barbecue, Lawn Mower
Dining Room, Kitchen	Cutlery, Plate

Table 1: Priority questions used in this experiment

4.3 Results

As stated before, the game did not manage to retain the participants. Three participants got to a daily streak of 2 and one participant only got a daily streak of 1.

The engagement review form gives us some insights as to why. All answers can be found in appendix E but these are the most notable results. None of the participants feel that completing a daily streak is very rewarding, and for most participants it is not rewarding enough to incentivize them

to play every day. It also does not help that most participants do not like the game, think it is boring and too repetitive, and think it feels pointless to play the game.

Fortunately, the game does a lot of other things really well. The participants loved the simplicity of the game. Overall, they think the game looks good and the images shown are very clear. They think the controls are intuitive and that the flow of the game is quite smooth. The engagement systems are overall well balanced too. Although it could use a little tweaking, the core gameplay seems to be fine. It simply needs additional systems to make the game feel less repetitive and more engaging.

As for the quality of the data, the four participants together played 66 batches, so they answered a total of 1320 questions. This means that over the span of two days, each participant played an average of 8.25 batches or 165 questions per day. This is more than what we expected, since we expected that players would only play until they completed the daily streak, which would be only 5 batches or 100 questions per day. Although this might sound good, it does explain why they felt the game was too repetitive; they kept playing for far longer than intended way past the daily goal, so they would only stop if they got fed up with the game.

Two out of the four participants actively tried to play untruthfully. They attempted a total of 13 batches, of which one got accepted by coincidence. They did not figure out how the system works or how to break it consistently. There were also no reports from any participant about the system rejecting batches that were played truthfully. From this experiment, it seems that the malicious input detection is quite solid.

Now we take a look at the results of the priority questions. The results are summarized in the following figures:

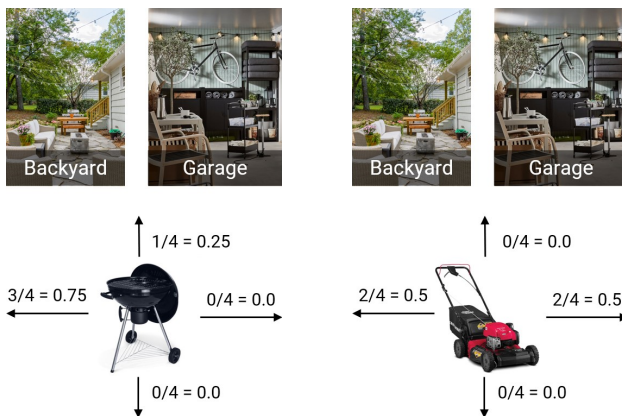


Figure 3: Priority question 1

Figure 4: Priority question 2



Figure 5: Priority question 3

Figure 6: Priority question 4

The associations in priority question 1 is what is to be expected. It is mostly associated with only the backyard because people use a barbecue in their backyard, but it is also associated with the garage because that is where a barbecue is stored. The results of priority questions 3 and 4 are also what can be expected, but the result of priority question 2 is not. The associations are divided in the extremes and no one associated the lawnmower with both or neither. Aside from that, the results overall seem to be good.

5 Responsible Research

There are a few ethical concerns that apply to Swipe for Science. These concerns will be discussed in this section.

5.1 Privacy

The application collects data from users, so privacy is naturally a concern. This is especially true because we are no security experts and because we did not have enough time to concern ourselves with implementing any security systems either. To tackle this concern, we made sure that no identifiable data would be collected. An id is used to save and retrieve which questions users have already answered to ensure that they do not answer a question multiple times, but no identifiable data is used for that; the id is randomly generated when data is first stored in the database. There is also no account system in place and the user is never prompted to enter a username of any kind. The only data that is stored in the database is the questions that were asked with the answers from the users.

5.2 Mobile Game Addiction

Another ethical concern that applies is mobile game addiction. A lot of games rely on operant conditioning to get people to play them (Vu, 2017). Operant conditioning is about reinforcing wanted behavior through rewards so that it is more likely to occur in the future (Skinner, 1938). A byproduct of this use is that some people can get addicted to these games. This is very bad for a players mental health, as it has been shown to be associated with social anxiety, depression, and loneliness (Wang et al., 2019).

This concern is quite tricky to tackle as it is in direct contradiction with the goal of a GWAP. A GWAP is supposed to be played a lot so that a lot of valuable data is collected, so in that aspect, the longer users play, the better. To incentivize the players, we had to implement systems that can be considered “addictive”, but we tried to limit it. There is an XP system in place where users earn XP by answering questions, but the most effective way to earn XP is with the daily streak. On one hand, it conditions users to play every day, but on the other hand, it deincentivizes users to play for the rest of the day and it instead incentivizes them to wait for the next day. This should create a habit in which users play every day, but only for a few minutes. Although some might consider this “too addictive”, we think this is a good balance and does not hinder the users in their personal lives.

Additionally, the experiment shows that players who do not like the game anymore, stop playing. Although this does mean that all the participants stopped playing, we are glad that none of them got addicted and that this game did not turn into a cheap Skinner box.

