

MASTER'S THESIS 2020

Usability testing on the web; Measuring how design impacts task performance times.

Stefan Eng

Elektroteknik
Datateknik

ISSN 1650-2884

LU-CS-EX: 2020-XX

DEPARTMENT OF DESIGN SCIENCES

LTH | LUND UNIVERSITY



MASTER'S THESIS
Design Sciences

LU-CS-EX: 2020-XX

**Usability testing on the web; Measuring
how design impacts task performance
times.**

Stefan Eng

Usability testing on the web; Measuring how design impacts task performance times.

Stefan Eng
atn08sen@lu.se

February 15, 2020

Master's thesis work carried out at MASSIVE.

Supervisors: John Deer, jdeer@company.com
Don Jeer, djeer@xy.lth.se

Examiner: Jane Doc, jane.doe@cs.lth.se

Abstract

Keywords: Usability testing, Web-application, Flask, HTML

Acknowledgements

Contents

1	Introduction	7
1.1	MASSIVE Entertainment A Ubisoft studio	7
1.2	Why usability testing?	7
1.3	Usability testing	8
1.3.1	Introduction	8
1.3.2	Evolution and state of the art	8
1.4	Running usability tests at larger scales	8
1.5	Report goals	8
1.6	Literary scope	8
2	Approach	9
2.1	Method	9
2.1.1	Software development methodology	9
2.1.2	Defining the initial concept	10
2.1.3	Paper prototypes for a first-draft interface	10
2.1.4	Gathering relevant test data	11
2.1.5	Evaluating impact and effectiveness	11
2.2	Theory	11
2.2.1	Information presentation and color	12
2.2.2	Representative simplified task models	12
2.2.3	Measuring time to task completion	12
2.2.4	Threshold for failure	12
2.3	Implementation	12
2.3.1	Platform software stack	12
2.3.2	Interface creation	12
2.3.3	Variable challenges	12

3	Evaluation	13
3.1	Results	13
3.1.1	Participation, test runs and success ratio	13
3.1.2	Test runs per user and participation activity	14
3.1.3	Test types, test order and task distribution	15
3.2	Discussion	15
3.2.1	Possible improvements	15
3.2.2	Threats to validity	15
4	Conclusions	17
	References	19

Chapter 1

Introduction

As the gaming industry continue to grow[!][!] so do the reported number of stress-related issues reported by the people working in the sector[!][!][!]. The initial idea for this report came from a managers observation that co-workers would abandon the digital communication software for a more hands-on approaches, such as post-it notes on a whiteboard, when the pressure got to a certain point.

Asking why, people stated that the software they were supposed to use for communicating and propagating the projects status throughout the team got in their way. Which is why they opted to use post-it notes, even though it has significantly worse communication bandwidth and is less accessible, at-least-it-works™.

1.1 MASSIVE

MASSIVE ENTERTAINMENT | A UBISOFT STUDIO

TODO: Expand section.

1.2 Why usability testing?

Initially, the plan was to make interface changes to the organization-software itself, but a question remained, how do you prove that it actually makes a difference?

TODO: Describe why usability testing, add scaling problem.

1.3 Usability testing

1.3.1 Introduction

Usability is traditionally done in person with the *over the shoulder* method which gives a good insight of what a participant does during a test. Further more, if the *thinking out aloud - method* is utilized correctly, the test-moderator should have a good insight into the participant thought-process during the test.

1.3.2 Evolution and state of the art

While effective(?) this method scales poorly with a one to one ratio between moderator and test participant. This project investigates the possibility of alleviating this scale constraint by utilizing a internet based platform to conduct usability tests of user interfaces online.

TODO: Expand this section.

1.4 Running usability tests at larger scales

TODO: Expand this section.

1.5 Report goals

1. Create a web based platform for usability testing.
2. Run one or several interface tests with real users on the platform.
3. Verify that the collected data shows a significant(?) impact on the studied variable(s) when parameters are changed. ?**TODO:** Shorten?

