An Analysis on the User Experience Implications of Username Formats for Recognition and Differentiation in Large Populations of Pseudonymous Individuals: Pre-Report

Fýrine Holm[1], J Pau[2], MoonshineAddict[3], Sego, G[4], and AliveR[4]

[1]Medium Sprinkler
[2]LoodleDoodles
[3]Unaliased
[4]Unaliased
[5]Unaliased
Of the **Sprinkler Nametag Study Research Group**

September 10, 2025

Preamble. This pre-report outlines the motivation, design, some preliminary results, and the implications for our username recognition study. We describe effectiveness, reaction time, and recall across pseudonym formats.

Figure 1: A pseudonymous user assigned the obfuscated username Healer358 describing their medical profession followed by their unique numeric ID, from Trial 1 of a civilization experiment run by ShotRush (2025b)

Contents

1 Disclaimer

v2.0 of the study is live as of the publication of this pre-report! The reader is asked to partake in the study regardless of whether or not they have been involved with the study in the past, or ShotRush's experiment. All submissions are highly valued and the basis of our coming report! The study platform is accessible from the link `https://sprinkler-nametag-test.fyr.li`
The main reason for the release of a pre-report before the study is finished gathering data is to make more people aware of the study, of which version 2 is going live simultaneously with this pre-report, which will hopefully convince more players to take v2.0 of the study and give us more data to work with on the final report. This pre-report is designed to be released prior to the completion of the final version of the study (and its accompanying report), and therefore we find it necessary to begin this document with a disclaimer that any conclusions one might be tempted to draw from the data in this pre-report are based on an inferior study methodology, and we advise patience until the final report is released. The contents of this pre-report are entirely preliminary to the final report and should not be taken as a substitute.

We have made the decision to exclude an abstract and conclusions section in this pre-report as our team has deemed it inappropriate to preempt the final report. We will refrain from drawing conclusions or interpreting the data that we have so far. We will instead focus on sharing our findings and analysis so far. This pre-report has been condensed to make it more easily digestible, as the extra details will be saved for the final report.

2 Plain Language Summary

This pre-report is being written partway through a series of studies undertaken by the Sprinkler Nametag Study research group. The pre-report details an ongoing experiment and investigation within the vanishingly narrow intersection of leading research in the field of User Experience and Interface (UX/UI) Design and ShotRush's Minecraft Civilization experiment (testing Joyful's[1] theory of specialization, a [digital] sociology theory wholly unrelated to UX/UI design). We seek to compare how different styles of username formatting and presentation could affect the ability of users to differentiate and recognize a single user's username in an ever-growing pool of usernames that the given user has been exposed to.

---

[1] Joyful has recently changed his name to Joyfuul (with two 'u's) for unknown reasons. Despite this, the research community continues to refer to his theory under the name he published this theory under, Joyful.

  We chose to follow this convention.

3 Public Significance Statement

This research is significant because it advances and promotes accessibility within communities of pseudonymous individuals, especially individuals with poor recall, or language/memory disorders such as dyslexia or Parkinson's disease. This research also benefits the Everyman, because our research looked to determine the systems that worked better or worse for everyone. If applied, this research may increase the amount of active social connections that someone is able to have in a pseudonymous environment, allowing for more deeply connected communities with wider engagement.

4 Background

The inspiration for this study (and indeed even the research team backing it) was the wealth of discontent shared by many participants about the username system. Although every participant seemed to have their own, unique suggestion for resolution (and perhaps distaste for any rival suggestion), nearly all of these suggestions varied wildly. We were made aware of this situation and we determined that it had been established that the current system may be inadequate, but simultaneously that there was no consensus on what the proper replacement would be. The environment that created this situation is worth briefly covering, in order to understand how so many participants could be similarly discontented yet so disjointed and divided as for a replacement.

4.1 ShotRush's Automatic Participant Pseudonyms

In ShotRush's ongoing digital sociology experiment, they deploy an iterative trial system. In these trials, the design of the pseudonym system was a simple, inoffensive, and logical format that upon cursory inspection does not invite criticism or show any obvious fault.

4.1.1 Rationale

ShotRush's experiment, like all experiments had confounding variables to control for. One of these was, unusually, the participants themselves. In the first trials of ShotRush's experiment, the total amount of players in attendance was few, and they had the ability to converse among themselves prior to the trials taking place, and even in-between sessions of an going trial. As a sociology experiment, the risk of these external interactions contaminating the results was unacceptably high (ShotRush, 2025a). ShotRush also explained that "[because of] the tendency for participants to gravitate towards recognizable personalities or content creators, all participants will remain anonymous during [the trials]" (ShotRush, 2025a). These reasons were the basis for the need to create an automated system for assigning pseudonyms to participants.

4.1.2 Design

The ShotRush pseudonym system has two iterations. In the zeroth trial, pseudonyms were rendered in the format `RoleXYZ` where `Role` was the name of that participant's specialization, and `XYZ` represent a random unique 3-digit identification number that would persist throughout the trial even when the participant changed careers. The resulting names of this system appeared as `Healer358`, `Guardsman904`, or `Farmer204`[2].

This system was redesigned following the zeroth trial, to prevent names from changing when participants re-skill themselves. In subsequent trials, the new system design rendered pseudonyms in the format `Participant_XYZ`, with roles now replaced by the one-size-fits-all `Participant`, an underscore, and the same 3-digit identifier (without leading zeroes), resulting in outputs such as `Participant_841`, `Participant_920`, and `Participant_77` (Holm, 2025b)[3].

4.1.3 Expected usage

In an interview with a member of the ShotRush team, they expressed that the intended purpose for their system design were initially to express the participant's role, and later not even that. The intent, as explained, was not for participants to use their assigned numbers as designations or monikers, but for participants to instead use eachother's avatar, actions, and mannerisms to come up with new names unlinked to external or real life baggage that could be used to deanonymize participants. The unwieldiness and indistinguishability of the provided pseudonyms was designed to intentionally incentive participants to generate their own working names in the moment for eachother.

4.1.4 Actual usage

In practice, through the course of several trials and sessions, participants relied very heavily on the numeric identifiers, using them as names. In one recorded case, over the span of 93 seconds, these unwieldy numeric identifiers were verbally uttered at least 18 times, Here is a transcript of this encounter.

This transcription shows the strain on communication that the reliance on these identifiers adds. (KrazyKTV, 2025, 9:32-11:05) Upon further investigation, barely any emergent names were used in Trial 0, the only two examples confirmed to have been widely used were `President Red`, named for his red-colored avatar, and `Fayman` [4] [5]

---

[2]These examples are taken from KrazyKTV (2025)

[3]This citation references a work of an author on this paper.

[4]This spelling is unconfirmed and the nickname may be spelled `Fairyman` or something similar. We only heard it spoken by KrazyKTV (2025).

[5]We were unable to find the origin of this nickname

```
Ninety-five? Ninety-five? Ninety-five?
...
Eight-eight-four? Let me go. Eight-eight-four, let me go.
Why is everybody...
Eight-eight-four? Let me go.
Eighty-four, stop. Eight-eight-four, stop.
Get Seven-fifty. Seven-fifty. Seven-fifty.
Eighty-four is doing this. Eighty-four is doing this.
Kill the dogs.
Which person is the-
Which person is the heretic?
Everybody's attacking.
It's Seven-fifty. Seven-fifty is attacking.
Seven-fifty is attacking. Seven-fifty. I think this out.
...
Is this him?
Get this. That's him. Oh, not me. Nine-zero-nine.
```

Table 1: A transcript of a 93-second interaction from ShotRush's Trial 0 (KrazyKTV, 2025, 9:32-11:05).

4.2   Barriers to replacing usernames in UX/UI Design

Usernames are one of the most critically important features of internet discourse and interaction between users. Despite this, their nature and presentation is rarely discussed, due in large part to the fact that very little has changed in the design and structure of usernames since the inception of the World Wide Web. Usernames have, from that inception, been unique string identifiers chosen by the user, typically constrained to the Latin alphabet and some maximum permissible length. While plenty have considered the impacts of different usernames from the perspective of the user choosing their own name, Across the entire history of online usernames, only a few changes have come up over time. One notable change made necessary by the growth of the internet is the inception of discriminators, which allow multiple users to occupy the same string by adding a numeric component, preventing common terms from becoming "taken". Aside from this small addition which is still not widely adopted, there has never been a pressing need for the design of usernames to fundamentally change.
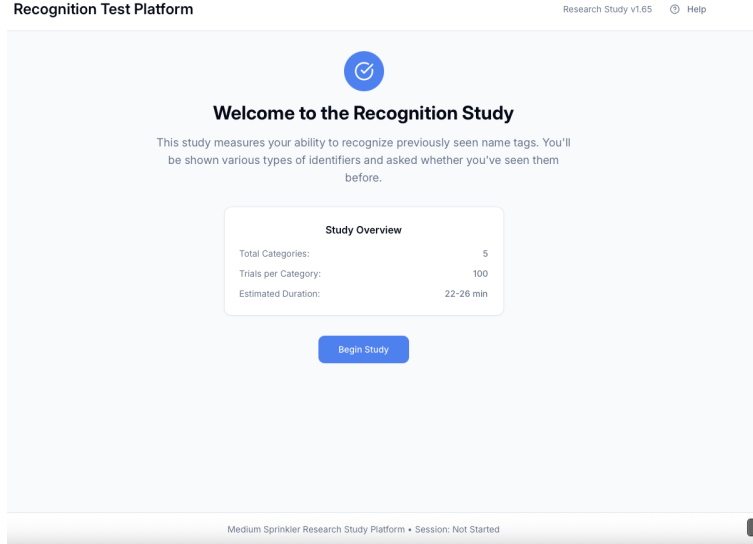
Figure 2: A screenshot of the Sprinkler Nametag Study participant landing page (Holm, 2025a).

5 Sprinkler Nametag Study

The Sprinkler Nametag Study[6], namesake of the research team authoring this pre-report, is an online research study platform website used to gather data on how well participants can recognize and tell apart different generations using the same format of usernames. Versions 1.0 and 1.65 of this test were conducted after September 4 and September 6, 2025, respectively.

5.1 Methodology and Design

This study presupposed that the primary issue with the ShotRush pseudonym system was the amount of pseudonyms generated by the system that could be differentiated, recognized, and remembered by the average player. Versions 1.0 and 1.65 only tested short-term memory, not long term memory. The study also didn't shuffle the order of category presentation per session.

5.2 Categories

For our study, we chose to track the following categories of pseudonym formats for analysis. These categories were used for all 1.X versions of the study.

