

# Paramount+ Research Plan

## Project Summary

Paramount+ wants its mobile app to be as usable as possible. The particular case examined here is to understand the usability of the app when users from our target market try to find a specific episode of a series that shows up in the “Keep Watching” section, what is affecting this usability, and how the usability of completing this task compares to the augmented mobile app. This test has both formative and summative aspects. It is formative in that qualitative feedback on the experience will be collected and analyzed for themes. It is summative in that metrics will be collected to evaluate the app’s usability and compare it to the usability of the augmented mobile app. The test will be experimental and exploratory; experimental in that users will be randomly assigned to either the Paramount+ mobile app group or the augmented Paramount+ mobile app (performs similarly to the Netflix app) group; exploratory in that qualitative feedback will describe the experience and pain points, and usability metrics will evaluate usability of the Paramount+ mobile app for this task.

## Research Goals

In an effort to improve the Paramount+ app and gain traction in the streaming market, we want to know how usable the Paramount+ mobile app is during the task of navigating from the home screen of the app to a specific episode of a series that appears in the “Keep Watching” section, and what is affecting usability.

We also want to know how the usability of completing this task compares to completing the same task in the augmented Paramount+ mobile app.

## Hypotheses

Task: Starting at the home screen of the Paramount+ / augmented Paramount+ mobile app and finding a specific episode of a show that appears in the “Keep Watching” section of the home screen

H1: Qualitative feedback:

- A. While completing the task on the Paramount+ mobile app, users will generate thematically similar negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction.

- B. While completing the task on the Paramount+ mobile app, more users will generate negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction than they will for the augmented mobile app.

## H2: System Usability Score (UMUX):

- A. After completing the task on the Paramount+ mobile app, users will rate the usability of the app below average on the Usability Metric for User Experience (UMUX).
- B. After completing the task on the Paramount+ mobile app, users will rate the usability of the app below the rating given to the augmented mobile app on the Usability Metric for User Experience (UMUX).

## H3: User actions:

- A. Completing the task in the Paramount+ mobile app will take more user actions than is required (6).
- B. Completing the task in the Paramount+ mobile app will take more user actions than it will in the augmented mobile app.

## H4: User errors:

- A. There will be errors made when completing the task in the Paramount+ mobile app.
- B. There will be more errors made when completing the task in the Paramount+ mobile app than in the augmented mobile app.

## H5: Time:

When completing the task in the Paramount+ mobile app, it will take more time than it will take in the augmented mobile app.

# Participants

## Screener

How/when will participants be screened?

## Questionnaire

- Do you watch television series on any streaming app?

- Answer should be “Yes”
- Do you ever use a streaming service with a mobile device (phone or pad)?
  - Answer should be “Yes”
- Do you have experience using the Paramount+ mobile app?
  - Answer should be “No”
- Are you between the ages of 18-50?
  - Answer should be “Yes”

## Sampling

Participants will be recruited through convenience sampling of friends and associates. Participants will be randomly assigned to either the Paramount+ app group or the augmented app group.

## Sample Size Calculation

Illustrated in Design and Procedures

## Design and Procedures

### **Task remains the same for all hypotheses**

Task: Starting at the home screen of the Paramount+ / augmented mobile app (depending on group assignment) and finding a specific episode of a show that appears in the “Keep Watching” section of the home screen

## H1: Qualitative Feedback

H1a While completing the task on the Paramount+ mobile app, users will generate thematically similar negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction.

**Design:** Qualitative exploratory

### **Sample size estimation: 5 participants needed**

Using the probabilistic model of problem discovery

$$P(x \geq 1) = 1 - (1 - p)^n$$

Constraints on available participants dictate a preference for no more than 5 participants per group. With this in mind, and an assumption that negative feedback will be fairly common, observing the sample size requirements table show that to have a 75% chance of observing a problem that occurs 25% of the

time, 5 participants are required. For a 75% chance of observing a problem that occurs 15% of the time, 9 would be required. For a 50% chance of observing a problem that occurs 15% of the time, again, 5 participants are required. Therefore, since I'm assuming negative feedback to be likely, 5 participants are needed to have a 75% chance of observing a problem that occurs 25% of the time.