5.3 Copyrights

Some other ethical concern in this game is copyrights. Currently, Swipe for Science uses several copyrighted images. We did not consider copyrights when searching for images, so we could focus on finding the clearest and best looking images. This is fine for now as the game will not be publicly released in its current form, but it must be taken into consideration if the game will ever become publicly available.

5.4 Reproducibility

An ethical concern that applies to all research is reproducibility. That should not be a problem for the design of the game. We are convinced that we have described how the game works in sufficient detail. The source code¹ is also available online, but it is a private repository, so access needs to be requested first.

As for the experiment, although the setup can be reproduced, the results can be very different. This is because the experiment has major shortcomings. There were only four participants, which is very low. The participants were also biased, as they all knew us personally and understood the importance of the research project. The group of participants was not very diverse either. The planned experiment period of five days would be far too short for to properly test for long term engagement as well. Additionally, there was not enough data collected to properly assess its quality.

All these shortcomings make the results and conclusions unreliable, as any could be the result of sheer coincidence. Further more extensive research is required to make any conclusive claims about the game and the results of the experiment should only be considered as an indication of the true performance of the game.

¹<https://gitlab.ewi.tudelft.nl/cse3000/2020-2021/rp-group-14/rp-group-14-jshlim>

6 Future Work

The most important aspect of this game that needs improvement is player engagement. There are many ways to improve this and we only mention a few of all the possible systems in this section.

6.1 Competitive leagues

One way to improve player engagement is by adding a competitive element to the game. This can, for example, be done by altering the existing league system to be more like the league system of Duolingo. Duolingo is an app for learning foreign languages in a gamified way. In both systems, the players will go up or down a league or stay in the same league depending on how much XP they have collected in a week. In Duolingo, however, the XP goals are dependent on other users. Every week players enter a pool with 29 other players and are ranked based on how much XP they have earned that week. The top 5 players progress to the next league and the lower 5 players degrade to the previous league. This is an excellent way to retain players, according to Truong (2020), so Swipe for Science could benefit from that too.

6.2 Alternative game modes

Alternative game modes not only would freshen up the gameplay loop, which would help the player engagement, but it would also give the opportunity to collect different types of data. One idea was to present the player with multiple contexts at a time. The player would then be able to select all the contexts with which he associated the concept, which could speed up the rate at which data is collected.

Another idea was to present the player with multiple concepts at a time. This is especially interesting for concepts which association can change if bundles with other concepts. An example of this would be that a sink can be associated with both a kitchen and a bathroom, but combined with a stove, it would only get associated with a kitchen.

A good idea that someone suggested in the engagement review is a blitz mode, where the player needs to complete as many batches as possible within a given time limit.

Alternative game modes can be combined with the other engagement systems as well. They can, for example, be unlocked by reaching a certain total XP goal or by climbing up in the leagues. Or perhaps there could be a bonus round after completing a daily streak.

6.3 Aesthetic upgrade

The game currently does not have a very unique and distinct style, but it can be upgraded without losing the simplicity. Additions like sound effect, animations, visual effects, landscape mode, and dark mode can give the game more character and will positively impact player engagement.

7 Conclusion

In this paper, we answered the research question is: “How can we elicit discriminative evidence using a single-player

game?” The result is the GWAP named Swipe for Science, the game that collects discriminative evidence when people play it. The main gameplay loop consists of swiping concepts up, down, left, or right based on how it is associated with the given contexts. Players are incentivized keep playing by various engagements systems such as the daily streak and league systems. The experiment shows that the game is experienced as simplistic and intuitive to control, and that it seems to collect data of overall good quality, but it does not manage to engage the players enough to retain them.

Bibliography

- Alexander, R. (2018). *Different Types of Knowledge: Implicit, Tacit, and Explicit*. Retrieved June 26, 2021, from <https://bloomfire.com/blog/implicit-tacit-explicit-knowledge/>
- Gong, Z., Zhong, P., & Hu, W. (2019). Diversity in Machine Learning. *IEEE Access*, 7, 64323–64350. <https://doi.org/10.1109/ACCESS.2019.2917620>
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28–39. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2018.01.004>
- Skinner, B. (1938). The behavior of organisms: an experimental analysis.
- Truong, D. (2020). *Gamification Elements in Developing Mobile Application* (Bachelor's Thesis). Retrieved June 27, 2021, from <http://urn.fi/URN:NBN:fi:amk-2020112524307>
- Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326.
- Von Ahn, L., & Dabbish, L. (2008). Designing Games with a Purpose. *Commun. ACM*, 51(8), 58–67. <https://doi.org/10.1145/1378704.1378719>
- Von Ahn, L., Ginosar, S., Kedia, M., & Blum, M. (2007). Improving image search with phetch. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, 4, IV–1209.
- Von Ahn, L., Kedia, M., & Blum, M. (2006). Verbosity: a game for collecting common-sense facts. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 75–78.
- Von Ahn, L., Liu, R., & Blum, M. (2006). Peekaboom: a game for locating objects in images. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 55–64.
- Vu, D. (2017). An Analysis of Operant Conditioning and its Relationship with Video Game Addiction.
- Wang, J.-L., Sheng, J.-R., & Wang, H.-Z. (2019). The Association Between Mobile Game Addiction and Depression, Social Anxiety, and Loneliness. *Frontiers in Public Health*, 7, 247. <https://doi.org/10.3389/fpubh.2019.00247>