1.6 Literary scope

This report draws and builds on information from the fields of usability testing, web design and interaction design.

Specifically, *Handbook of usability testing : how to plan, design, and conduct effective tests*[1] and *Usability testing essentials. [Electronic resource] ready, set- test!*[2], provide a contrast between traditional and modern approaches to usability testing and how to perform them.

Don't make me think, revisited : a common sense approach to web usability[3] provides a concise and interesting summary of no-nonsense approaches to web design from a usability perspective. Last but not least *The design of everyday things*[4] introduces both *user-centered-* and *interaction-design* together with the concept of *affordances*,

Chapter 2

Approach

2.1 Method

This section expands on the methodology behind the initial concept and creation of the testing platform, as well as the process of generating and evaluating data from the usability tests performed by the participants.

2.1.1 Software development methodology

The goal was to adopt an agile development process[!] for the creation, evaluation and improvement of the usability testing platform.

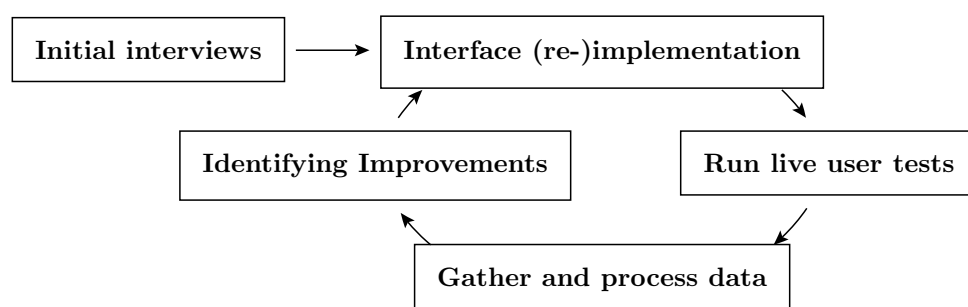


Figure 2.1: Concept, development, testing and improvement cycle.

This methodology is centered around creating a minimal working prototype[!] that is put in front of real users as soon as possible. The feedback data generated by the users should then be feed back into the design to improve the next prototype version, which is tested again. This cycle should then be repeated until either the software is satisfactory or the time is up.

2.1.2 Defining the initial concept

In order to perform usability tests that can be measured and validated, there needs to something for users to interact with. Since the subject of communication under pressure was the initial focus, the suggestion of helping managers reduce the stress for their team came up.

After interviewing a couple of managers, the following ideas were gathered:

- An easy way to see if a co-worker is assigned more work than they have available hours.
- Calendar overview where it is possible to determine if there are hot-spots where lots of results need to be produced at the same time.
- A concise way to identify if there are critical tasks that, if delayed, would delay other tasks that depend on it.
- The possibility to identify a group or teams strengths and assign task types accordingly.

2.1.3 Paper prototypes for a first-draft interface

Even though the initial interface setup was very bare-bones, it was still important to get it in front of users, in this case, test participants, as soon as possible.



Figure 2.2: Interface drafts 1.1, 1.2, 1.3 and 1.4.

?TODO: These figures should probably be under result??

Four mock-up interfaces were created based on *<insert part about ui design reference here>* and presented to five(?) colleagues for evaluation. The evaluation was conducted as an in-person interview where the interviewee were asked to voiced their thoughts out aloud. After the initial reaction and thought about each of the designs, the interviewee was asked to pick, according to them, the most suitable design.

After the initial pick, the interviewee was presented with *<insert part about ISO-standard design here>*, which they had to arrange in order of most to least important according to their own views. After prioritizing the different design attributes, the participant was again asked to pick what they felt was the most suitable design.

TODO: Expand with information about ISO design standard

2.1.4 Gathering relevant test data

There is a two-fold goal that the collected test-data needs to solve:

- Generate a quantifiable value that can be tracked in order to evaluate the participants performance during tests.
- Aggregate participant feedback about the platform in order to improve the next version.