**5 participants needed**

H1b While completing the task on the Paramount+ mobile app, more users will generate negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction than they will for the augmented mobile app.

**Design:** Experimental, between subjects

**Sample size estimation:**

**6 participants per group, 12 total**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a two-sample, two-tailed t-test

$$N = 2z^2s^2/d^2, d=e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 80% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = 2z^2s^2/(e*s)^2$$

$$z = 1.282$$

$$e = 0.8$$

$$N = 5.14, \text{ round up to } 6$$

Iteration using t-distribution with N sourced from equation using z

$$df \text{ (degrees of freedom)} = 2(n - 1) = 2(6 - 1) = 10, \text{ therefore}$$

$$t = 1.372$$

$$e = 0.8$$

$$N = 5.88, \text{ round up to } 6$$

**6 participants per group, 12 total**

## H2: Usability Metric for User Experience (UMUX)

H2a After completing the task on the Paramount+ mobile app, users will rate the usability of the app below average on the Usability Metric for User Experience (UMUX).

**Design:** Comparison to SUS benchmark derived from population sampling (68 is average)

### **Sample size estimation:**

**4 participants (80% confidence), 6 participants (90% confidence)**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a one-sample, one-tailed t-test

$N = z^2 s^2 / d^2$ ,  $d = e(s)$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 90% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

**Therefore  $N = z^2 s^2 / (e * s)^2$**

$z = 1.645$

$e = 0.8$

$N = 4.22$ , rounded up to 5

Iterations using t-distribution with N sourced from equation using z

Final iteration

df (degrees of freedom) =  $n - 1 = 7 - 1 = 6$ , therefore

$t = 1.943$

$e = 0.8$

$N = 5.90$ , round up to 6

**6 participants (90% confidence)**

H2b After completing the task on the Paramount+ mobile app, users will rate the usability of the app below the rating given to the augmented mobile app on the Usability Metric for User Experience (UMUX).

**Design:** Experimental, between subjects design

**Sample size estimation:**

**6 participants per group, 12 total**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a two-sample, two-tailed t-test

$$N = 2z^2s^2/d^2, d=e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 80% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = 2z^2s^2/(e*s)^2$$

$$z = 1.282$$

$$e = 0.8$$

$$N = 5.14, \text{ round up to } 6$$

Iteration using t-distribution with N sourced from equation using z

$$df \text{ (degrees of freedom)} = 2(n - 1) = 2(6 - 1) = 10, \text{ therefore}$$

$$t = 1.372$$

$$e = 0.8$$

$$N = 5.88, \text{ round up to } 6$$

**6 participants per group, 12 total**

### H3: User actions

H3a Completing the task in the Paramount+ mobile app will take more user actions than is required (6).

**Design:** Comparison to “benchmark” of 6 actions to task completion

**Sample size estimation:**

**4 participants (80% confidence), 6 participants (90% confidence)**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a one-sample, one-tailed t-test

$$N = z^2 s^2 / d^2, d = e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 90% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = z^2 s^2 / (e * s)^2$$

$$z = 1.645$$

$$e = 0.8$$

$$N = 4.22, \text{ rounded up to } 5$$

Iterations using t-distribution with N sourced from equation using z

Final iteration

$$df \text{ (degrees of freedom)} = n - 1 = 7 - 1 = 6, \text{ therefore}$$

$$t = 1.943$$

$$e = 0.8$$

$$N = 5.90, \text{ round up to } 6$$

**6 participants (90% confidence)**

H3b Completing the task in the Paramount+ mobile app will take more user actions than it will in the augmented mobile app.