Appendix

A App Icon



Figure 7: Swipe for Science Launcher Icon

B Screenshots

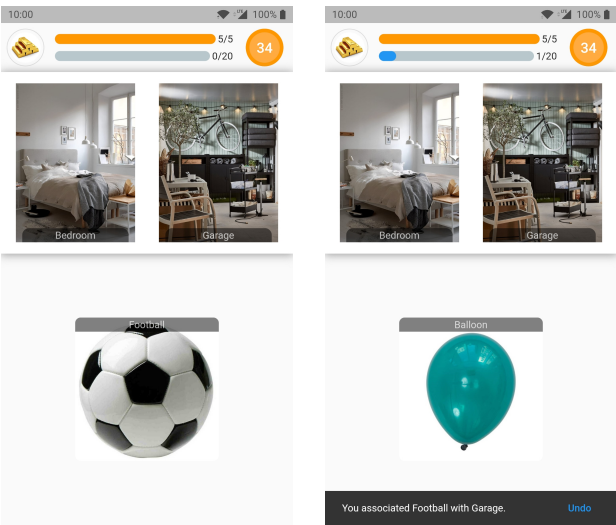


Figure 8: Game Screen

Figure 9: Swipe completed



Figure 10: Batch Accepted

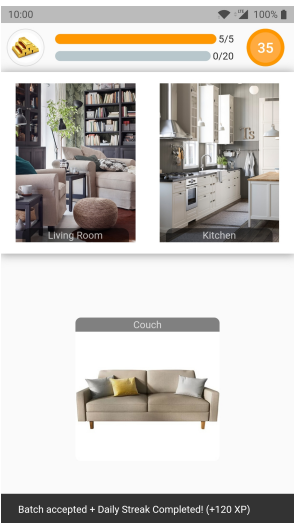


Figure 11: Streak Completed

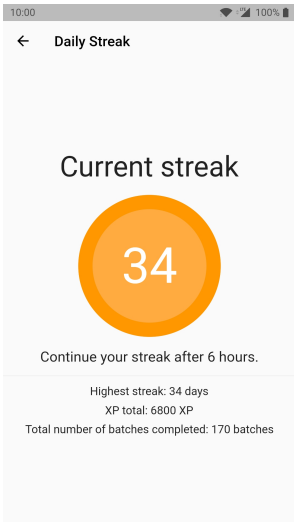


Figure 12: Daily Streak Screen

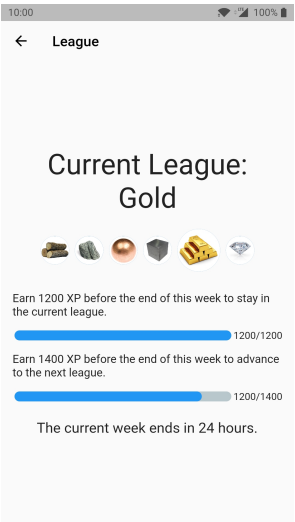


Figure 13: League Screen

C Ground truths for each room

Backyard	Inflatable Swimming Pool, Tree
Balcony	
Bathroom	Toilet Paper, Toilet
Bedroom	Alarm Clock, Bed
Dining Room	
Garage	Bicycle Pump, Tires
Hallway	Coat Rack
Home Office	Office Chair
Kitchen	Colander, Frying Pan, Stove
Laundry Room	Laundry Detergent Pods, Washing Machine
Living Room	Couch

Table 2: Overview of all ground truths

D Data set of labelled images currently used in Swipe for Science

D.1 Contexts



Figure 14: Backyard



Figure 15: Balcony



Figure 16: Bathroom



Figure 17: Bedroom



Figure 18: Dining Room



Figure 19: Garage



Figure 20: Hallway



Figure 21: Home Office



Figure 22: Kitchen



Figure 23: Laundry Room



Figure 24: Living Room

D.2 Concepts

The concepts that are used as ground truths are marked with an asterisk (*). See Appendix C for an overview of all the ground truths.