The first problem is solved by sticking with a established tradition within usability testing[!][!], measuring completion time. In order to keep track of their test-runs, participant are assigned an anonymous id-string after acknowledging that they have read the initial information. The anonymous id is registered as used in the database and the id is stored in the participants browsers web-session.

As a participant starts a new task, the database will register the start-time and link it to the aforementioned anonymous id. If the participant completes the task, the stop-time will be recorded in the same database post, and the difference between these two values will be used as the time for task completion.

Acquiring feedback data is done through a post-test survey with ten set questions (1-5) and a free-from dialog box for additional input. TODO: Expand and reword last section

2.1.5 Evaluating impact and effectiveness

TODO: expand this section

2.2 Theory

The theory contains color-theory and the experience from managerial positions.

2.2.1 Information presentation and color

2.2.2 Representative simplified task models

2.2.3 Measuring time to task completion

2.2.4 Threshold for failure

2.3 Implementation

Writing a web-application with Flask + python + sqlite3. HTML5 and CSS with a dab javascript.

2.3.1 Platform software stack

2.3.2 Interface creation

2.3.3 Variable challenges

Chapter 3

Evaluation

3.1 Results

3.1.1 Participation, test runs and success ratio

The test-site went live 2020-01-24 and the link was initially shared through Facebook. On 2020-01-27 the link was shared on the MASSIVE internal mailing list, boosting the participation significantly. In a total, 101 participants got past the initial information page over a five day period.

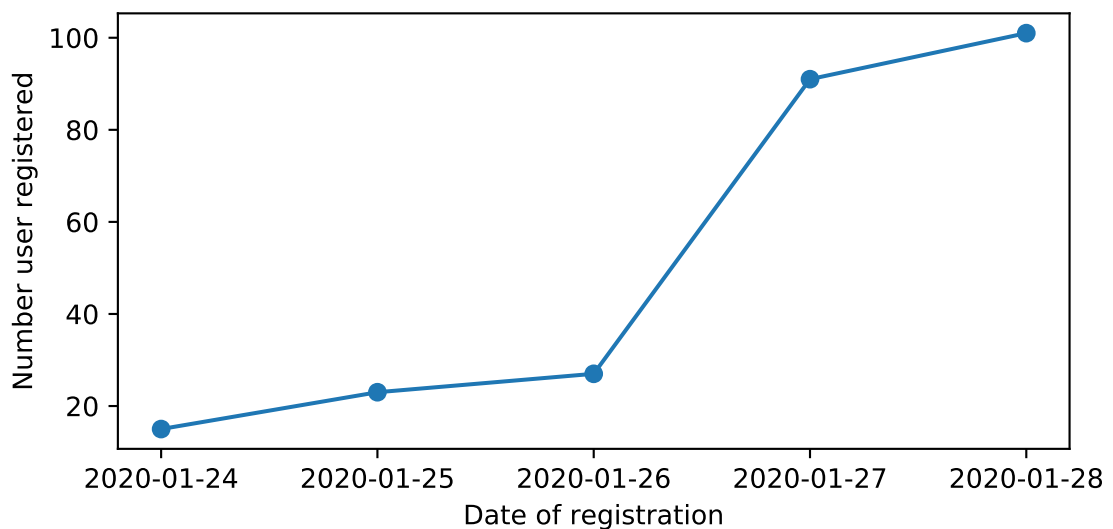


Figure 3.1: Total registered participants over time.

In total 698 tests were run, where 537 were answered correctly, 3 never produced an answer, leaving 158 incorrect answers. Looking only on the test runs without any user- or task-correlation, the chance that any given test run produces the correct answer is $\sim 76.5\%$.

