Roll20 Usability Test Results

Prepared by David Clipson, 8/21/2022

Executive Summary

This study evaluates the usability of a new prototype for the Roll20 website. Remote usability tests were conducted with six participants in August 2022. All participants have existing knowledge of Dungeons & Dragons, have played D&D online in the past, are male, and are between the ages of 25 and 45. Each participant was asked three background questions, performed eight tasks using a Figma prototype, and responded to a System Usability Scale (SUS) survey. Each session took between 20 and 30 minutes. Data was collected regarding task success, time-on-task, and perceived ease-of-use.

Key findings

- Overall, the prototype appears to be a significant improvement over the current system.
- All participants completed all tasks successfully without assistance, yielding a 100% task success rate.
- Most tasks were completed in under 10 seconds, but there were some notable outliers.
- Users reported that most tasks were very easy, with an average Single Ease Question (SEQ) score of 6.5 out of 7.
- Users reported that the system is highly usable, with an average System Usability Scale (SUS) score of 85.8%.
- Users needed more time and reported lower SEQ values for the first task, which is largely attributed to the need to become oriented with the system's overall structure.
- Users needed more time and reported lower SEQ values for the Dice Roller task, which is a more complex task that also involves learning a new part of the system.
- Users would benefit from clearer visual feedback when an action has been successfully performed, particularly when casting spells.
- Users need access to more information about spell effects when casting spells.
- Some parts of the prototype need clearer affordances. In particular, some buttons are missing hover effects and/or tooltips.
- Users would prefer more consistency between the Attacks and Spells tabs.

Table of Contents

Executive Summary	1
Table of Contents	2
Introduction	3
Research Goals	3
Methods Participants Scenario and Tasks Test Procedure Data Collected Metrics	4 4 5 6
Results Question 1: Are users able to perform dice rolls successfully? Question 2: Are users able to perform dice rolls efficiently? Question 3: Do users find rolling dice to be easy? Additional Results	7 7 8 10
Discussion	11
Recommendations	12
Appendix A: Consent Form for Usability Participants	14
Appendix B: Pilot Study Test Script	

Introduction

Roll20 is a website that allows players of tabletop roleplaying games to unite online. It includes a complete set of tools needed to play many games, including Dungeons & Dragons 5th Edition. D&D 5E is by far the most popular game currently supported by the site, representing over 50% of all user campaigns. D&D 5E has seen a significant rise in popularity in recent years, partially due to pop culture sensations such as Stranger Things and Critical Role. For this reason, providing an excellent user experience for D&D 5E players is important for Roll20's future.

In 2021 I conducted user interviews, surveys, and usability tests to identify the best and worst aspects of Roll20's user experience for D&D 5E players. In a survey of 112 users, I established baseline satisfaction scores for a range of product areas:



Roll20 Satisfaction Survey Results

From this research I concluded that many users found Roll20's support for performing dice rolls to be neither usable nor satisfying. Based on these findings, I created mockups and prototypes of possible solutions to improve the experience of rolling dice within Roll20. I created redesigned versions of both the Dice Roller and Character Sheet modules within Roll20, which function together to meet the user's dice rolling needs.

Research Goals

This study attempts to address the following questions regarding my prototype:

- 1. Are users able to perform dice rolls successfully?
 - It is critically important that users are able to perform a variety of dice rolls with accuracy and confidence.
- 2. Are users able to perform dice rolls efficiently?
 - Rolls happen frequently during a D&D session. Therefore it is important that users can perform rolls relatively quickly.
- 3. Do users find rolling dice to be easy?
 - In order for Roll20 to stay competitive, users must find it easy to use.

Methods

Participants

For this study, I planned to include between 5 and 7 participants. In total 6 participants were included in the study.

The following requirements were applied when selecting participants:

- Participants must be familiar with D&D 5E.
 - D&D has a lot of rules and terminology. Users who are unfamiliar with standard game concepts may have trouble performing tasks as a result.
 - Participants should have played at least one D&D 5E session to be included in the study.
- Participants must have access to a desktop or laptop computer with internet and a microphone.
 - Tests will be conducted remotely. Participants will need to have access to the necessary technology to facilitate the test.
 - Participants must indicate they have the required technology to be included.
- Participants must be willing to consent to audio and screen recording during the session.
 - Audio and screen recordings will be reviewed after the live sessions to facilitate deeper analysis.
 - Participants must indicate willingness to consent to being recorded to participate.
- Participants must be at least 18 years of age.
 - The target market for Roll20 is primarily adults. Additionally, users under 18 would not be able to give consent to be recorded.
 - Participants must indicate they are at least 18 years of age to participate.
- Participants must be fluent in English.
 - Studies will be performed in English.
 - Participants must indicate they are fluent in English to participate.

