Improving Media Discovery on Amazon Prime Video: A Usability Evaluation

Author: Olu Ogunmeru

ABSTRACT

This project investigates the usability of Amazon Prime Video in helping users find movies or TV shows to watch based on recommendations or search. The original interface revealed several design problems including lack of predictive aiding in search, high information access cost, poor visibility of filtering options, and inconsistent navigation (Ogunmeru, 2025). A redesigned prototype was created using Figma, introducing predictive search suggestions, hover previews, and tabbed filtering (Included, Rent or Buy, Combined). A within-subjects experiment with six participants evaluated both versions. Participants were tasked with finding a movie included in their subscription. Metrics included task completion time (continuous), success rate (categorical), and subjective usability (NASA-TLX). Results showed faster task times and lower workload ratings with the new interface. This report details the methodology, results, and implications for streaming interface design.

INTRODUCTION

Amazon Prime Video is a major streaming platform, yet its usability for media discovery often frustrates users (Ogunmeru, 2025). Observations and interviews revealed frequent issues with search accuracy, hidden filters, and navigation inconsistencies (Ogunmeru, 2025). The task examined in this study is: finding a movie or TV show to watch based on recommendations or search.

To improve this, a new interface was developed addressing the following problems:

1. Poor predictive search suggestions.
2. High cognitive load from reading multiple movie descriptions.
3. Hidden "Free to Me" toggle (replaced with clearer tab navigation).
4. Inconsistent "See More" behaviors across categories.

Goal: Evaluate if the redesigned interface improves usability during media discovery.

Hypotheses:

1. Users will complete the task faster using the redesigned interface.
2. Users will be more successful at finding a free movie.
3. Users will report lower cognitive workload (NASA-TLX) using the redesigned interface.

METHODS

Participants:

Six participants (ages 19 to 56, mean age = 31 years old) who are active users of streaming platforms.

The Inclusion criteria was that the participants have basic familiarity with streaming services and are individuals who use the platform occasionally and are not deeply familiar with advanced features, settings, or extensive content navigation.

Experiment Setup:

Tests were conducted on a MacBook in-person. Participants interacted with clickable mock-ups in Figma simulating the redesigned interface and the original interface natively. Data was recorded with a stopwatch and screen recorder.

Experiment Design:

The **independent variable (IV)**, *interface version*, was chosen to examine the impact of design improvements on usability with the two levels being the original and redesigned interface.

The **continuous dependent variable**, *task completion time*, provides a clear, objective measure of efficiency.

The **categorical dependent variable**, *task success*, was chosen to assess whether participants were able to find a movie included in their subscription.

The decision to use a *within-subjects* **design** was made to minimize individual variability across participants, since each person served as their own control. This allowed for more direct comparison of performance between the original and redesigned interfaces.

Finally, the *NASA-TLX subjective measure* was selected to capture perceived cognitive workload, offering insights into how mentally demanding each version of the interface felt to users.

Confound Control: The interface order was counterbalanced to minimize learning effects, ensuring that any improvements in performance were not due to familiarity with the task itself. Participants were randomly assigned to begin with either the original or redesigned interface. Using the same device for all sessions controlled for hardware or screen-size-related performance differences, while keeping the task consistent ensured that observed effects could be attributed solely to interface design.

Experiment Procedure:

1. Participants first signed the consent form.
2. They then viewed task instructions:
   "Find a movie you haven't seen before that is included in your subscription that you feel you would enjoy watching."
3. Used Interface A (randomized order).
4. NASA-TLX survey completed.
5. Repeated steps 3-4 with Interface B.
6. Completed post-experiment questionnaire.

RESULTS

A paired-samples t-test was conducted to compare task completion time and NASA-TLX cognitive workload between the original and redesigned interfaces.

Task Completion Time: Participants completed the task significantly faster using the redesigned interface ($M = 11.46$, $SD = 2.35$) than the original interface ($M = 20.29$, $SD = 8.36$), $t(5) = 3.52$, $p = .017$ (two-tailed), Cohen's $d = 1.43$, indicating a large effect size. This represents a 43.5% reduction in time.

NASA-TLX Scores: Participants also reported significantly lower cognitive workload for the redesigned interface ($M = 0.17$, $SD = 0.09$) compared to the original ($M = 1.33$, $SD = 0.20$), $t(5) = 25.71$, $p < .001$, Cohen's $d = 10.49$, indicating an extremely large effect size.