1. Category 1: format `Participant_XXX` such as `Participant_841`

---

[6]This study was fully funded by the Sprinkler Nametag Research Group and sponsored by Fýrine Holm, one of the authors of this pre-report.
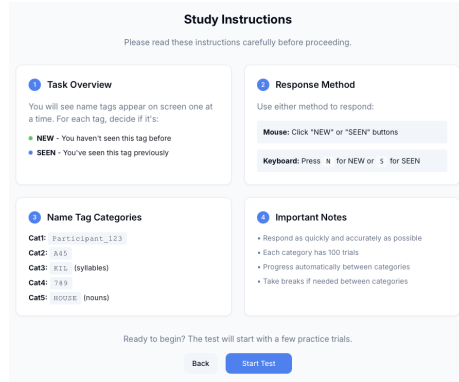
Figure 3: A screenshot of the Sprinkler Nametag Study participant onboarding page (Holm, 2025a).

2. Category 2: format $\boxed{\texttt{LXX}}$ such as `G45`

3. Category 3: format $\boxed{\texttt{SYLLABLE}}$ such as `ZAG`

4. Category 4: format $\boxed{\texttt{XXX}}$ such as `841`

5. Category 5: format $\boxed{\texttt{NOUN}}$ such as `HOUSE`

These categories will be referred to be their numbers for the remainder of this pre-report. The reader is advised to remember their order.

5.3 v1.0 Set-up and Study Results

The study was set-up in a hurry on September 4, 2025 and immediately went public to start gathering data. This was version 1.0, evaluating the 5 original categories. Version 1.0 tracked users on their average, max, and min response time per category as well as percentage accuracy across 49 trials per category per user[7]. This version of the study has no security measures to prevent from intentional deception by participants. Due to the fact that 1.0 did not track the individual series of accuracy, only the average accuracy, as well as the aggregate stats for response times but not the individual response time per test, we gathered less data from v1.0 than v1.65 per user.

Additionally, in the first version of the test, users were presented with no interstitial screen in between the testing categories to reduce test fatigue, while in v1.65 there is a new interstitial screen in between categories that offers users a rest. This means that we could expect the results of the data gathered from v1.0 of the test to be tainted by test fatigue.

---

[7]The user was presented with 50 trials per category, but due to a software bug on the client, the first response to each category was not recorded properly.

This 1.0 version of the test received 41 responses that were reviewed, analyzed, and those 40 [8] responses were charted in this pre-review to demonstrate their patterns.

Category	Mean (%)	Median (%)	Mode (%)
1	74.33673469	73.46939	73.46939
2	78.01020408	79.591835	85.714287
3	87.04081633	89.79592	97.959185
4	69.48979592	69.38776	73.46939
5	93.57142857	95.91837	100

Figure 4: v1.0 Table of Average Scores

5.3.1 Category 1 Data

As noted in Figure 4, scores for Category 1 averaged 74.3%. For a fuller picture, however, Figure 5 below compares scores to response time. Note how most scores clump around 60-85% and response times around 1-2 seconds.

Figure 5: v1.0 Individual Participant Results (Cat 1)

[8] A keen-eyed reader may note that 40 is less than 41. After careful review of the responses to version 1.0 of the test, we removed one result from the final analysis on heavy suspicion of tampering due to wildly unreasonable values that deviated sharply from the other responses by orders of magnitude and hinted to us that the participant who submitted it had likely intended to intentionally skew the final results by performing well, if average on their category of choice and taking inhuman amounts of time to get half of all the questions wrong on all other categories. We did not take this removal lightly.

Additionally, focusing only on score distribution, the Figure 6 shows the distribution of scores by intervals of 5%. Note the peak of 13 scores around 70-75%.

Figure 6: v1.0 Distribution of Scores (Cat 1)

5.3.2 Category 2 Data

Scores for Category 2 averaged 78%, as seen in Figure 4. Figure 7 on the next page compares accuracy to response time. Note how most scores clump around 75-85% and response times around 1-2 seconds.
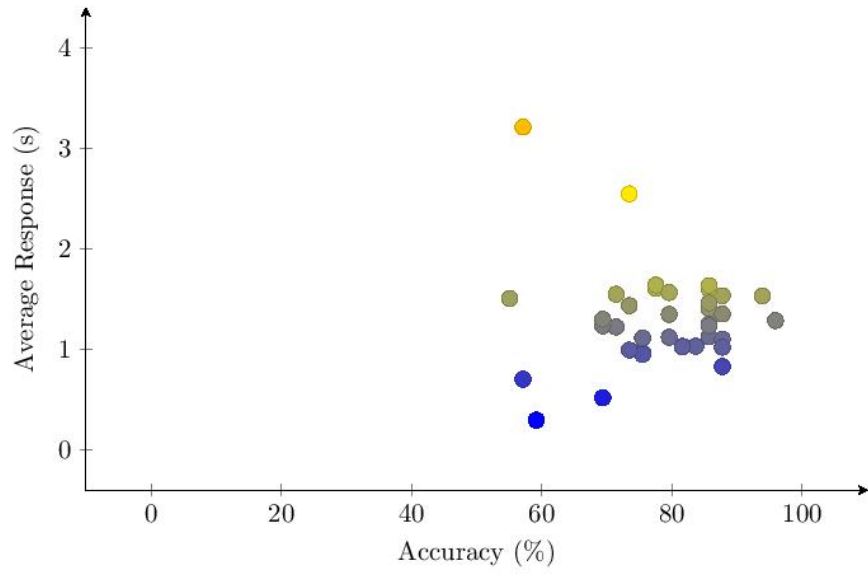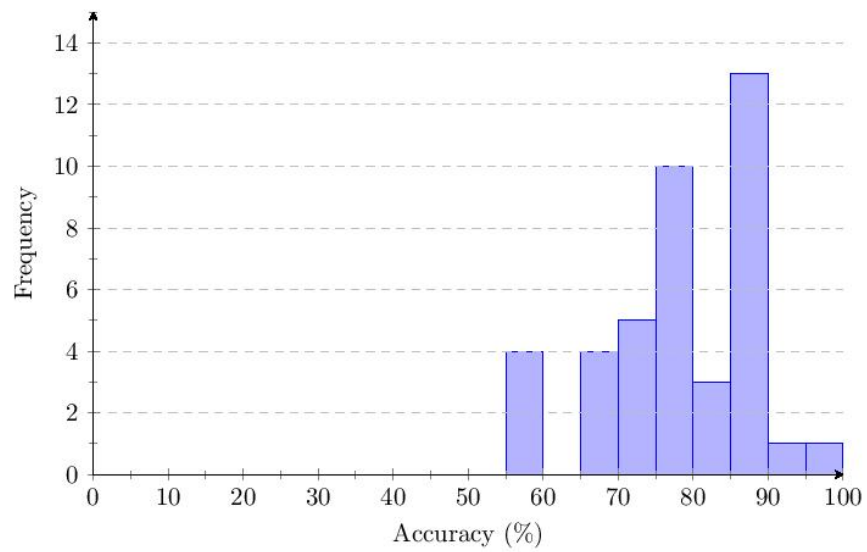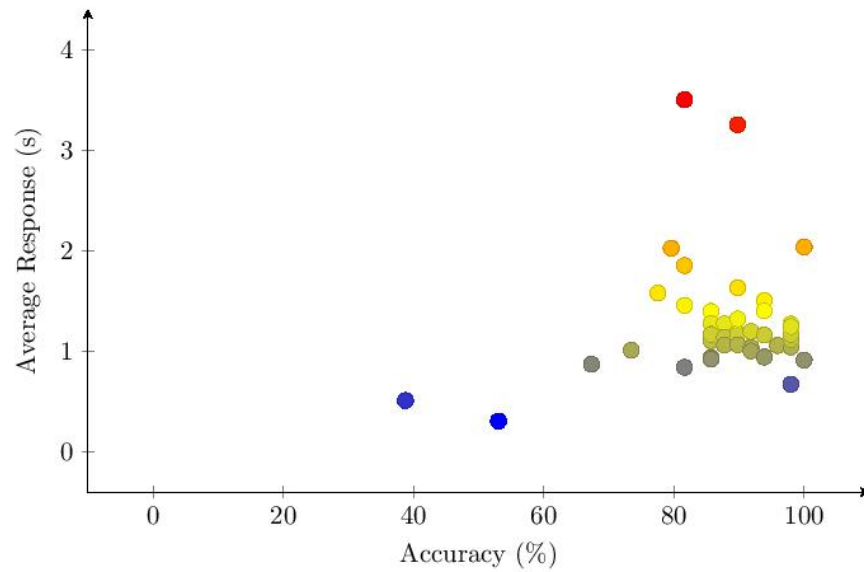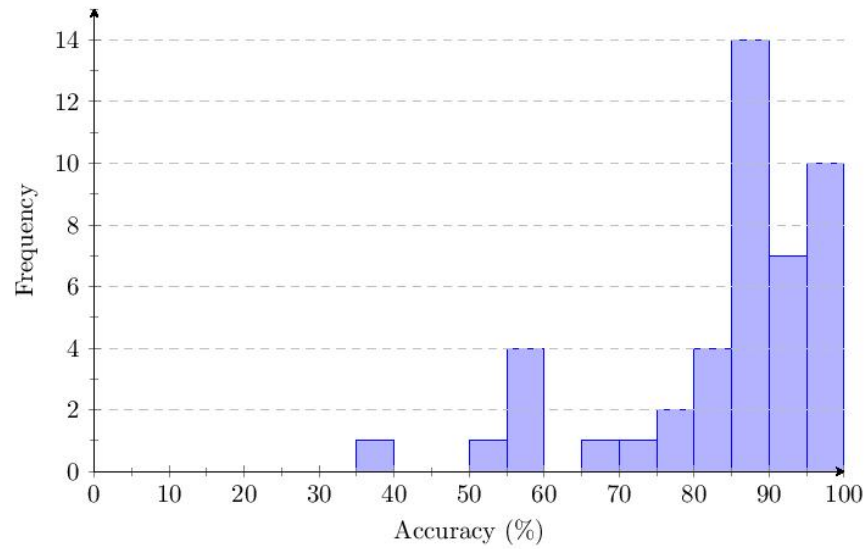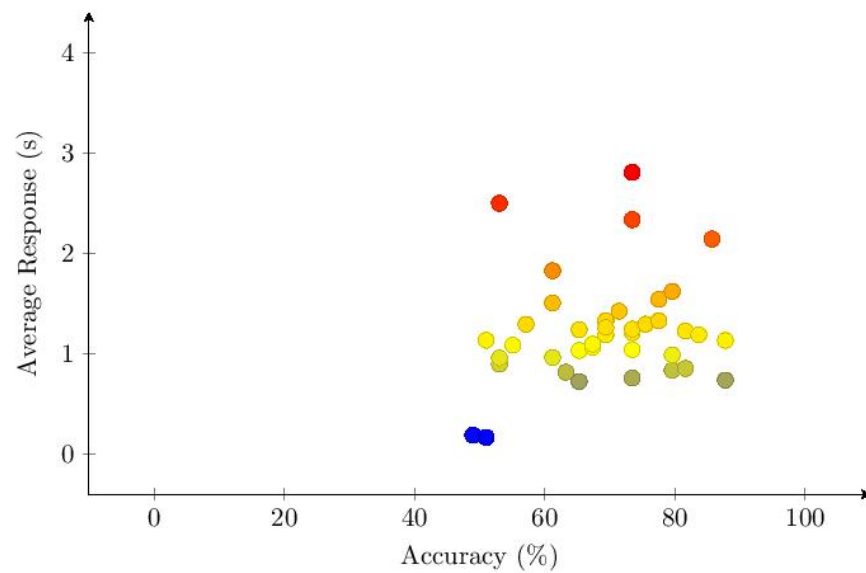
Figure 7: v1.0 Individual Participant Results (Cat 2)