Participants were recruited from my personal network of D&D players. All participants have played D&D 5th Edition online before, are male, and are between 25 and 45 years old.

Scenario and Tasks

Participants were asked to imagine a scenario in which they were playing a game of D&D online. Participants were asked to complete a series of tasks that all involve rolling dice. The tasks were presented in the same sequence for each participant.

Users were asked to perform the following sequence of tasks:

- Make a Perception roll.
 - This is a common roll that can be found in the "Skills" section.
- Make an Initiative roll.

- This is a common roll that can be found in the "Stats and Saves" section.
- Make an attack and damage rolls for an attack spell.
 - These are common rolls that can be found in the "Spells" section.
- Cast a spell with neither attack nor damage rolls (e.g., Shield).
 - This is also in the "Spells" section but requires a different user interaction.
- Make an attack and damage rolls for a weapon.
 - These are common rolls in the "Actions" section.
- Make an ad hoc roll.
 - This tests the Dice Roller control, which is separate from the character sheet.

For each task, I recorded data related to the success, efficiency, and perceived ease of these tasks.

Test Procedure

The study was conducted via remote usability tests. Participants and the moderator connected using their own computers. Audio communication and screen sharing was facilitated using Discord. All sessions were recorded (video and audio) using OBS Studio.

At the start of each session, users were asked a few background questions which may provide insights when analyzing results. Users were provided with an interactive Figma prototype so they could perform tasks directly. For each task, the moderator narrated the task verbally and instructed the participant to start. Users were asked to perform the task and state when they believed they had completed the task.

Tasks were considered successful if they were completed accurately without any help from the moderator. If the user failed to complete the correct tasks or needed to ask the moderator for assistance, the task was considered a failure.

After each task, the moderator asked the user to rate that task using the Single Ease Question (SEQ). I used the 7-point version of the questionnaire, with "1" representing "very difficult" and "7" representing "very easy." The user was also asked whether they have any questions or feedback about the task they just performed.

After all tasks were attempted, the moderator asked for any additional questions or comments the participant may wish to share. The moderator then sent the participant a link to the System Usability Scale (SUS) questionnaire, asked the participant to fill out the survey as the final step in the test, and thanked the participant for their time. The questionnaire included a free response question as well to collect any additional feedback.

After each session, the recordings were reviewed to transcribe quotes and to measure time-on-task for each task. Time-on-task was measured as the time between when the moderator instructed the participant to start and the time when the user stated they had completed the task.

Data Collected

Background data for each participant was collected at the beginning of the session:

- How many hours have you spent playing D&D 5th Edition in your lifetime?
- How often have you played D&D online within the past 12 months?
- Which category includes your current age?

The usability of the system was measured using the following metrics:

- Task success rate
- Time-on-task
- Single Easy Question (SEQ) scores for each task
- System Usability Scale (SUS) score for the system

All task-based measurements were taken by the moderator while reviewing the video recordings of the sessions.

- Task success was recorded using a binary scale (pass or fail) on a per-task basis.
- Time-on-task was measured on a per-task basis by comparing the timestamps between when the moderator said "start" and when the participant provided a verbal indication of being done.
- The SEQ was asked after each task using a 7-point scale.

The System Usability Scale (SUS) questionnaire was sent to the participant using Qualtrics. SUS scores were transcribed manually and then computed using a spreadsheet. Each question was scored on a scale of 0-6. The total score was multiplied by 1.666 to calculate a percentage value.

Metrics

The research questions, metrics, and data analysis for this study are summarized in the following table:

Research Question	Metrics & Data	Data Analysis
1. Are users able to perform dice rolls successfully?	Metric: Task success rate Data: Binary success per task	Mean, Standard Deviation
2. Are users able to perform dice rolls efficiently?	Metric: Time-on-task Data: Time in seconds per task	Mean, Geometric Mean, Standard Deviation
3. Do users find rolling dice to be easy?	Metrics: Perceived ease and usability Data: SEQ scores, SUS scores	Mean, Standard Deviation, comparison to industry standard

Results

Question 1: Are users able to perform dice rolls successfully?

