Usability testing

Owner	John Reber
Status	COMPLETE
Version	1.2
Last updated	23/08/2018

Introduction

Usability can be defined as:

> *The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*
>
> *ISO 9241-11*

The delivery team embedded tester is expected to help the team uncover any usability issues as early in the process as possible using both their experience and a guide such as Jakob Neilsen's heuristics and usability metrics.

This can be done through static analysis of wireframes, designs, flat html pages etc. on test harnesses or through exploratory testing [link] (using tester charters and personas).

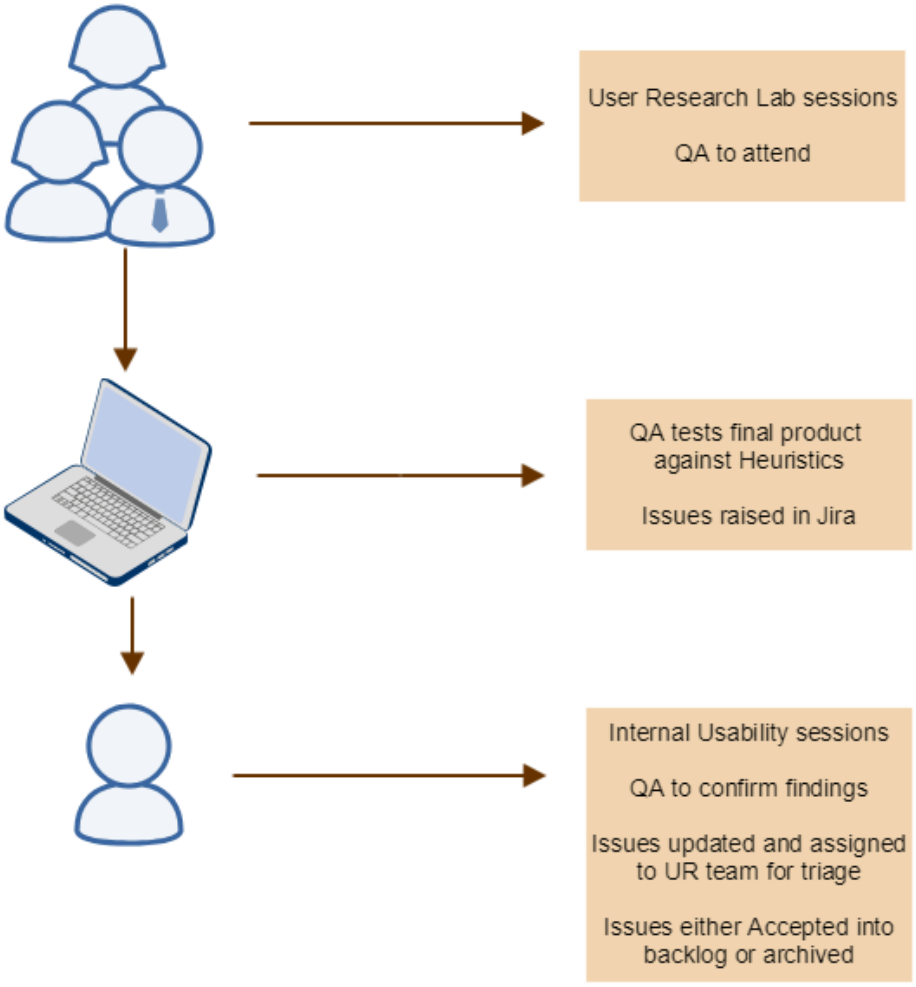
Approach

The User Research team are responsible for facilitating and organising Usability sessions with citizens via our selected providers. A QA will attend and observe where time allows.

From the QA team a usability tester will also run their own usability focused tests using Jakob Neilsen's heuristics as a guide. If any usability issues are discovered that violate the heuristics then 'lowest' priority bugs (labelled 'Usability_Testing') will be raised.