For score comparison, Figure 8 below depicts score distribution with intervals of 5%. Note the primary peak with 14 scores at 85-90%.
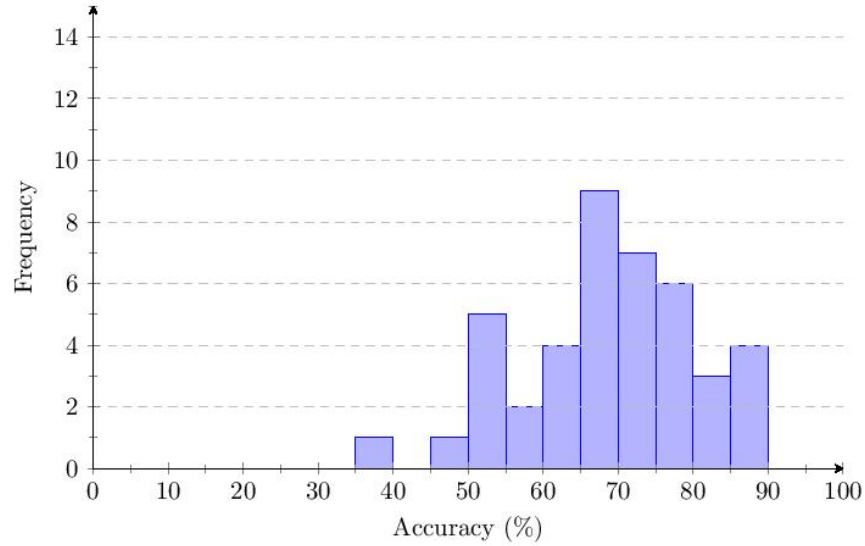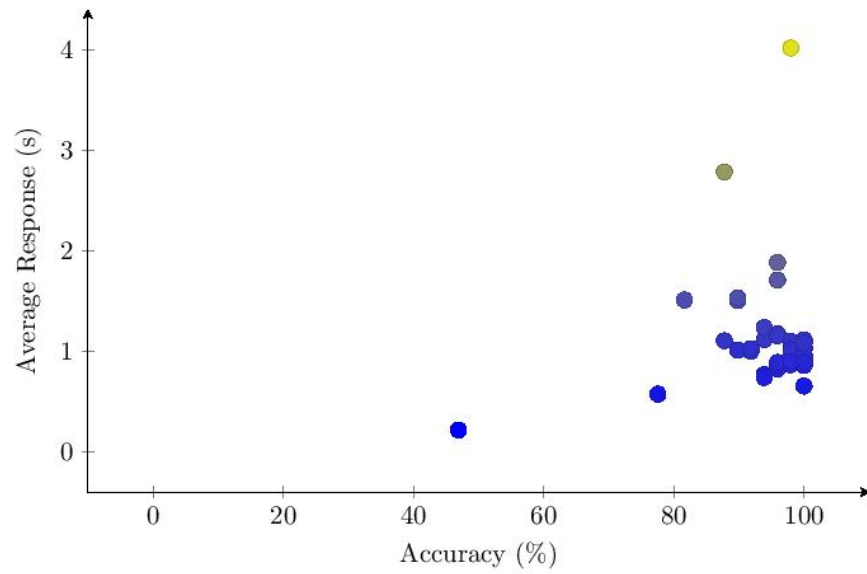

Figure 8: v1.0 Distribution of Scores (Cat 2)

5.3.3 Category 3 Data

Scores for Category 3 averaged 87%, as seen in Figure 4. Figure 9 below compares accuracy to response time. Note how most scores clump around 80-100% and response times around 1-1.5 seconds.

Figure 9: v1.0 Individual Participant Results (Cat 3)

For score comparison, Figure 10 on the next page depicts score distribution with intervals of 5%. Note the primary peak with 14 scores at 85-90%.

Figure 10: v1.0 Distribution of Scores (Cat 3)

5.3.4 Category 4 Data

Scores for Category 4 averaged 69.5%, as seen in Figure 4. Figure 11 below compares accuracy to response time. Note how most scores clump around 50-90% and response times around 1 seconds.

Figure 11: v1.0 Individual Participant Results (Cat 4)

For score comparison, Figure 12 below depicts score distribution with intervals of 5%. Note the primary peak with 9 scores at 65-70%.
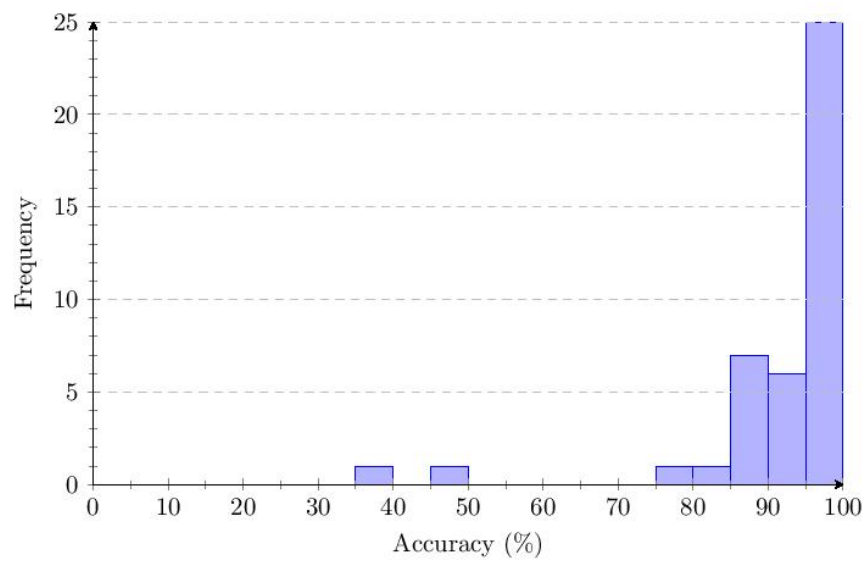

Figure 12: v1.0 Distribution of Scores (Cat 4)

5.3.5   Category 5 Data

Scores for Category 5 averaged 93.6%, as seen in Figure 4. Figure 13 below compares accuracy to response time. Note how most scores clump around 90-100% and response times around 1 seconds.

Figure 13: v1.0 Individual Participant Results (Cat 5)

For score comparison, Figure 14 below depicts score distribution with intervals of 5%. Note the primary peak with 25 scores at 95-100%.

Figure 14: v1.0 Distribution of Scores (Cat 5)

5.3.6 Comparing Categories

Below shows both a scatter plot (Figure 15) and histogram (Figure 16) comparing the results of the five tested categories in v1.0.
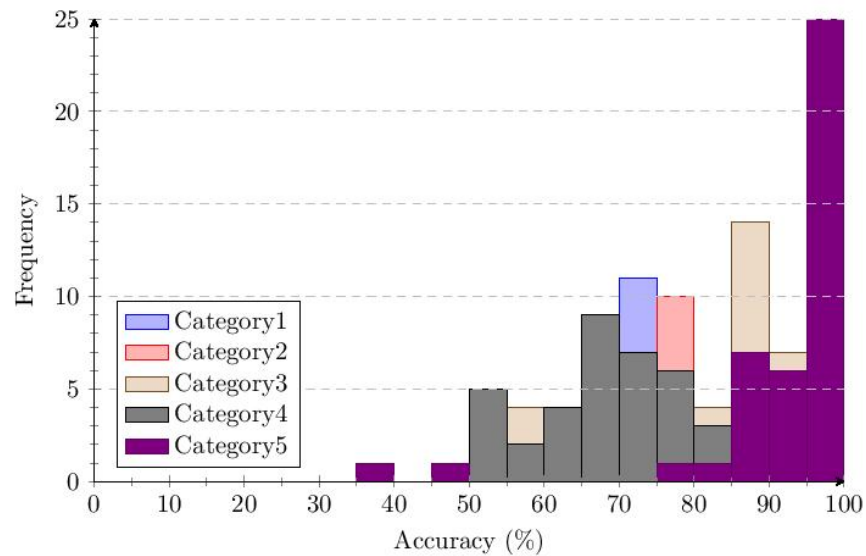

Figure 15: v1.0 Individual Participant Results

Figure 16: v1.0 Distribution of Scores

5.4 v1.65 Methodological Changes and Study Results

In contrast to the first version of the test, version 1.65 included an interstitial screen in-between the testing categories to reduce testing fatigue. The trials gathered per category was also doubled from 50^9 to 100. This means that we suspect the data from v1.65 to be higher quality. However, due to its increased length and a reduced marketing effort, only 28 responses were received.

Additionally, new data was collected from participants in this iteration. Specifically, this included data on the scores and response times by individual question, rather than the just the overall score of v1.0.

Category	Mean (%)	Median (%)	Mode (%)
1	69.86956522	72	75
2	74.60869565	76	79
3	87.65217391	90	91
4	68.04347826	70	70
5	96.26086957	97	96

Figure 17: v1.65 Table of Average Scores

5.4.1 Category 1 Data

Scores for Category 1 averaged 69.9%, as seen in Figure 17. Figure 18 below compares accuracy to response time. Note how most scores clump around 50-80% and response times around 1-2 seconds.