**Design:** Experimental, between subjects

**Sample size estimation:**

**6 participants per group, 12 total**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a two-sample, two-tailed t-test

$$N = 2z^2s^2/d^2, d=e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 80% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = 2z^2s^2/(e*s)^2$$

$$z = 1.282$$

$$e = 0.8$$

$$N = 5.14, \text{ round up to } 6$$

Iteration using t-distribution with N sourced from equation using z

$$df \text{ (degrees of freedom)} = 2(n - 1) = 2(6 - 1) = 10, \text{ therefore}$$

$$t = 1.372$$

$$e = 0.8$$

$$N = 5.88, \text{ round up to } 6$$

**6 participants per group, 12 total**

## H4: User errors

H4a There will be errors made when completing the task in the Paramount+ mobile app.



**Design:** Estimation of true proportion of users who commit at least one task error in the population

**Sample size estimation:**

H4b There will be more errors made when completing the task in the Paramount+ mobile app than in the augmented mobile app.

**Design:** Experimental, between subjects

**Sample size estimation:**

**6 participants per group, 12 total**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a two-sample, two-tailed t-test

$$N = 2z^2s^2/d^2, d=e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 80% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = 2z^2s^2/(e*s)^2$$

$$z = 1.282$$

$$e = 0.8$$

$$N = 5.14, \text{ round up to } 6$$

Iteration using t-distribution with N sourced from equation using z

$$\text{df (degrees of freedom)} = 2(n - 1) = 2(6 - 1) = 10, \text{ therefore}$$

$$t = 1.372$$

$$e = 0.8$$

$$N = 5.88, \text{ round up to } 6$$

**6 participants per group, 12 total**

## H5: Time

H5 When completing the task in the Paramount+ mobile app, it will take more time than it will take in the augmented mobile app.

**Design:** Experimental, between subjects

### **Sample size estimation:**

**6 participants per group, 12 total**

Due to constraints on available participants, test metrics will be less conservative (lower confidence in findings, looking for a large difference)

Using a two-sample, two-tailed t-test

$$N = 2z^2s^2/d^2, d=e(s)$$

**N** = number of participants

**z** = assigned value for level of confidence in findings (we don't want to be less than 80% confident in the findings)

**s** = Standard deviation, variability in data (value we don't have)

**d** = observed difference

**e** = effect size (Looking for a large effect,  $e = 0.8$ )

$$\text{Therefore } N = 2z^2s^2/(e*s)^2$$

$$z = 1.282$$

$$e = 0.8$$

$$N = 5.14, \text{ round up to } 6$$

Iteration using t-distribution with N sourced from equation using z

$$df \text{ (degrees of freedom)} = 2(n - 1) = 2(6 - 1) = 10, \text{ therefore}$$

$$t = 1.372$$

$$e = 0.8$$

$$N = 5.88, \text{ round up to } 6$$

**6 participants per group, 12 total**

# General Procedures

## Introduction, NDA, Consent to Record

Thanks for being here today! As you know, my name is Booker and I'm working on a school project doing user experience research. My goal here is to get your feedback on the design of a streaming app. I want you to evaluate how easy the product is to use, and get your perspective on things that you like or dislike. I'll start with some questions, and then we'll go through the task, and then I'll have a few more questions at the end.

First, I'd like to **share my screen** with you and get your **verbal** consent to participate in this study:

---

Please read this form and provide verbal consent.

In this usability test:

- You will be asked to perform certain tasks on a phone app.
- I will also conduct an interview with you during this session/study.

Participation in this usability study is voluntary. This remote session will be recorded. All video and audio that is recorded will remain strictly confidential. The descriptions and findings may be used to evaluate the phone app. However, at no time will your name or any other identification be used.

You can withdraw your consent to the study and stop participation at any time.

If you have any questions after today, you can contact Booker Harrap (User Experience Researcher) at this email address: [booker.harrap@cgu.edu](mailto:booker.harrap@cgu.edu)

Once you have read and understood the information on this form and had all of your questions answered, please indicate that you consent to participation and being recorded.

---

Ok, now that I am recording, could you please once again give consent to participate in this study and be recorded?

---

The most important thing to keep in mind is that we're **testing the mobile app, not you.**

If there's anything difficult or unclear, that's exactly the type of thing I want to hear about. I want to hear your open and honest feedback on what is clear and what is unclear to you, and in particular, anything you find confusing or difficult to understand. Please think out loud as you move through the task and before you make any selections.