Figure 25: Alarm Clock*



Figure 26: Balloon



Figure 27: Barbecue



Figure 28: Bed*



Figure 37: Clock

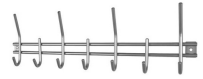


Figure 38: Coat Rack*



Figure 29: Bicycle Pump*



Figure 30: Books



Figure 39: Colander*



Figure 40: Computer



Figure 31: Bowl



Figure 32: Broom



Figure 41: Couch*



Figure 42: Cutlery



Figure 33: Bucket



Figure 34: Calculator



Figure 43: Dog



Figure 44: Drill



Figure 35: Cat



Figure 36: Chair



Figure 45: Flashlight



Figure 46: Fly Swatter



Figure 47: Fly



Figure 48: Football



Figure 57: Knife



Figure 58: Ladder



Figure 49: Fridge



Figure 50: Frying Pan*



Figure 59: Lamp



Figure 60: Laundry Detergent Pods*



Figure 51: Glasses



Figure 52: Hammer



Figure 61: Lawn Mower



Figure 62: Microphone



Figure 53: Headphone



Figure 54: Ice Skates



Figure 63: Mirror



Figure 64: Mouse

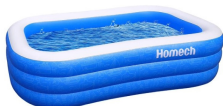


Figure 55: Inflatable Swimming Pool*



Figure 56: Keys



Figure 65: Newspapers



Figure 66: Office Chair



Figure 67: Painting



Figure 68: Pen



Figure 77: Road Cone



Figure 78: Rock



Figure 69: Phone Charger



Figure 70: Piano



Figure 79: Ruler



Figure 80: Safe



Figure 71: Pickaxe

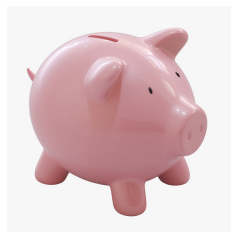


Figure 72: Piggy Bank



Figure 81: Screwdriver



Figure 82: Ship in Bottle



Figure 73: Plate



Figure 74: Pogo Stick



Figure 83: Shoes



Figure 84: Shovel

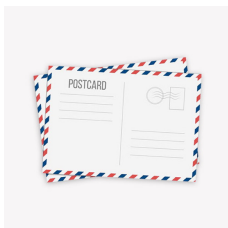


Figure 75: Postcard



Figure 76: Potted Plant



Figure 85: Snorkel



Figure 86: Stool



Figure 87: Stove*



Figure 88: Sunflower



Figure 97: Toy Slide



Figure 98: Tree*



Figure 89: Sunglasses



Figure 90: Switch



Figure 99: TV



Figure 100: Vacuum



Figure 91: Table



Figure 92: Tires*



Figure 101: Wallet



Figure 102: Wardrobe



Figure 93: Toilet Paper*



Figure 94: Toilet*



Figure 103: Washing Machine*



Figure 104: Watering Can



Figure 95: Toothbrush



Figure 96: Towel

E Engagement Review

The following 6 pages are the questions from the engagement review. The 10 pages after that contain all the answers from the engagement review.

Swipe for Science Engagement Review

Hello everyone! Looks like no one made it past a 2-day streak and everyone's daily streak has expired by now. The game is apparently not as engaging as I intended it to be, so now I need to figure out why that is. I would appreciate it very much if you would take a few minutes to help me out by filling in this form.

First there are a few general questions about the game.

***Required**

1. The controls of the game are... *

Mark only one oval.

	1	2	3	4	5	
Very confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very intuitive

2. In terms of overall aesthetics, the game looks and feels... *

Mark only one oval.

	1	2	3	4	5	
Very off-putting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very appealing

3. The images used in this game are... *

Mark only one oval.

	1	2	3	4	5	
Very vague	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very clear

4. The flow of the game feels... *

Mark only one oval.

	1	2	3	4	5	
Very clunky	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very smooth

5. In terms of effort, playing the game feels... *

Mark only one oval.

	1	2	3	4	5	
Very demanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very relaxing

6. Do you like playing the game? *

Mark only one oval.

	1	2	3	4	5	
Not at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Totally

**Questions about
the engagement
systems**

These questions are about the XP system, the daily streak system, and the league system, and how these systems affect your engagement with the game.

7. Answering 20 questions per batch is... *

Mark only one oval.

	1	2	3	4	5	
Too little	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too much

8. Completing a batch feels... *

Mark only one oval.

	1	2	3	4	5	
Pointless	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very rewarding

9. Completing 5 batches to increment the daily streak is... *

Mark only one oval.

	1	2	3	4	5	
Too little	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too much

10. Completing the daily streak feels... *

Mark only one oval.

	1	2	3	4	5	
Pointless	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very rewarding

11. Receiving 100 XP per completed daily streak (in comparison to 20 XP per completed batch) is... *

Mark only one oval.

	1	2	3	4	5	
Too little	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too much

12. Waiting 18 hours to continue the daily streak is... *

Mark only one oval.

	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too long

13. Having 24 hours to increment the daily streak is... *

Mark only one oval.

	1	2	3	4	5	
Too stingy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too generous

14. Having 7 days to progress through the leagues is... *

Mark only one oval.

	1	2	3	4	5	
Too short	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too long

15. The XP goal to progress from the Wooden league to the Stone league (600 XP in 7 days) is... *

Mark only one oval.

	1	2	3	4	5	
Too low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too high

16. Why did you not (continue to) play the game every day? *

Tick all that apply.

- ☐ I forgot about it.
- ☐ I didn't have enough time for it.
- ☐ I don't like playing the game.
- ☐ Playing the game feels pointless.
- ☐ Completing a daily streak does not feel rewarding (enough).
- ☐ I had already collected enough XP to progress to the next league.

Other: ☐ _____

17. What changes would you make to the game so that you would play it every day? *

18. What other changes would you make to the game to make it more engaging? *

Some last few questions

These are some last few open questions. You're almost done!

19. What do you like about the game? *

20. What do you dislike about the game? *

21. Do you have anything else you want to say about the game?

This content is neither created nor endorsed by Google.