---

[9] Only 49 of which were actually properly logged and sent back to the server for analysis due to a bug.
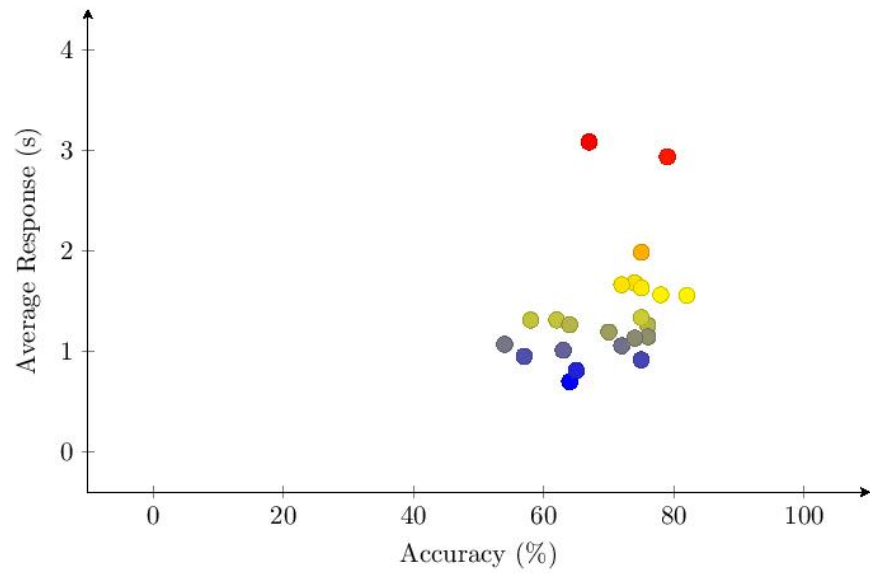
Figure 18: v1.65 Individual Participant Results (Cat 1)

For score comparison, Figure 19 below depicts score distribution with intervals of 5%. Note the primary peak with 8 scores at 75-80%.
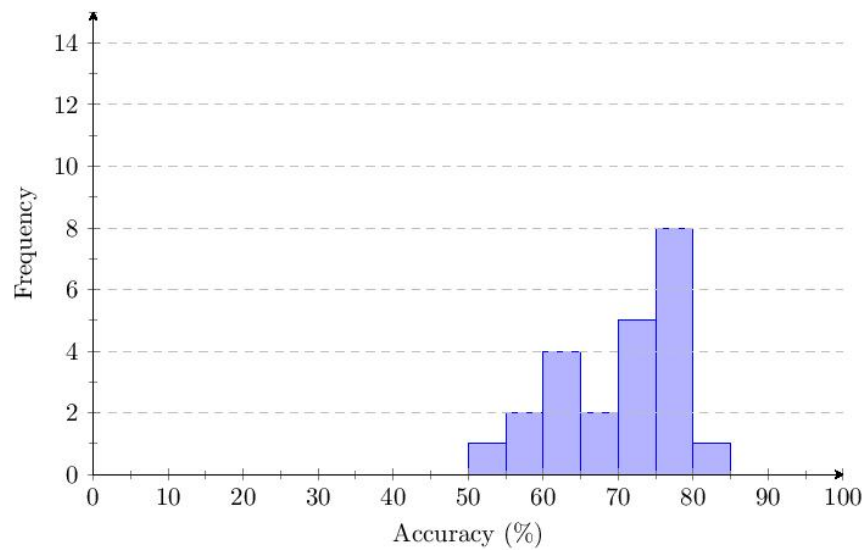

Figure 19: v1.65 Distribution of Scores (Cat 1)

As mentioned earlier, due to the methodological changes, it was possible to graph the average score per question over time. Figure 20 shows a drop of about 50% in accuracy over time for Category 1.
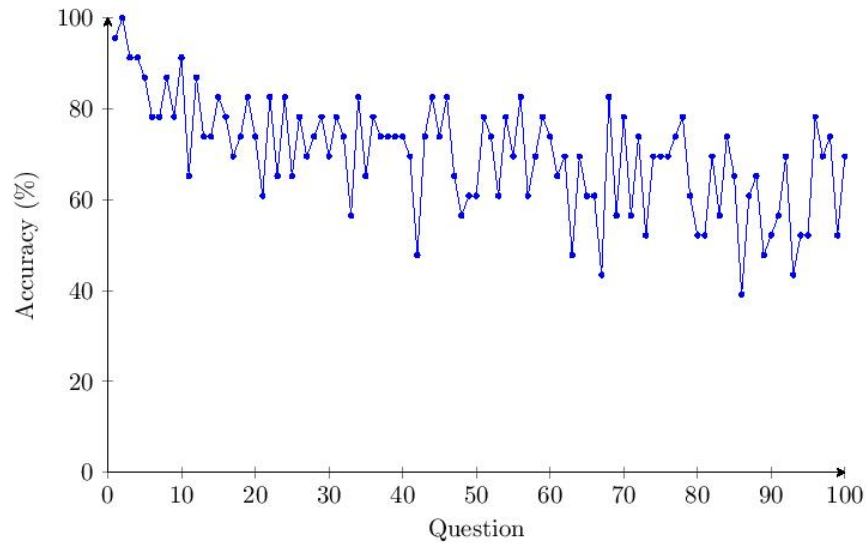

Figure 20: v1.65 Accuracy (%) per Question (Cat 1)

In Figure 21 seen on the following page, there was no noticable trend or changes over time for response time. The large spikes seen in the data could be attributed to participants choosing to spend more time considering their answer for certain questions.

Figure 21: v1.65 Response Time per Question (Cat 1)

5.4.2 Category 2 Data

Scores for Category 2 averaged 74.6%, as seen in Figure 17. Figure 22 below compares accuracy to response time. Note how most scores clump around 60-85% and response times around 1-2 seconds.
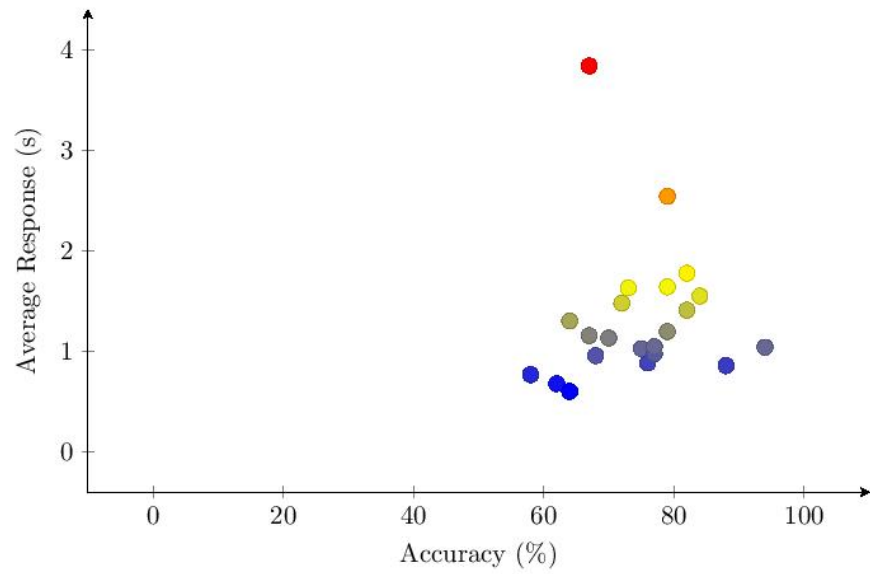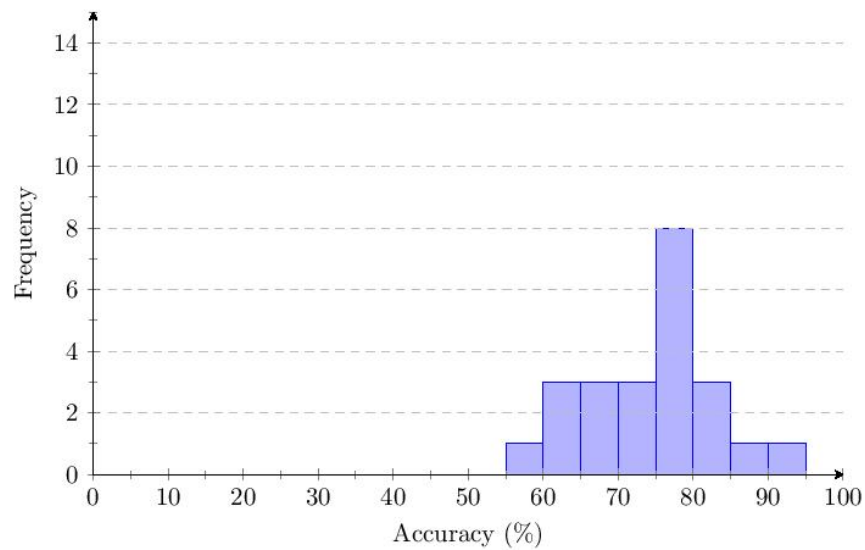
Figure 22: v1.65 Individual Participant Results (Cat 2)

For score comparison, Figure 23 below depicts score distribution with intervals of 5%. Note the primary peak with 8 scores at 75-80%.

Figure 23: v1.65 Distribution of Scores (Cat 2)

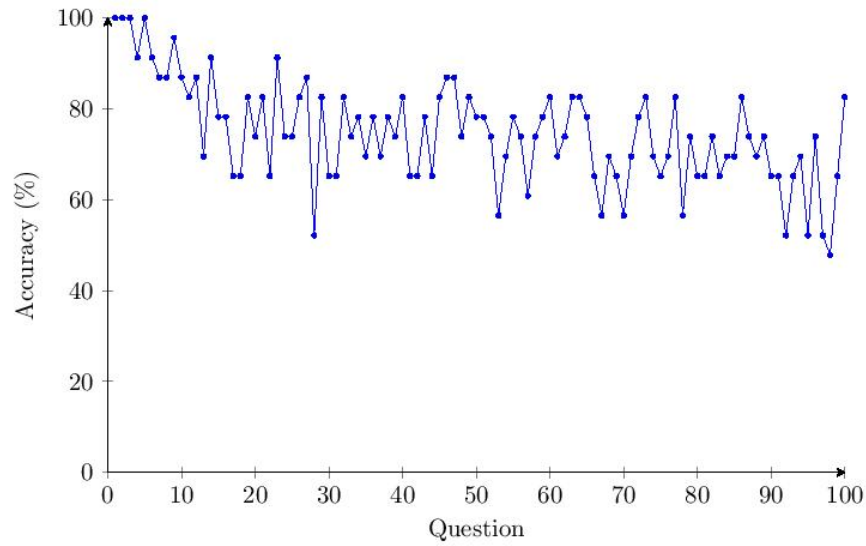Figure 24 shows a drop of about 40% in accuracy over time for Category 2.