We are set to go until [00:00] today. Does that still work for you?

Any **questions** before we begin?

## Setting up the task

The scenario is this: You were just watching a TV show you like on your phone and you left it playing while you were doing something and missed an episode. You want to watch it now, and you need to go back one episode and watch it from there. For the sake of this task, we will say the show is "Frasier."

Please click on the link.

This is an app prototype that is scrollable just like a phone app.

Reminding you to think out loud, can you get to the previous episode of Frasier?

Thank you, great!

That's it, now I just have some follow up questions.

## Post-task questions

Do you have any comments about the task you just did?

Any questions?

Now I'd like you to answer some survey questions. It's 4 questions in total.

---

## Usability Metric for User Experience (UMUX)

Rate your answers from 1 to 7, 1 being "strongly disagree", and 7 being "strongly agree", with 4 being neutral.

1. The app's capabilities meet my requirements
  2. Using this app is a frustrating experience
  3. The app is easy to use
  4. I have to spend too much time correcting things with this app
- 

## Debrief

This usability test was designed to assess whether there is a usability problem with the Paramount+ app when trying to get to a specific episode of a show that is entered from the "Keep Watching" section of the home screen. [If you used the augmented app, you were in the comparison group, working on a system that I hypothesized to be more usable than the Paramount+ app] Your feedback and commentary will be analyzed along with other participants' comments for thematic similarities. Your in-app actions will also be analyzed: whether you completed the task, how many actions it took, how many errors you made, and the time it took to complete the task. These metrics will be compared to participants in the other app group (augmented / Paramount+).

Thank you so much for your participation!

Have you got any questions for me before we wrap up?

## Analysis

### H1: Qualitative Feedback

H1a While completing the task on the Paramount+ mobile app, users will generate thematically similar negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction.

**Design:** Qualitative exploratory

**Analysis:** Thematic analysis of user feedback. Data will be coded for valence, subject matter, and section of task. Confidence interval for true proportion.

H1b While completing the task on the Paramount+ mobile app, more users will generate negative feedback about the process that indicate low effectiveness, efficiency, or satisfaction than they will for the augmented mobile app.

**Design:** Experimental, between subjects

**Analysis:** Thematic analysis of user feedback, Data will be coded for valence, subject matter, and section of task. Quantification of negatively valenced comments along themes and task section, comparison between groups with a two-sample t-test.

## H2: Usability Metric for User Experience (UMUX)

H2a After completing the task on the Paramount+ mobile app, users will rate the usability of the app below average on the Usability Metric for User Experience (UMUX).

**Design:** Comparison to SUS benchmark derived from population sampling

**Analysis:** One sample, one-tailed *t*-test.

H2b After completing the task on the Paramount+ mobile app, users will rate the usability of the app below the rating given to the augmented mobile app on the Usability Metric for User Experience (UMUX).

**Design:** Experimental, between subjects design

**Analysis:** Two-sample, two-tailed *t*-test.

## H3: User actions

H3a Completing the task in the Paramount+ mobile app will take more user actions than is required (6).

**Design:** Comparison to “benchmark” of 6 actions to task completion

**Analysis:** One-sample, one-sided  $t$ -test, potentially log-transforming positively skewed data.

H3b Completing the task in the Paramount+ mobile app will take more user actions than it will in the augmented mobile app.

**Design:** Experimental, between subjects

**Analysis:** Two-sample, two-sided  $t$ -test

#### H4: User errors

H4a There will be errors made when completing the task in the Paramount+ mobile app.

**Design:** Estimation of true proportion of users who commit at least one task error in the population

**Analysis:** Adjusted-Wald confidence interval (errors > 0)

H4b There will be more errors made when completing the task in the Paramount+ mobile app than in the augmented mobile app.

**Design:** Experimental, between subjects

**Analysis:** Two-sample, two-sided  $t$ -test

#### H5: Time

H5 When completing the task in the Paramount+ mobile app, it will take more time than it will take in the augmented mobile app.

**Design:** Experimental, between subjects

**Analysis:** Two-sample, two-tailed  $t$ -test.