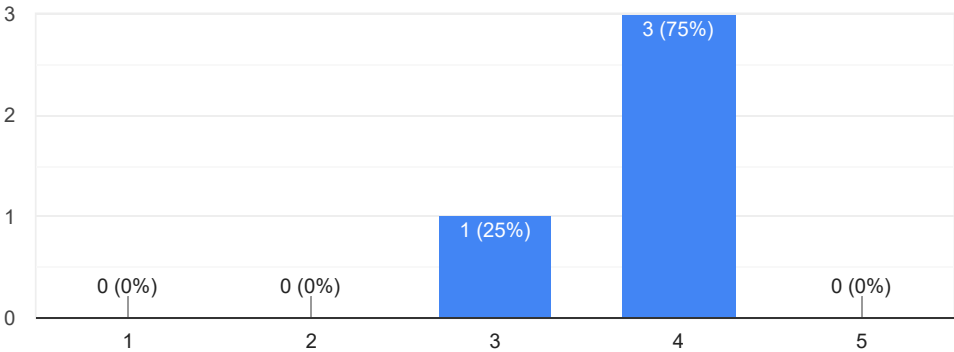
Google Forms

Swipe for Science Engagement Review

4 responses

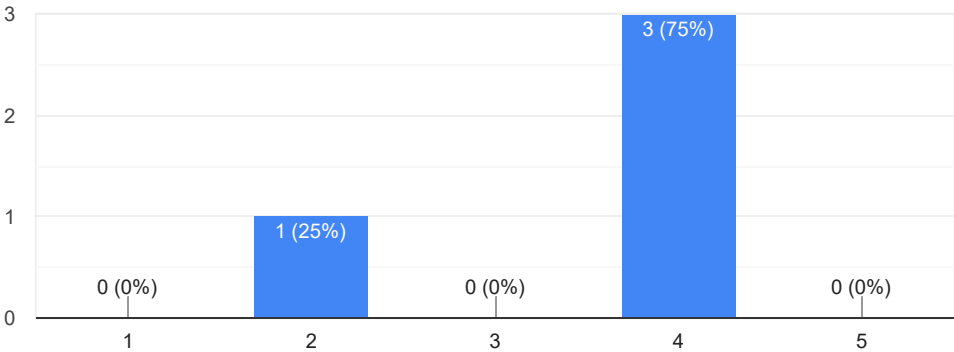
The controls of the game are...

4 responses



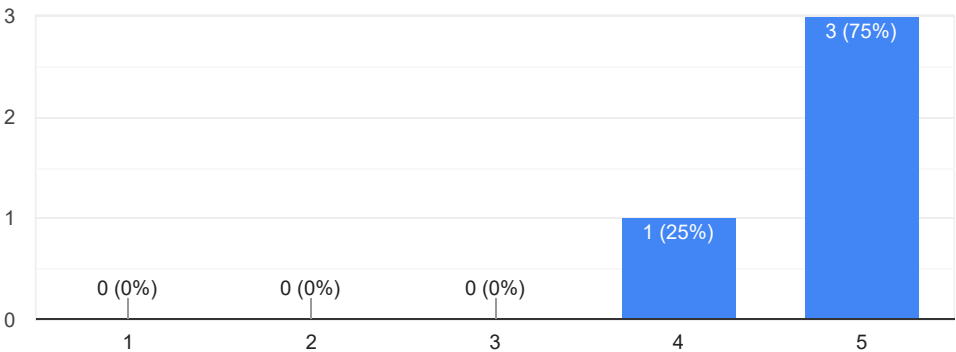
In terms of overall aesthetics, the game looks and feels...

4 responses



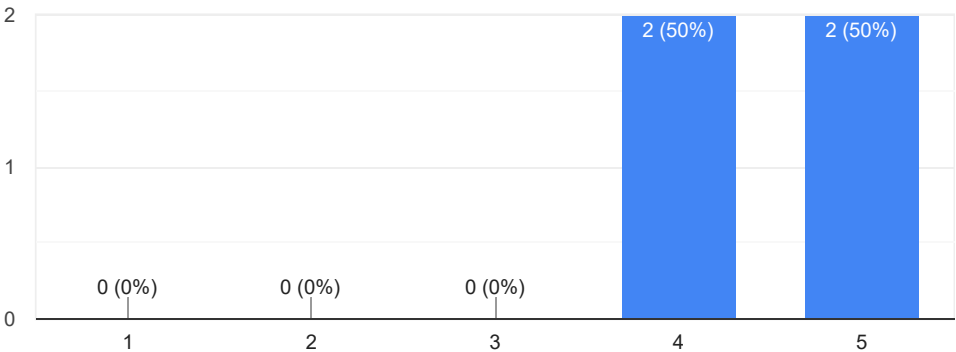
The images used in this game are...

4 responses



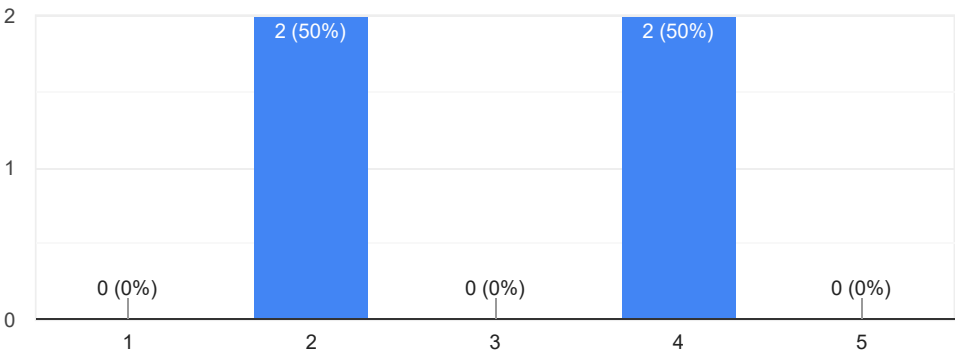
The flow of the game feels...