Figure 24: v1.65 Accuracy (%) per Question (Cat 2)

Outliers with response times of over 2 minutes were noted under Questions 41 and 98. They have been removed from the presented data in Figure 25 to more accurately show average response times for most participants.
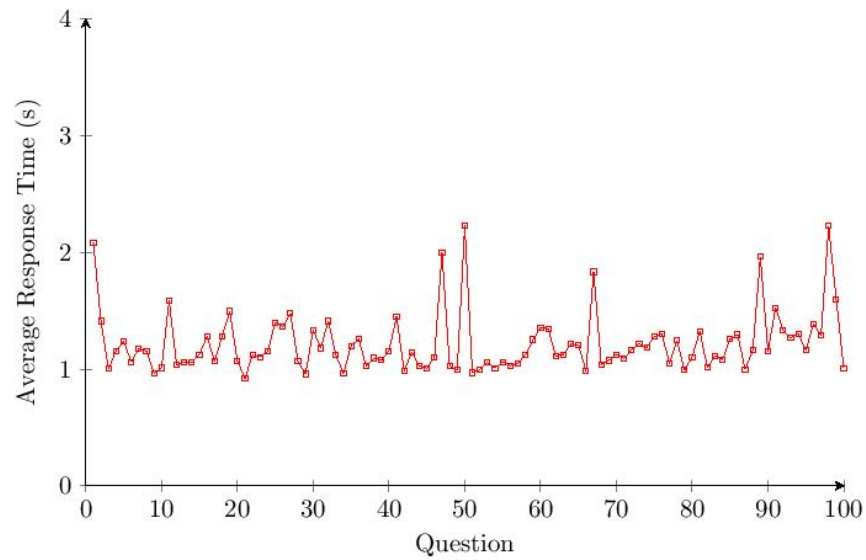

Figure 25: v1.65 Response Time per Question (Cat 2)

5.4.3 Category 3 Data

Scores for Category 3 averaged 87.7%, as seen in Figure 17. Figure 26 below compares accuracy to response time. Note how most scores clump around 80-95% and response times around 1-1.5 seconds.
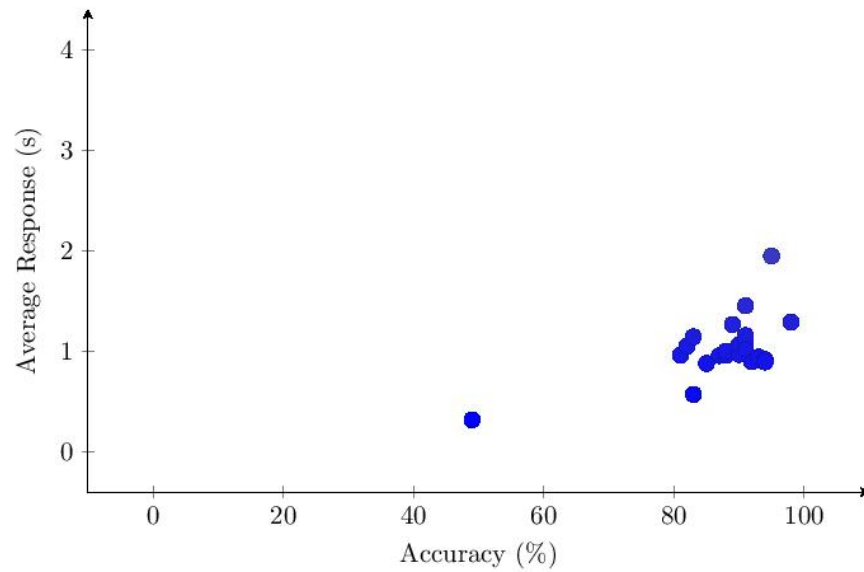

Figure 26: v1.65 Individual Participant Results (Cat 3)

For score comparison, Figure 27 below depicts score distribution with intervals of 5%. Note the primary peak with 11 scores at 90-95%.
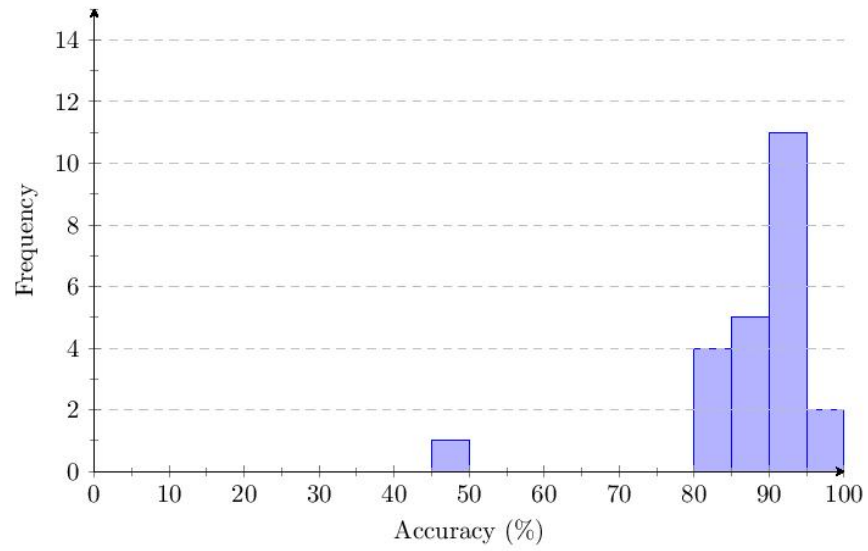
Figure 27: v1.65 Distribution of Scores (Cat 3)

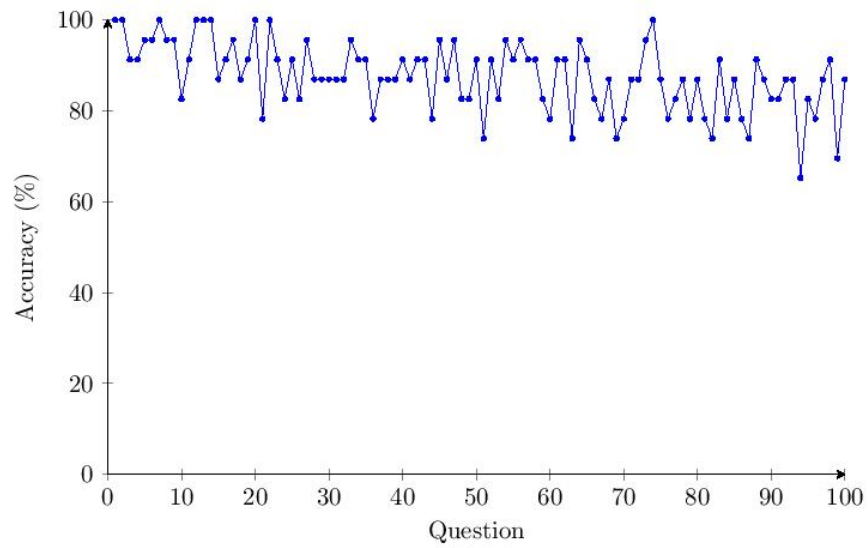Figure 28 shows a drop of about 20% in accuracy over time for Category 3.

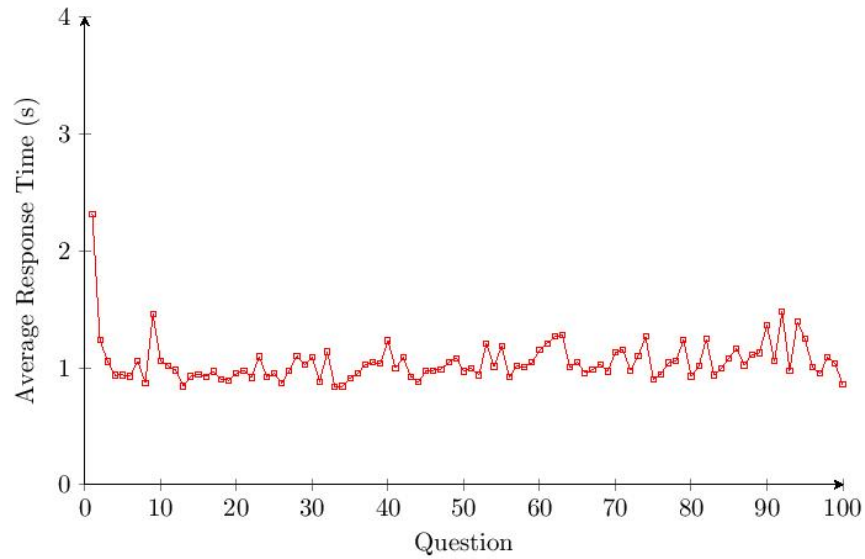Figure 28: v1.65 Accuracy (%) per Question (Cat 3)

Figure 29: v1.65 Response Time per Question (Cat 3)

An outlier with a response time of around 35 minutes was noted under Question 2. It has been removed from the presented data to more accurately show average response times for most participants.