All participants were successful in completing all tasks, yielding a 100% success rate. Participants did not make any errors nor did they require any assistance in order to complete the tasks. Based on these results, I can confidently say that users were able to perform dice rolls successfully.

Question 2: Are users able to perform dice rolls efficiently?

In order to understand whether users are able to complete tasks efficiently, I recorded and analyzed the average time-on-task for each of the tasks:





Overall these results seem to indicate that users are able to learn and use the interface reasonably quickly. Most tasks took under 10 seconds on average for users to complete, which is good given that users were learning this interface for the first time. However, there are some notable outliers, which will be addressed in the discussion section below.

Question 3: Do users find rolling dice to be easy?

We can look at Single Ease Question (SEQ) and System Usability Scale (SUS) results to understand how participants rated the system in terms of easiness and usability.

SEQ was recorded on a per-task basis using a scale of 1-7. The results are summarized below:



As shown above, several tasks scored an SEQ of 7 across all participants, which is the highest possible score. Most tasks scored an SEQ well above 5.5, which is the industry average value for SEQ. It is notable that the final task (Task 8) had the lowest average SEQ score by a significant margin and also had the greatest amount of variance among participants.

All participants filled out a modified System Usability Scale (SUS) at the end of the session. The results of this survey are summarized below:



Overall, the average SUS score was 85.8% with a standard deviation of 4.5%. According to the Usability.gov website, a score above 68% is considered above average. Based on this point of reference, 85.8% is significantly above average and seems to indicate that users rated the usability of the system to be very good.

Additional Results

I noticed that the longest tasks also had lower average SEQ scores. Based on this observation I hypothesized that there may be a correlation between time-on-task and SEQ score. I performed a linear regression analysis between these two metrics to see if a correlation really exists.

SEQ score is a subjective value ranging from 1 to 7 which is chosen by the participant. It's a discrete scale using integer values. In this test, only one SEQ score was below 5. All other SEQ values fell between 5 and 7 (inclusive).

Time-on-task represents the amount of time it took for the participant to complete the task and verbally indicate they are done. These values represent real seconds and are accurate within 0.1 seconds.



Based on the available data, there does seem to be some correlation between SEQ scores and time-on-task in seconds. Intuitively this makes sense, as tasks that take longer to complete may be perceived as being more difficult.

Discussion

Overall, I feel this study was successful and yielded useful, meaningful results. However, the data reveals some potential issues with both the prototype and the testing methodology that are worth discussing.

Task 1 took significantly longer than most other tasks. Based on my personal observations and what users reported verbally, I attribute this difference to users needing time to become oriented and getting used to the lagginess of the Figma prototype. In retrospect, it is not surprising that this task took longer due to these factors. However, the presence of these confounding factors makes it hard to understand exactly how easy or time-consuming the task itself would be if it were not the first task. One possible solution to this problem is to change the order of tasks across participants so there isn't one task that's always first. Alternatively, a similar task could be added later in the scenario so the task can be measured without the need for users to become oriented.

Task 3 also took an above-average amount of time and had one of the lowest average SEQ scores. Many participants struggled with the fact that there are actually two possible solutions for this task. Not wanting to make a mistake caused participants to hesitate and second-guess

themselves. In terms of testing methodology, this could possibly be addressed by rephrasing the task in a way that directs the participant to one of the solutions in particular. Alternatively, I may need to revisit the design itself to eliminate this potential source of confusion.

Task 8 was the biggest outlier, both in terms of time-on-task and SEQ. This result is not surprising, as Task 8 was significantly more complex than other tasks and involved an entirely different part of the interface. Whereas Tasks 1-7 all focus on the character sheet sidebar, Task 8 focused on the new Dice Roller control. The task itself also involves more clicks than the previous tasks.

Because Task 8 was the first (and only) task involving the Dice Roller, it required users to become reoriented and learn a new part of the interface, similar to Task 1. As a result, this task was not very successful in measuring how efficient the system is. Instead, the task was more successful in measuring how long it takes to discover and learn the system. To address this in the future, additional tasks involving the dice roller should be included. The first Dice Roller task will provide insight regarding how long it takes to use the system for the first time, whereas subsequent Dice Roller tasks will provide a more accurate representation of how efficient the system is after the initial learning curve.