4 responses



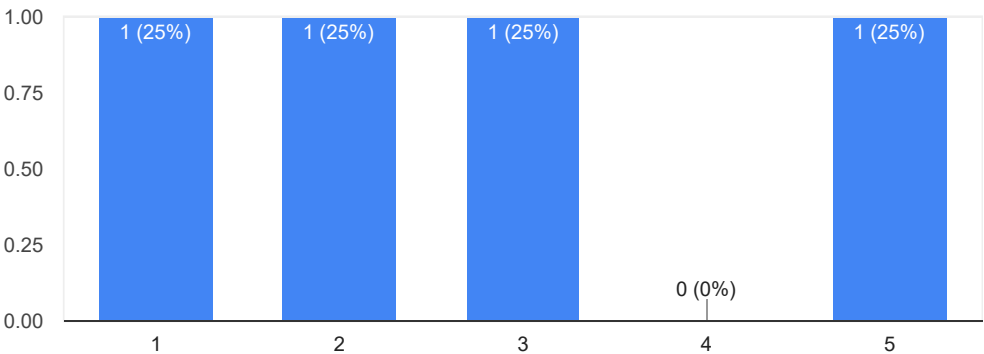
In terms of effort, playing the game feels...

4 responses



Do you like playing the game?

4 responses

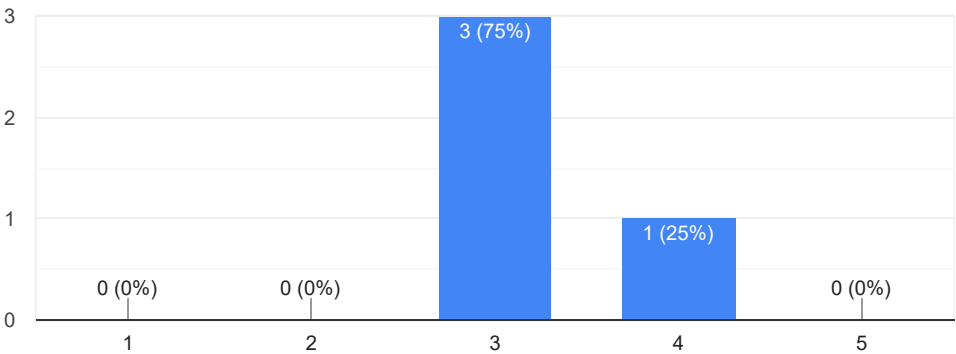


Questions about the engagement systems



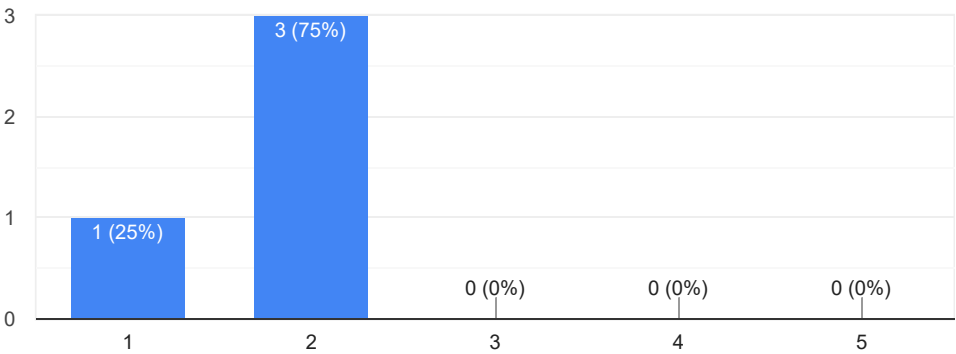
Answering 20 questions per batch is...

4 responses



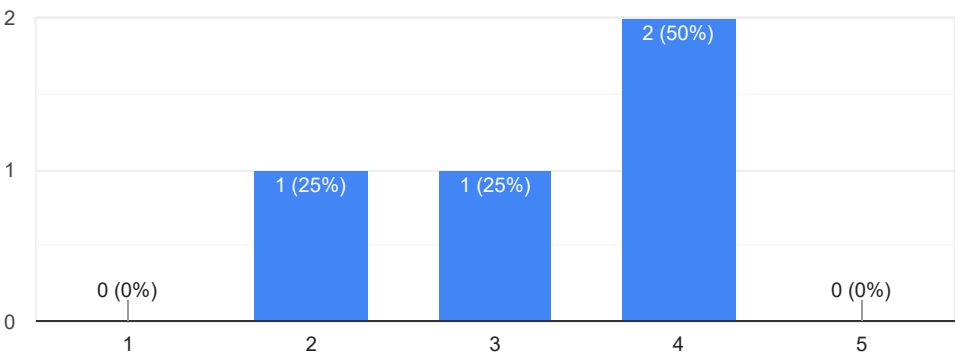
Completing a batch feels...