5.4.4   Category 4 Data

Scores for Category 3 averaged 68%, as seen in Figure 17. Figure 30 below compares accuracy to response time. Note how most scores clump around 50-85% and response times around .5-1.5 seconds.
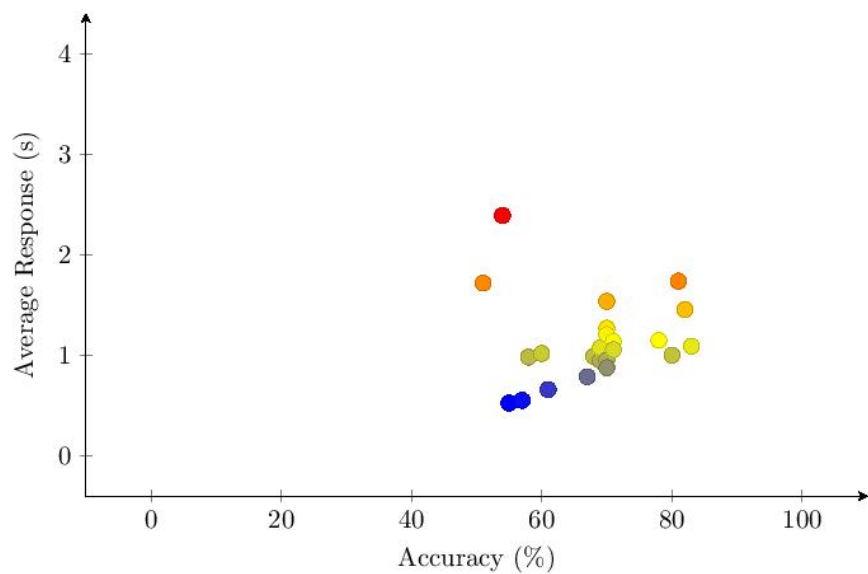
Figure 30: v1.65 Individual Participant Results (Cat 4)

For score comparison, Figure 31 below depicts score distribution with intervals of 5%. Note the primary peak with 7 scores at 70-75%.
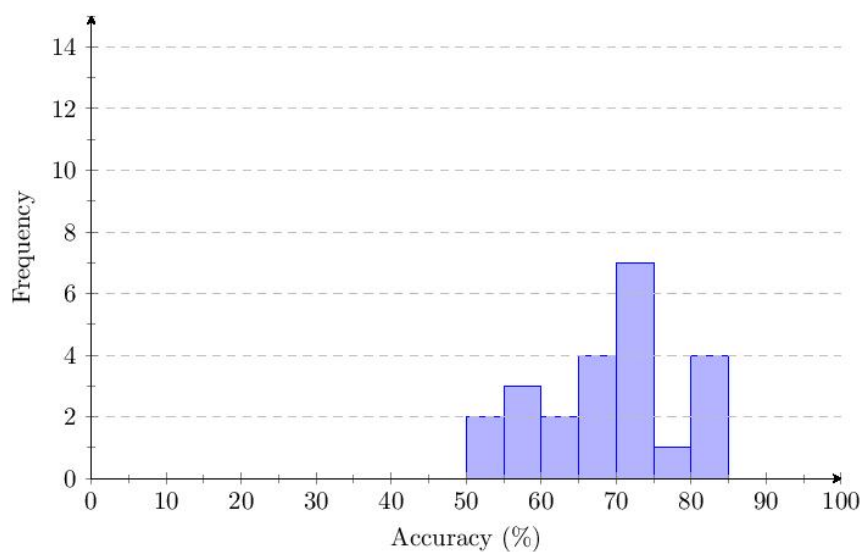

Figure 31: v1.65 Distribution of Scores (Cat 4)

Figure 32 shows a drop of about 40% in accuracy over time for Category 4.
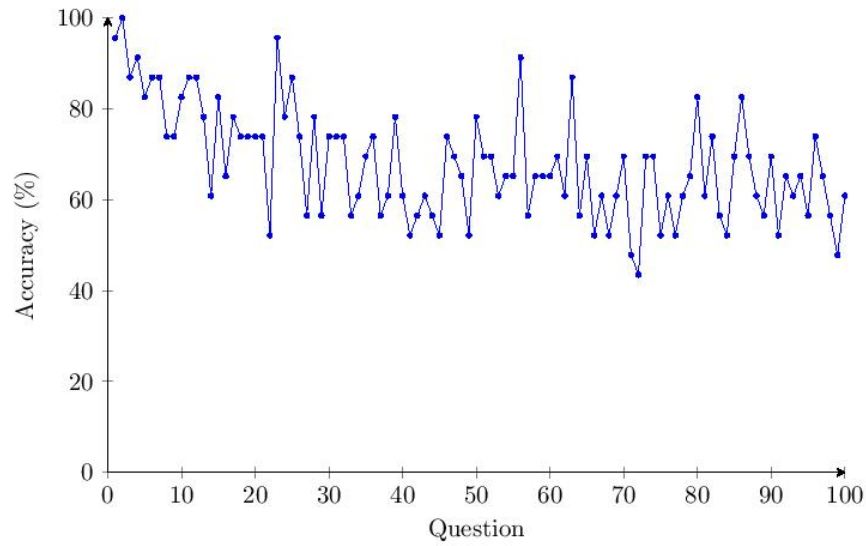
Figure 32: v1.65 Accuracy (%) per Question (Cat 4)

As seen in Figure 33, the response time for Category 4 only increases by a small percentage over time.

Figure 33: v1.65 Response Time per Question (Cat 4)

5.4.5 Category 5 Data

Scores for Category 3 averaged 96.3%, as seen in Figure 17. Figure 34 below compares accuracy to response time. Note how most scores clump around 90-100% and response times around 1 second.

Figure 34: v1.65 Individual Participant Results (Cat 5)

For score comparison, Figure 35 below depicts score distribution with intervals of 5%. Note the primary peak with 7 scores at 95-100%.

Figure 35: v1.65 Distribution of Scores (Cat 5)

Unlike previous categories, Figure 36 shows no drop in accuracy over time for Category 5.

Figure 36: v1.65 Accuracy (%) per Question (Cat 5)

As seen in Figure 37, the response time for Category 7 only increases by a small
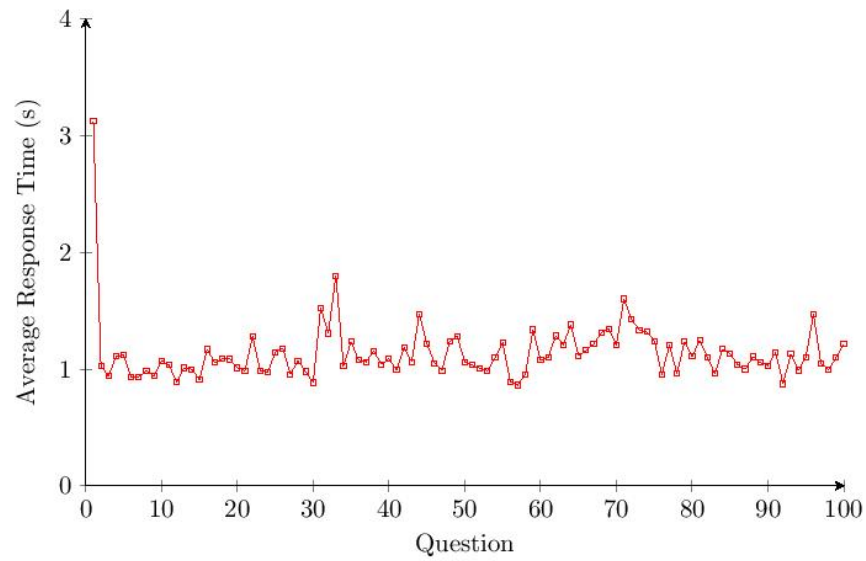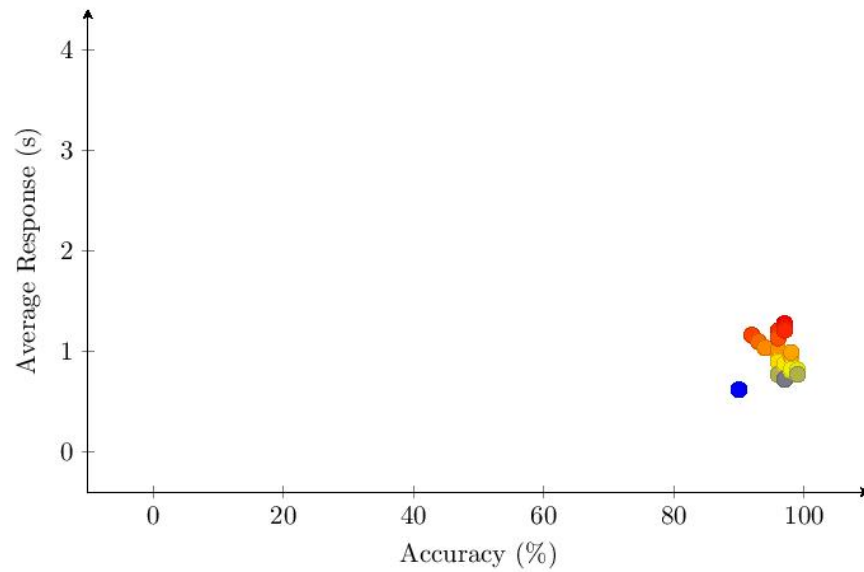
percentage over time.

Figure 37: v1.65 Response Time per Question (Cat 5)

5.4.6 Comparing Categories

Below shows both a scatter plot (Figure 38) and histogram (Figure 39) comparing the results of the five tested categories in v1.65.