Overall, I believe the issues that affected Task 1 and Task 8 likely affected all of the tasks to a lesser degree. In particular, each of these tasks were unique and required the user to learn something new on the fly. None of the tasks required users to interact with parts of the interface that they'd already learned. Consequently, my time-on-task data is probably more closely connected with the concepts of discoverability and learnability as opposed to the intended goal of measuring efficiency. To improve this in the future I believe it would help to ask users to perform the same task more than once. For example, I could simply ask the users to repeat the exact same sequence of tasks twice within a single testing session. The first run through the sequence would primarily reveal how long it takes to learn the system, whereas the second run through the tasks would show how quickly a user can perform the tasks if they are already familiar with the system. Given that the second run would likely be very fast, it would definitely be feasible without causing too much fatigue for participants or moderators.

Recommendations

Overall, the results from this study seem to indicate that my prototype is a significant improvement over the current Roll20 website. With an average SUS score of 85.8 and a task success rate of 100%, I am confident that the current design allows users to roll dice successfully and with ease.

There is always room for improvement and I do have a few key take-aways based on the data and user feedback collected during this study:

• The presence of a "dice" icon in the toolbar caused confusion for several participants. This icon corresponds to Roll20's current Dice Roller control. It should either be removed from the prototype or wired up to open the new Dice Roller. In order to make the transition easier for existing users, I would likely opt for the latter.

- Several users noted inconsistency between the Attack and Spell tabs. The Spell tab includes a "Cast" button for every spell, whereas weapons in the Attack tab have no corresponding button. Providing an "Attack" button for rows in the Attack tab would make these similar views even more consistent.
- A few buttons on the "Stats & Saves" tab don't have hover effects. This is a clear omission and should simply be added to the prototype.
- When rolling hit dice and damage dice together (using the "Cast" button), it's not clear which dice are for the hit and which are for the damage. Some participants suggested using color-coding to visually separate these dice. This seems like a good solution.
- Many participants were unsure about what the "Shield" spell actually does. Because there are no dice associated with this spell, users were less confident about whether it had been used successfully. I believe the system needs clearer, more prominent visual feedback about the effects of spells being cast. One possible solution is making the user's most recent action more visually pronounced in chat. It may also be helpful to add a "feed" of the user's most recent actions in a separate location, which would double as a place to highlight recent Dice Roller results.
- Many users were confused by the icon used to represent the "piercing" damage type. This icon needs to be reworked to be more intuitive. A mini usability study could be conducted solely for the purposes of determining which visual metaphors are most widely and consistently understood for different damage types.
- Several users asked what the "A" and "D" buttons on the Dice Roller mean. Most were able to guess, but this can be made clearer by adding tooltips to the buttons.

I expect that implementing fixes to address these issues will further improve user satisfaction and efficiency when rolling dice using this new system.

Consent Form for Usability Participants

Please read and sign this form.

Website

https://roll20.net/

Study

The purpose of this study is to understand how people use the Roll20 website.

The study will be conducted via an online session using your computer and microphone. During the session you will be interviewed about the site and asked to find information or complete tasks using the site. You may be asked to interact with a mockup or prototype of the site that includes features or changes that differ from the current live version.

The session will last up to 45 minutes. You may choose to discontinue the study at any time without any penalty or loss of benefits.

Risks and Benefits

Your participation in this study will help us adapt the Roll20 website to the needs and desires of its users. Participation in this study will not affect the standing of your Roll20 account in any way.

Alternatives

There are no alternative procedures or treatments that are pertinent to this study.

Confidentiality

The session will be conducted using screen capture and audio recording technology. Your name and responses will be recorded by the study administrator. All information and recordings collected during this study are for research purposes only. Your name and identifying information will not be used for any other purpose.

All recordings and personally identifiable information will be stored securely, will not be shared with any third parties, and will be destroyed within one year of the study date.

Significant Risks and Compensation

There are no significant risks associated with participation in this study.

Contact

Please reach out to the study administrator for any of the following reasons:

- Questions or concerns about the study
- Any discomfort experienced during the study
- Questions about your rights as a study participant

Study Administrator

Name: David Clipson Email: <u>dclipson@gmail.com</u>

Voluntary Participation

Participation in this usability study is strictly voluntary. Refusal to participate will involve no penalty or loss of benefits. You may discontinue participation at any time without penalty or loss of benefits.

Statement of Informed Consent

By signing this form, I agree to to the following:

- I understand that participation in this usability study is voluntary.
- I agree to immediately raise any concerns or discomfort during the session to the study administrator.
- I understand and consent to the use of audio and screen capture recordings.
- I have read and understand the information on this form.
- Any questions I have about the session have been answered.