4 responses



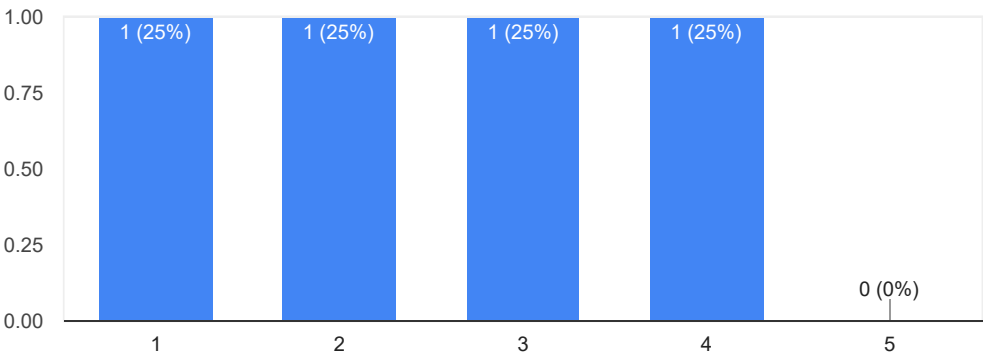
Completing 5 batches to increment the daily streak is...

4 responses



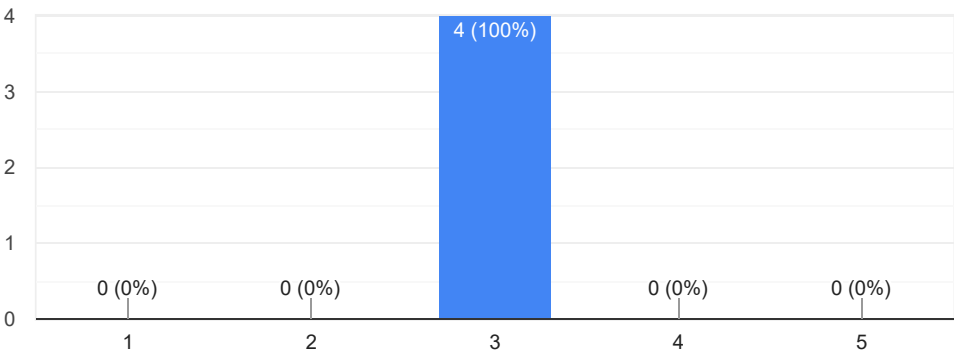
Completing the daily streak feels...

4 responses



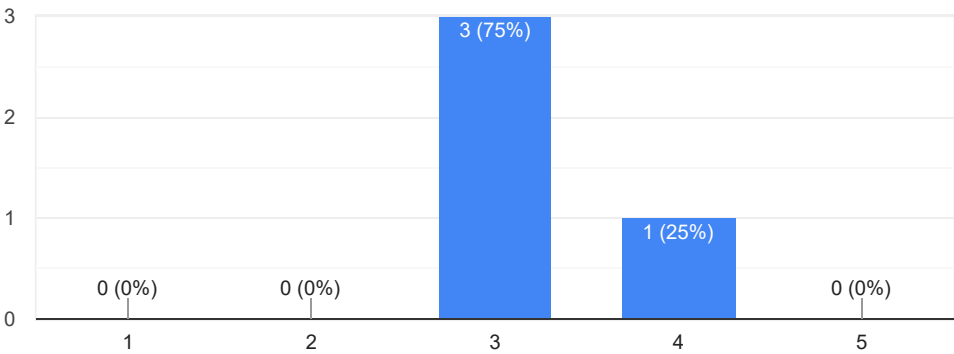
Receiving 100 XP per completed daily streak (in comparison to 20 XP per completed batch) is...

4 responses



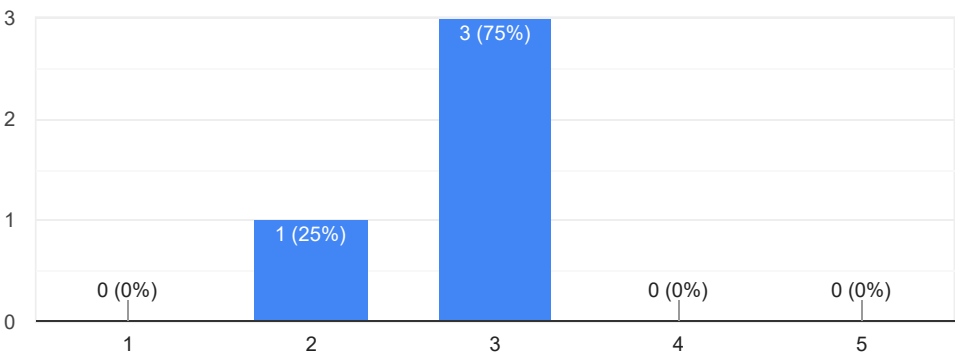
Waiting 18 hours to continue the daily streak is...