Figure 38: v1.65 Individual Participant Results

Figure 39: v1.65 Distribution of Scores

30

5.5 v1.X Meta-Analysis of Results

For some statistics, it may be of interest if not use to analyze the differences between the results of version 1.0 and version 1.65, as well as to c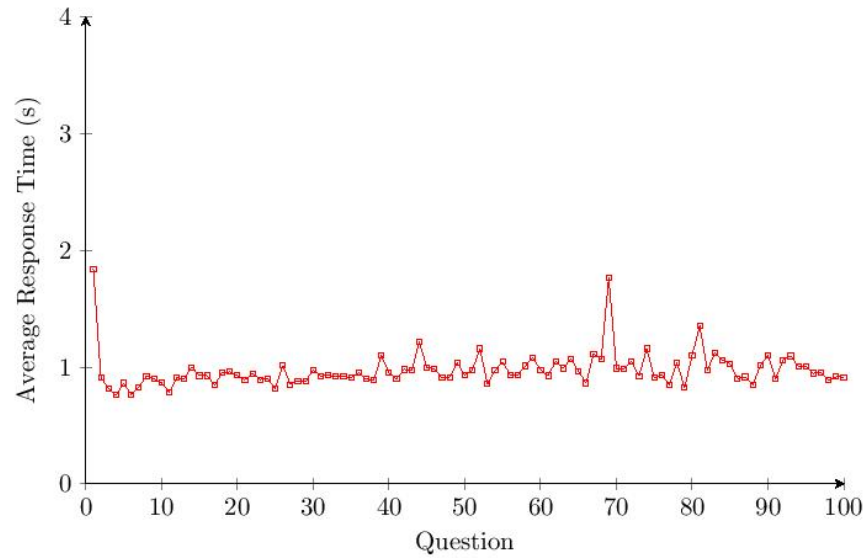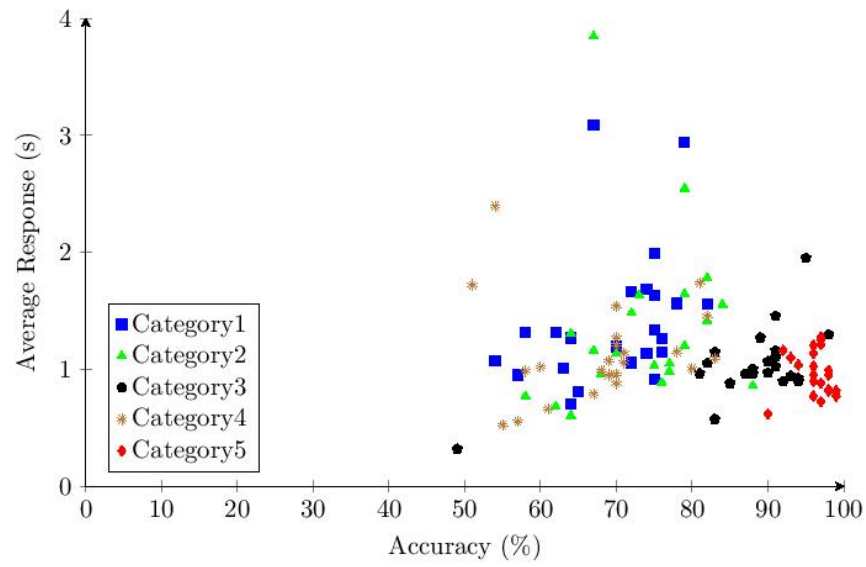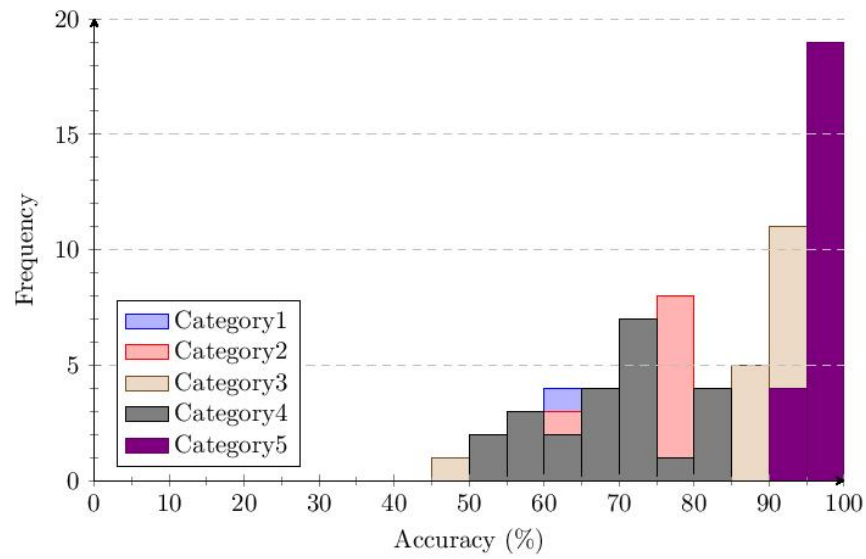ombine both datasets into one to get a larger amount of data to analyze when dealing with calculations or graphing methods that could pay excessive respect to outliers when the pool of total data points is too few.

Category	Mean (%)	Median (%)	Mode (%)
1	72.78156948	73.46939	73.46939
2	76.90816389	77.55102	85.714287
3	87.36734736	89.79592	97.959185
4	69.03220681	69.69388	73.46939
5	94.60650572	96	100

Figure 40: Combined v1.0 & v1.65 Table of Average Scores

5.5.1 Category 1 Data

Figure 41 combines the Category 1 scores of v1.0 and v1.65 together for greater comparison. Note how most scores clump around 60-90% and response times around 1-2 seconds.

Figure 41: Combined (v1.0 & v1.65) Individual Participant Results (Cat 1)

5.5.2 Category 2 Data

Figure 42 combines the Category 2 scores of v1.0 and v1.65 together for greater comparison. Note how most scores clump around 60-90% and response times around 1-2 seconds.

Figure 42: Combined (v1.0 & v1.65) Individual Participant Results (Cat 2)

5.5.3 Category 3 Data

Figure 43 combines the Category 3 scores of v1.0 and v1.65 together for greater comparison. Note how most scores clump around 80-100% and response times around 1-2 seconds.

Figure 43: Combined (v1.0 & v1.65) Individual Participant Results (Cat 3)

5.5.4 Category 4 Data

Figure 44 combines the Category 4 scores of v1.0 and v1.65 together for greater comparison. Note how most scores clump around 50-85% and response times around 1-2 seconds.

Figure 44: Combined (v1.0 & v1.65) Individual Participant Results (Cat 4)

5.5.5 Category 5 Data

Figure 45 combines the Category 5 scores of v1.0 and v1.65 together for greater comparison. Note how most scores clump around 90-100% and response times around 1 second.
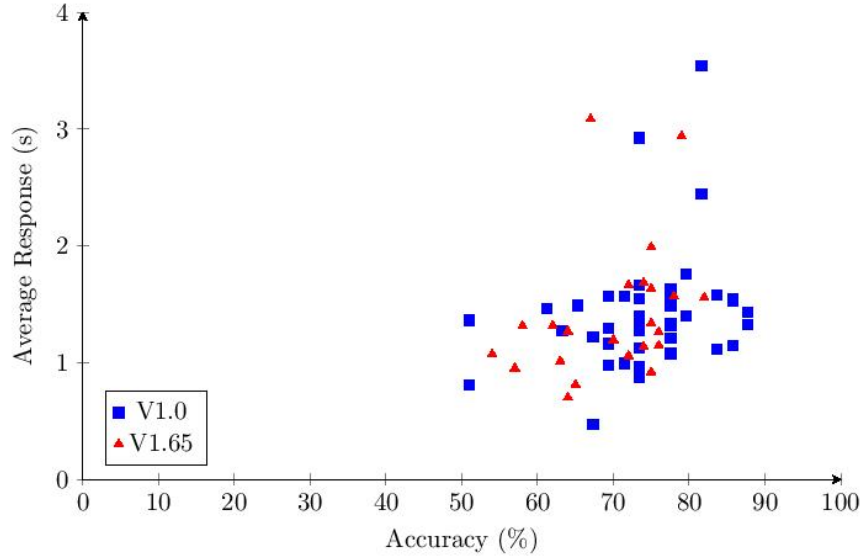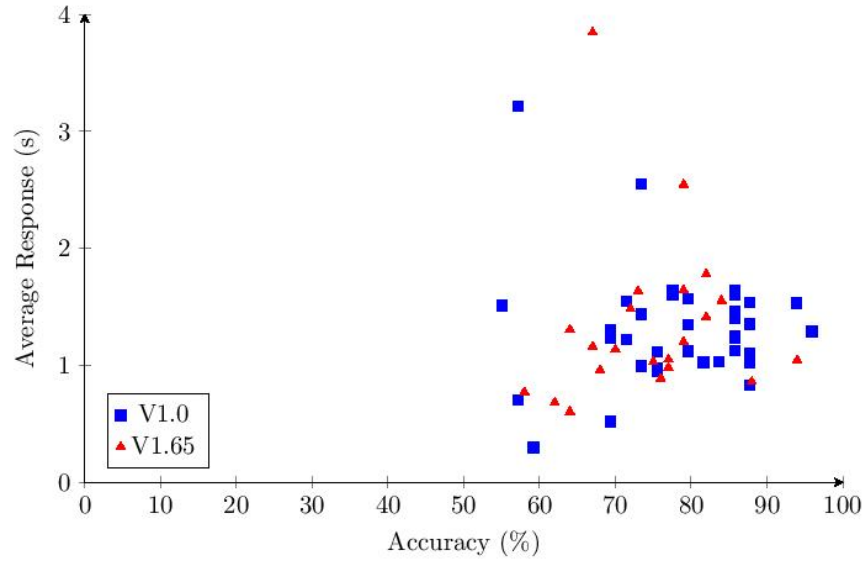
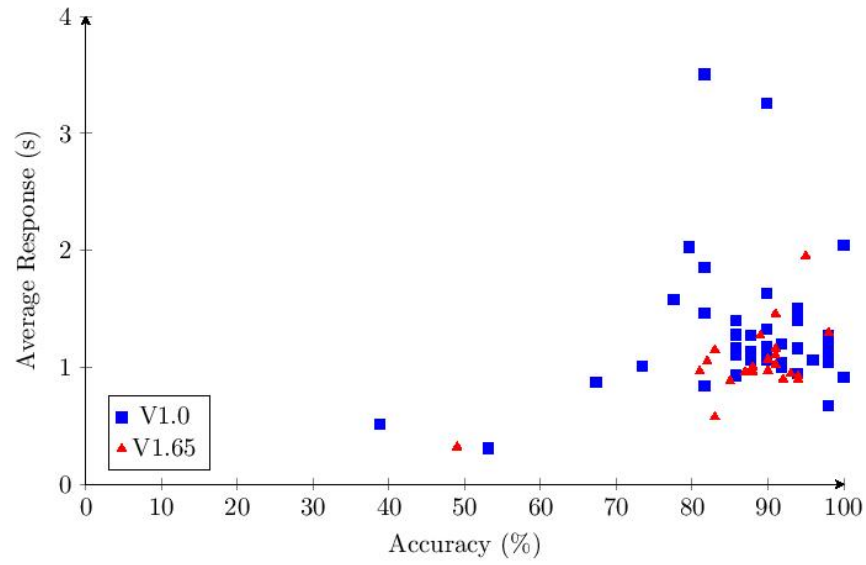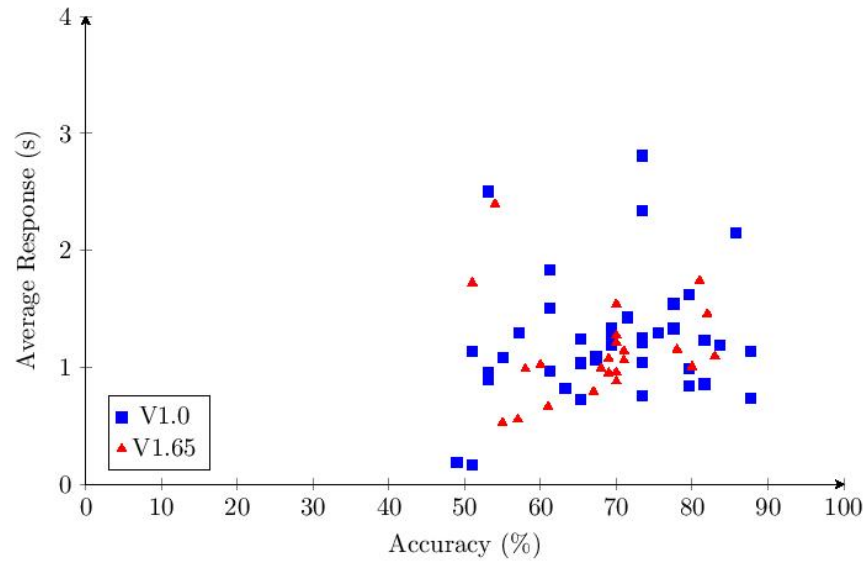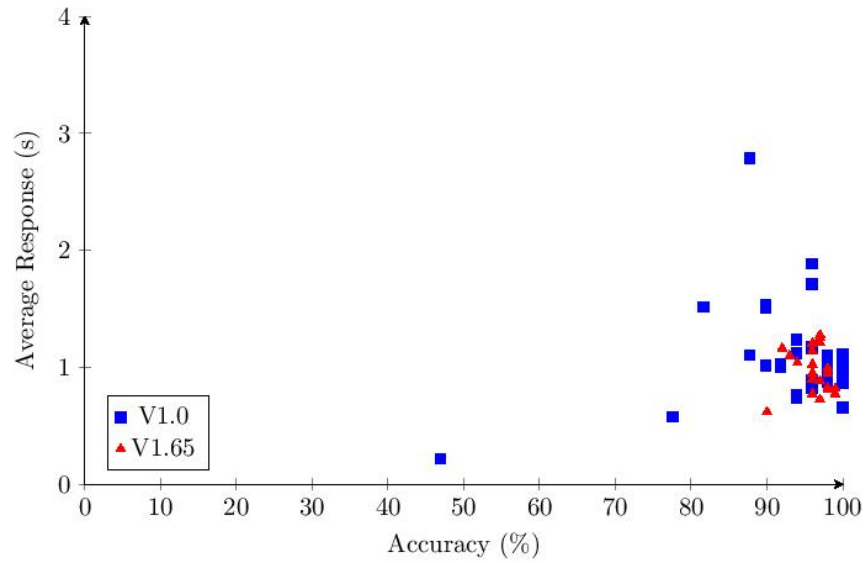Figure 45: Combined (v1.0 & v1.65) Individual Participant Results (Cat 5)