Fig 1.2A in Appendix A shows the mean task times and NASA-TLX scores with standard deviation.

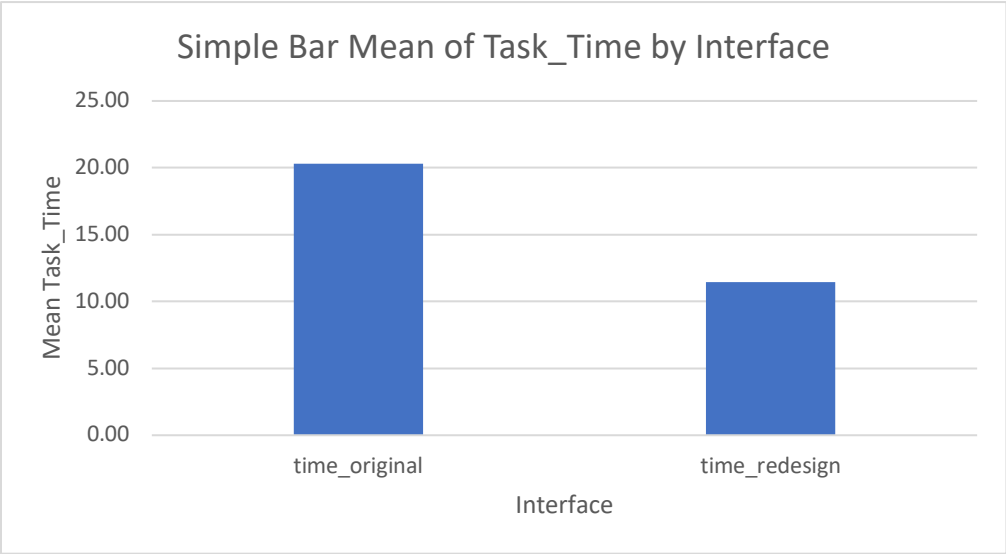


Fig 1.0. Bar Graph for Paired t-test.

DISCUSSION

The results of this study strongly support the hypotheses that the redesigned interface improves usability for the selected task. The significant decrease in task completion time suggests that users were able to navigate the new layout more efficiently, likely due to clearer content filtering (Included, Rent or Buy, Combined), improved layout consistency, and more intuitive predictive search features. These elements reduced the need for trial-and-error searching and redundant clicking.

The dramatic drop in NASA-TLX scores provides strong evidence that the redesigned interface reduced cognitive workload. This is likely the result of minimizing information access cost, increasing predictability of interface behavior, and improving visual salience. Participants not only completed the task faster, but reported it felt easier, less frustrating, and more streamlined.

While the sample size was small, the consistent direction and strength of effects suggest that the redesign has substantial potential to improve user outcomes on larger scales. Future work could explore these design changes across diverse user demographics or examine long-term engagement with the interface.

REFERENCES

Ogunmeru, O. (2025). *Assignment 1[Unpublished course assignment]*. COMP 3008, School of Computer Science, Carleton University.

Ogunmeru, O. (2025). *Assignment 2 [Unpublished course assignment].* COMP 3008, School of Computer Science, Carleton University.

Ogunmeru, O. (2025). *Assignment 3 [Unpublished course assignment].* COMP 3008, School of Computer Science, Carleton University.

Ogunmeru, O. (2025). *Assignment 4 [Unpublished course assignment]*. COMP 3008, School of Computer Science, Carleton University.