4 responses



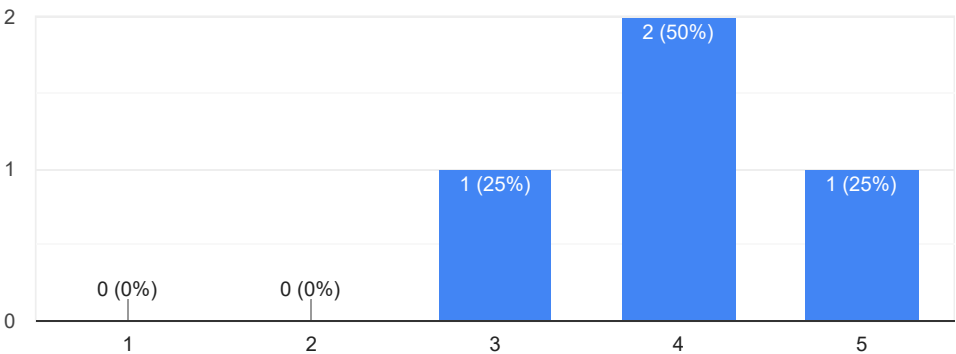
Having 24 hours to increment the daily streak is...

4 responses



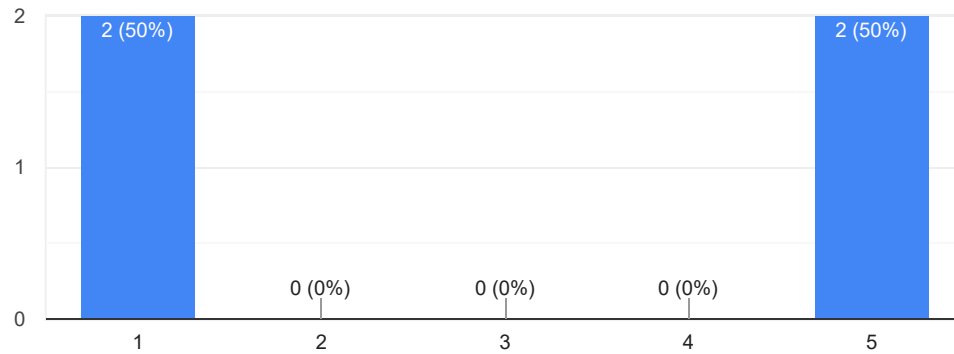
Having 7 days to progress through the leagues is...

4 responses



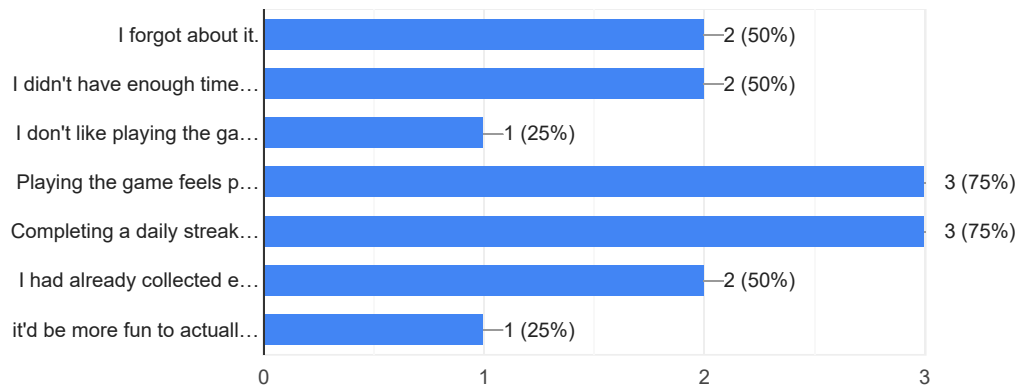
The XP goal to progress from the Wooden league to the Stone league (600 XP in 7 days) is...

4 responses



Why did you not (continue to) play the game every day?

4 responses



What changes would you make to the game so that you would play it every day?

4 responses

Make more variation in the subjects and rooms. It felt a bit repetitive

Some sort of progression system where you can unlock things/features/gameplay. The current system is just cosmetic, and personal. Nothing new of value is awarded when reaching a new rank.

Some extra competition

Make it more fun! Make it a game you would like to play! (Sound fx, colors, visual effects, interaction, give it character, cooler name, crazy power ups, give users a goal, give users freedom of tactics)

Look up what other games did before you and what they did right, maybe look up analyzations of those games. (YT: Game Maker's Toolkit)

What other changes would you make to the game to make it more engaging?

4 responses

Make it maybe also in landscape mode

Leaderboard, some competitive factor.

More game modes? Perhaps blitz or a special daily challenge, maybe a gamemode where you should get as many right as possible by guessing what another player thinks.

See previous

Some last few questions



What do you like about the game?

4 responses

The easy swiping

The interface is clean and smooth

The simplicity

It's clear to read and it controls easy

What do you dislike about the game?

4 responses

The repetitiveness of the subjects

Get's boring and repetitive quite quickly

The cheat detection! Screw you xD it's pretty good.

It feels pointless but mostly just not that fun

Do you have anything else you want to say about the game?

2 responses

I like it. Very well done. Obv i got no time to play it. Sadly.

I like the idea of helping an AI (maybe use that in the game itself)

This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#).

Google Forms