 Today's date:

 Please print your name:

Please sign your name:

Thank you for your time and participation in this study.

Pilot Study Test Script

Introduction

Thank you for participating in this study for the Roll20 website. This study is designed to help us improve the design of the website to better need the needs of our users. To get started, I'd like to ask a few questions about you and your history with D&D.

- In your lifetime, how many hours have you spent playing as a Player Character in the Dungeons and Dragons 5th Edition system? (Only include time spent playing with a group.)
 - I have played less than 1 hour.
 - I have played between 1 and 5 hours.
 - I have played between 6 and 20 hours.
 - I have played between 21 and 50 hours.
 - $\circ~$ I have played between 51 and 100 hours.
 - I have played for over 100 hours.
- D&D can be played remotely or virtually, meaning users rely on software to connect instead of being in the same room. Within the past 12 months, how often have your D&D sessions been primarily remote or virtual?
 - Over 75% are virtual
 - Between 51% and 75% are virtual
 - Between 26% and 50% are virtual
 - Under 26% are virtual
 - Not applicable; No D&D sessions in the past 12 months.
- Which category includes your current age?
 - Under 18 years
 - 18-25 years
 - o 26-35 years
 - 36-45 years
 - 46-55 years
 - 56-65 years
 - 66 years or older

Pre-Test

Next I will show you an interactive prototype of a redesign for the Roll20 website that we are currently working on.

- This is not a finished product, but you can interact with it like a normal website by clicking on things.
- I will ask you to complete a series of tasks. For each task, I will read the task aloud completely and then ask you to start. When asked to start, please complete the task

using the prototype. When you believe you have completed the task, please say "Done" or "Ready."

- After each task, I will ask for your thoughts and opinions about the task.
- We are testing the design, not you. Don't worry if you make mistakes or get stuck.
- Please complete tasks as accurately and quickly as possible. If you get stuck, I'll give you time to figure things out on your own.
- If you get really stuck or feel like you're not sure what to do, feel free to ask for help.

Scenario

Imagine you are playing Dungeons and Dragons 5th Edition as a Level 2 Wizard. You are using the Roll20 website so you can play with your friends online. Your adventuring party is exploring a forest that is rumored to be inhabited by goblins.

Tasks to perform

- 1. As you explore the forest, the Dungeon Master asks everybody in the party to make a Perception check. Make a Perception roll.
 - Single Ease Question (SEQ) #1. (See next section for details.)
 - Do you have any questions or comments about the task you just performed?
- 2. You see a group of hostile goblins. Combat is beginning. The Dungeon Master instructs you to "roll for initiative." Make an initiative roll.
 - SEQ #2
 - Do you have any questions or comments about the task you just performed?
- 3. It's your turn during combat. You want to attack one of the goblins with a magical Chromatic Orb. Make a roll to see if the Chromatic Orb hits.
 - SEQ #3
 - Do you have any questions or comments about the task you just performed?
- 4. Your Chromatic Orb spell hits. Make a roll to determine how much damage the Chromatic Orb inflicts.
 - SEQ #4
 - Do you have any questions or comments about the task you just performed?
- 5. One of the goblins attempts to hit you with a club. You might be able to block the hit using the Shield spell. Cast the Shield spell.
 - SEQ #5
 - Do you have any questions or comments about the task you just performed?
- 6. It's your turn again. You are running out of spell slots, so you attack with your dagger. Make a roll to see if your dagger hits.
 - SEQ #6
 - Do you have any questions or comments about the task you just performed?
- 7. Your dagger hits. Make a roll to determine how much damage the dagger inflicts.
 - SEQ #7
 - Do you have any questions or comments about the task you just performed?

- 8. After combat you decide to take a long rest. During the rest your "Portent" ability activates, allowing you to roll two d20 dice and save the results for later. You know there isn't a section for this on your character sheet. Make an ad hoc roll with two d20 dice.
 - SEQ #8
 - Do you have any questions or comments about the task you just performed?

Single Ease Question

After each task, a 7-point SEQ will be asked:

• On a scale of 1-7, with 1 being very difficult and 7 being very easy, how would you rate the task you just performed?

Post-Test

Immediately following the last task, the user will be thanked for their time and provided with a Qualtrics survey containing the following:

- 1. A modified System Usability Scale (SUS) using a 7-point scale.
- 2. "Do you have any questions or comments about the prototype you just tested?"
- 3. "Thank you very much for your time and participation in this study."