5.5.6 Category Comparison

Below shows both a scatter plot (Figure 46) and histogram (Figure 47) comparing the results of the five tested categories across both v1.0 and v1.65.
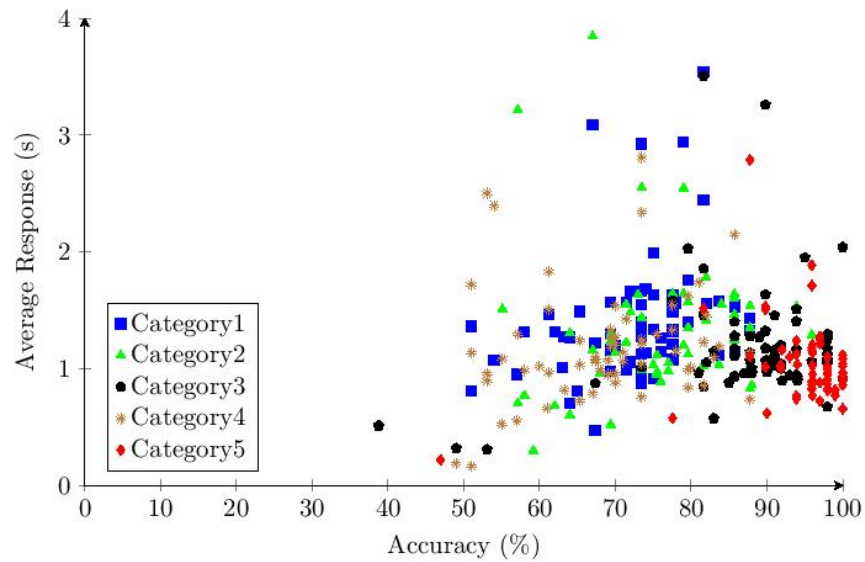
Figure 46: Combined (v1.0 & v1.65) Individual Participant Results

Figure 47: Combined (v1.0 & v1.65) Individual Participant Results

5.6 v2.0 (Ongoing)

The version 2.0 of the study is live as of 1 October, 2025, which is the informal publication date of this pre-report. The study platform is being released to the public simultaneous to this pre-report, in the hopes that their simultaneous publication and referencing each other may aid in improving interest in this research project and study.

The changes to v2.0 were selected to improve on and account for the shortcomings in the prior versions without taking too long to develop.

5.6.1 Request for Comment

We held an informal Request for Comment (RfC) to gather community feedback on some proposed changes to the survey, specifically to find new categories to add for further testing, to gather feedback on categories to remove, and to gather feedback on the length of each category in trials.

This RfC resulted in the recommendations to add a category in the format of `Adjective+Noun`, a category which incorporates color associations to assess the usefulness of avatar visual information in aiding users to identify and distinguish other users, beyond the raw text of their username. There was also, surprisingly, broad support for increasing the amount of trials per category, lengthening the total survey time.

5.6.2 Decisions of Sprinkler Research Group on v2.0

With aid of the recommendations from the RfC, we decided to make the following changes to v2.0 from v1.65, broadly following the RfC recommendations. We determined that it was a good idea to add `Adjective+Noun` as a category, and this is present in v2.0.

We could not find a practical method to incorporate the color association idea, although we think it holds merit, that wouldn't double the total test length with re-trials of all categories and without introducing potential confounding, so for this reason it has been excluded. We agree that v2.0 will be able to be more deeply analyzed with longer categories, so we have decided to double the number of trials per category to 200. It is very possible that this will decrease the number of respondents who fill out the entire survey, but the responses that we do get will be more accurate.

The RfC also recommended the removal of Category 1 due to its similarity with Category 4, and we accepted this proposal. We also chose to remove Category 2 due to the insufficient total number of possible usernames (Only 2600 unique names)[10] which would not be useful or scalable in real digital environments with an unknown quantity of users.

Additionally, we decided to replace Category 3 with a new category of our own making, due to Category 3's similarly small total namespace (2400 unique names)[11], and its similarity to the concept of this new category. That new category is the output of a custom script running a Markov chain trained on english words, with outputs filtered to ensure that no single result is one letter away from an english word nor two letters away from an inappropriate word.[12] This technique has been demonstrated to produce at least 50,000 unique names, and can likely produce far more, so it is sufficient for the purposes of this survey.

Category 4 was kept in order to have a metric to directly compare to the results of Category 1 and Category 4 from v1.0 and v1.65, even though the total range of possible values is evidently a mere one-thousand. The reason for choosing to keep Category 4 over categories 3 or 1 is simply because we determined Category 4 the most similar to both categories 1 and 4 in versions 1.0 and 1.65, meaning it gives us the largest surface to compare against between study versions for a single category. Finally, to round the number of categories off to a nice and even 5, we added a category that samples from a dataset of 6,864,472 `Minecraft` usernames in a list online by Yunowe (2022). We checked that every name in this dataset was unique and added it as Category 8.

---

[10]This figure is obtained from the calculation 26*10*10

[11]This figure is obtained from the calculation 20*6*20

[12]Its source will be included with the final version of the report once the data from v2.0 has been gathered.

5.6.3   Final v2.0 spec

The final spec for v2.0 consists of a new category list. [13]

1. Category 7: format `Markov output` such as `Scassle`

2. Category 6: format `Adjective+Noun` such as `Medium Sprinkler`

3. Category 8: format `Minecraft username` such as `w4nderlost`

4. Category 4: format `XXX` such as `779`

5. Category 5: format `NOUN` such as `HOUSE`

Including the other non-categorical changes, here is the total summary of changes made to v2.0:

1. Category 1 has been replaced with Category 6.

2. Category 3 has been replaced with Category 7.

3. Category 2 has been replaced with Category 8.

4. The number of trials recorded per survey has been changed to 201 (double plus a "gimme" at the beginning)

5. Respondents will no longer receive immediate feedback on right or wrong answers in order to reduce pressure or stress and avoid affecting performance

6. Category 5 had previously too small a data set of nouns to draw from. We increased this from around 200 to around 7000.

5.6.4   Methodological Improvements

This study could be further improved upon by testing the participant's long-term memory as well. We chose not to do this to prevent having to force participants to provide de-anonymizing identifying information about themselves. However, there is likely interesting data to be garnered here. It is also worth further study to compare the potential for color associations or avatar appearances to aid in recognition and differentiation of users when there is no visible username, the username itself is color-coded, or when custom colors simply are characters in the username itself.

---

[13]Do note that the categories in this list have non-sequential numerical IDs due to our choice to not discard the older data. In order to ensure that this data is still able to be easily incorporated into final analysis of the study, including data from all versions, the new added categories have been given the next available number.

5.6.5 Category Updates

6 Next steps

We refuse to draw conclusions in this pre-report until the data gathering and analysis of 2.0 has been completed. Due to the relatively low sample size and some potential systematic flaws in the study design, we cannot have full confidence that the patterns the data suggests are fully determined by and descriptive of the efficiencies of the underlying models. Readers interested in a more in-depth and data-based conclusion should subscribe to our future research to see our final report.

If any reader has information, questions, or suggestions for us, then such a reader is advised to contact us on Discord @fyrine or @loodledoodles.

References

Holm, F. (2025a). *Sprinkler nametag test.* Retrieved September 12, 2025, from https://sprinkler-nametag-test.xn--frine-qva.li

Holm, F. (2025b, September 11). *Trial 2 session 0 Minecraft civilization experiment (ShotRush)* [Video]. YouTube. Retrieved September 11, 2025, from https://www.youtube.com/watch?v=ZbzCPjFoLeQ

KrazyKTV. (2025, July 31). *The fall of a Minecraft civilization — CivLabs* [Video]. YouTube. Retrieved September 9, 2025, from https://www.youtube.com/watch?v=anFQXIOD4rg

ShotRush. (2025a, July 8). *Fixing the problem with simulating society in Minecraft* [Video]. YouTube. Retrieved September 9, 2025, from https://www.youtube.com/watch?v=Z76QnXLxT0M

ShotRush. (2025b, August 1). *We did a true civilization experiment.* [Video]. YouTube. Retrieved September 9, 2025, from https://www.youtube.com/watch?v=SU1BUyw9ETs

Yunowe. (2022, July). *Minecraft username list (6,800,000+)* [Thread post with MediaFire download link to username list]. Retrieved September 30, 2025, from https://hypixel.net/threads/minecraft-username-list-6-800-000.5032332/