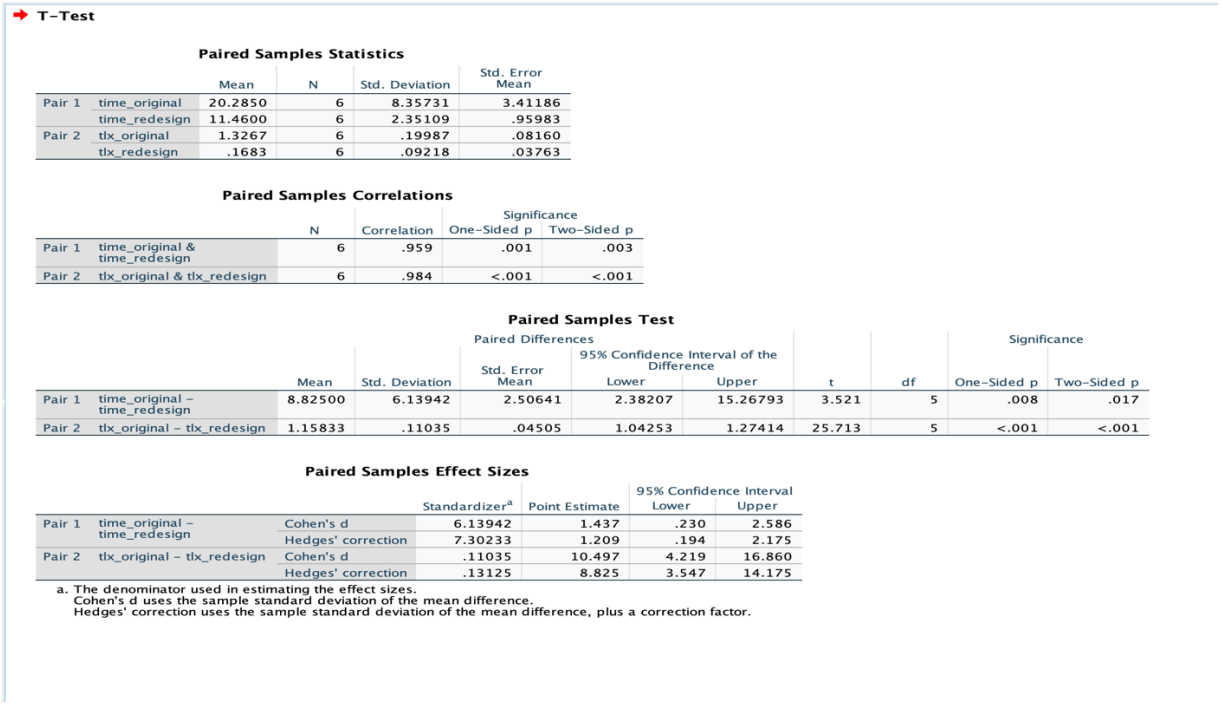
APPENDIX A

The table below shows the results of the experiment.

NOTE: The NASA-TLX score represents the overall cognitive workload on a scale of 0 to 20. 20 meaning high cognitive workload.

Participant	Original Interface (s)	Task Completed	NASA-TLX	Redesigned Interface (s)	Task Completed	NASA-TLX
P1	21.44	Yes	1.30	11.92	Yes	0.15
P2	35.12	Yes	1.67	15.10	Yes	0.34
P3	22.34	Yes	1.45	13.10	Yes	0.20
P4	17.31	Yes	1.20	9.78	Yes	0.10
P5	11.98	Yes	1.15	8.97	Yes	0.10
P6	13.52	Yes	1.19	9.89	Yes	0.12

Fig 1.A Experiment results



T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	time_original	20.2850	6	8.35731	3.41186
	time_redesign	11.4600	6	2.35109	.95983
Pair 2	tlx_original	1.3267	6	.19987	.08160
	tlx_redesign	.1683	6	.09218	.03763

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	time_original & time_redesign	6	.959	.001	.003
Pair 2	tlx_original & tlx_redesign	6	.984	<.001	<.001

Paired Samples Test

		Paired Differences					t	df	Significance One-Sided p	Significance Two-Sided p
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper				
Pair 1	time_original – time_redesign	8.82500	6.13942	2.50641	2.38207	15.26793	3.521	5	.008	.017
Pair 2	tlx_original – tlx_redesign	1.15833	.11035	.04505	1.04253	1.27414	25.713	5	<.001	<.001

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	time_original – time_redesign	Cohen's d	6.13942	1.437	.230	2.586
		Hedges' correction	7.30233	1.209	.194	2.175
Pair 2	tlx_original – tlx_redesign	Cohen's d	.11035	10.497	4.219	16.860
		Hedges' correction	.13125	8.825	3.547	14.175

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Fig 1.2A Paired t-test results.

APPENDIX B

Participant Consent Form Link:

https://carletonu.az1.qualtrics.com/jfe/form/SV_6D9wO4muCjr2Noy

NASA-TLX Survey Link:

https://carletonu.az1.qualtrics.com/jfe/form/SV_bvlXrBXZmKzUTsy

Post Experiment Questionnaire Link:

https://carletonu.az1.qualtrics.com/jfe/form/SV_cI4MmJctu0RD0Tc