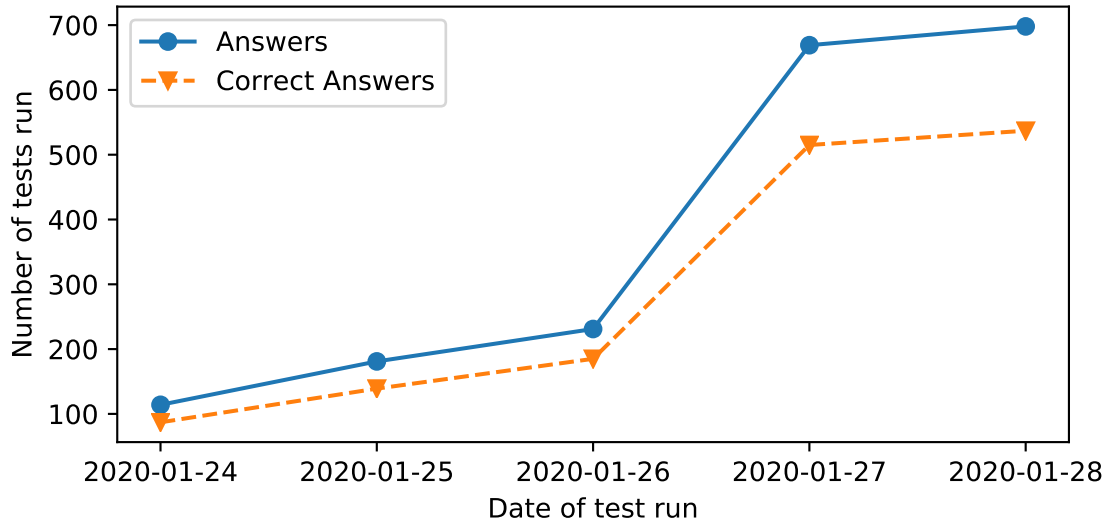


Figure 3.2: Total number of tests run overt time.

3.1.2 Test runs per user and participation activity

The number of recommended test was five, which when completed, allowed the participants to continue to the final survey. However, there was nothing stopping each participant from doing more or less than the suggested number.

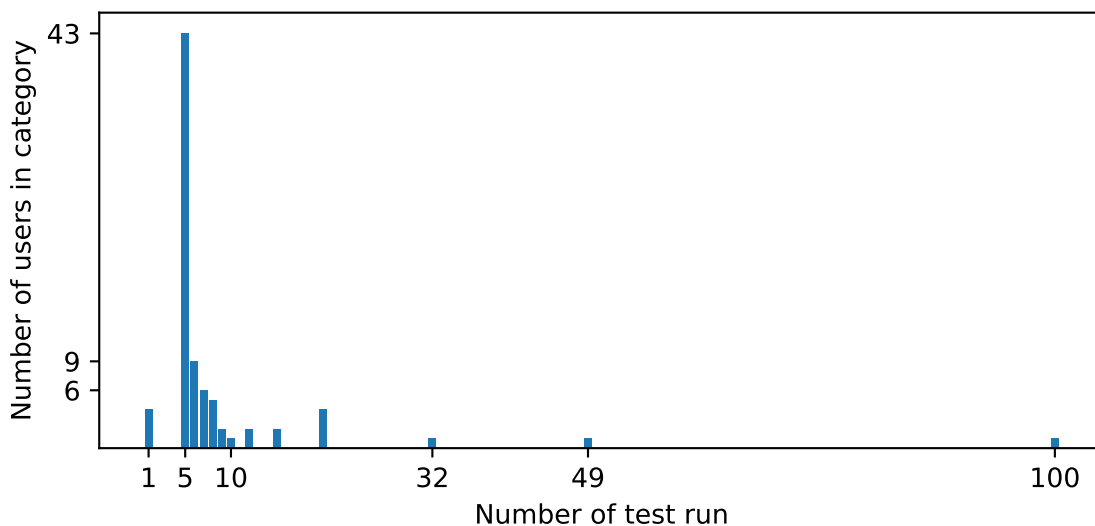


Figure 3.3: Number of users that performed a specific number of tests. (Tabulated data available in appendix.)

Of the 101 number of participant sessions started, 81 of them ran at least one test, which means 20 ran no tests. Further more, 47 completed between five and nine tests, and 12 completed ten or more.

3.1.3 Test types, test order and task distribution

Since the participants were free to choose any combination of tests, it's interesting to know which of the tasks were run the most.

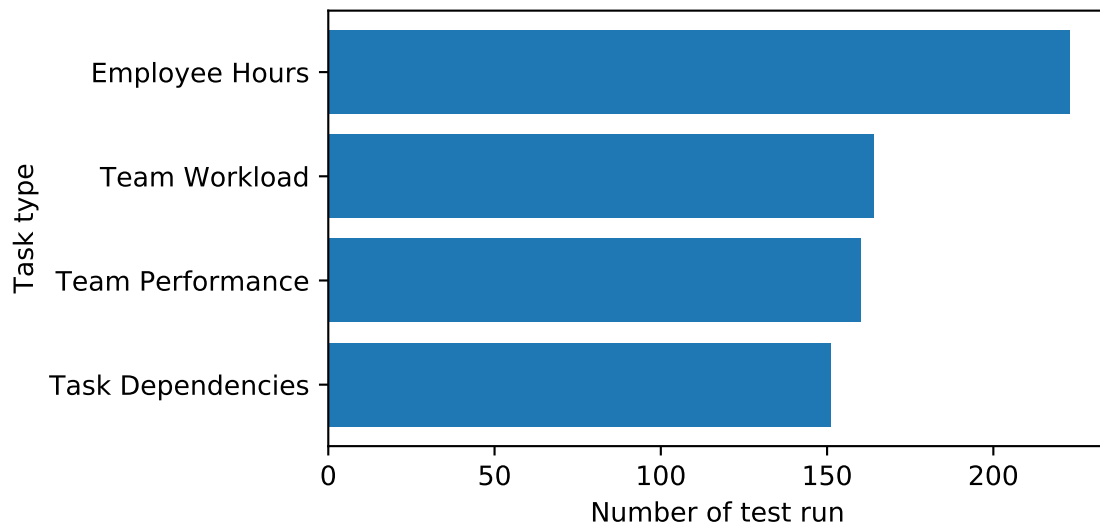


Figure 3.4: Test run distributed on task-types.

TODO: Continue exploring the data, time distribution? Which task was failed the most? Are there thresholds in the variables where tasks start to fail more often? Is there a most preferable order to do the tasks? Which task is the most popular in the 5-run category? ...

3.2 Discussion

3.2.1 Possible improvements

3.2.2 Threats to validity

Chapter 4

Conclusions

Did it have an significant impact? Was the web the correct platform? What could be done better over the internet? Recording screen and voice? (Javascript, since it's already used, pull up some statistics?)

References

- [1] Jeffrey Rubin. *Handbook of usability testing : how to plan, design, and conduct effective tests*. Wiley technical communication library. Wiley, 1994. ISBN: 0471594032. URL: <http://ludwig.lub.lu.se/login?url=https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.1401070&site=eds-live&scope=site>.
- [2] Carol M Barnum. *Usability testing essentials. [Electronic resource] ready, set– test!* Morgan Kaufmann Publishers, 2011. ISBN: 9780123750921. URL: <http://ludwig.lub.lu.se/login?url=https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.6121937&site=eds-live&scope=site>.
- [3] Steve Krug. *Don't make me think, revisited : a common sense approach to web usability*. New Riders, 2014. ISBN: 0321965515. URL: <http://ludwig.lub.lu.se/login?url=https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.3201511&site=eds-live&scope=site>.
- [4] Donald A. Norman and Donald A Norman. *The design of everyday things*. Basic Books, 2013. ISBN: 0465072992. URL: <http://ludwig.lub.lu.se/login?url=https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.3750650&site=eds-live&scope=site>.