Additionally a usability tester will be tasked with running internal 'Summative' usability sessions on the released product. The purpose of these sessions is to discover whether first time users of the service make the same observations that the usability tester has found in their own assessment. If the findings are validated then further detail can be added to the JIRA ticket. A by product of the session may be the uncovering of further bugs.

This information will be fed back to the delivery teams and their respective User Researchers.



Usability Essentials

This page provides usability guidelines and methodologies to help improve communication between HMCTS Reform online services and their target users.

Term “usability” is to referred to what was then a number of vague and subjective attributes of a product, collectively known as “user friendly characteristics”.

Usability can be defined as:

> ***The extent to which a product can be used by specified users to achieve specified goals with***

> ***effectiveness, efficiency and satisfaction in a specified context of use.***

This encompasses the three critical elements of:

- ***Specific users* - not just any user, but the specific ones for whom the product is designed.**
- ***Specified goals* - these specific users have to share their goals for the product, meaning that the product's goals represent their goals.**
- ***A specific context of use* - the product has to be designed to work in the environment in which these users will use it.**

  ***Effectiveness* and *efficiency* support the users need to achieve a goal for using the product with accuracy and speed.**

User experience

You can use the honeycomb to determine what facets you want to learn from when you conduct usability testing.

Designers can also use the honeycomb as the basis for discussion about what elements are most important to build into your products so that the user experience is a positive one.

"Focus on the user, not the product."

- **Useful**: Your content should be original and fulfill a need
- **Usable**: Site must be easy to use
- **Desirable**: Image, identity, brand, and other design elements are used to evoke emotion and appreciation
- **Findable**: Content needs to be navigable and locatable onsite and offsite
- **Accessible**: Content needs to be accessible to people with disabilities
- **Credible**: Users must trust and believe what you tell them

Usability heuristics

Jakob Nielsen's 10 principles for interaction design have been around for more than 20 years but are still are good guide to usability when building web sites and applications.

1	**Visibility of system status**	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2	**Match between system and the real world**	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order. For example the term 'select' over 'click' is more relevant for the majority of users who access our services via Touchscreen devices. Likewise, we should consider carefully terms such as 'view' and 'read' as opposed to 'access' for our consumers with accessibility needs.
3	**User control and freedom**	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4	**Consistency and standards**	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5	**Error prevention**	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
6	**Recognition rather than recall**	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
7	**Flexibility and efficiency of use**	Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8	**Aesthetic and minimalist design**	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9	**Help users recognize, diagnose, and recover from errors**	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution. Help and documentation. Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.
10	**Help and documentation**	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Design & Content

Design	These will be owned by GOV.UK in the live service but have been designed and developed to use within the Uses standard GOV.UK guidance page template which includes in page navigation linking multiple guidanc Content has been written in guideline from GDS
Patterns	**Use GOV.UK guidance page template e.g. 'Appeal to the Social Security and Child Support Tribunal'** **https://govuk-prototype-kit.herokuapp.com/docs/tutorials-and-examples** **https://govuk-elements.herokuapp.com** **https://www.gov.uk/service-manual/design**
Interactions	Navigate between guidance pages using the in page navigation Click on 'Start' button to start an application

Content	

GDS Specific Checks

The following is a list of usability related GDS requirements extrapolated from the Digital Service Standard and supporting pages.

https://www.gov.uk/service-manual/technology/designing-for-different-browsers-and-devices#test-for-compatibility

"You must also make sure there's an obvious way for users to report problems so you can carry out additional testing and make adjustments to your service."

Formative and Summative Testing

Formative testing (small studies)

Usually done whilst the product is in development, with a goal of diagnosing and fixing problems, typical based on small studies, and repeated during development.

- **Reveals what the users like** - this is important so they're not lost as the product moves through development.
- **Great at ending arguments** - developers can find out what works best for users, not from vocal or powerful team members think will work best.
- **Does not provide metrics or statistics -** but gives great insight as to what the developer can put into action right away.

To get good results from small studies, you need to incorporate the following essentials elements:

1. **Define a user profile -** you need to pick one subgroup of the user population, create a profile of this user, and make this the basis for recruiting participants for the study. This is the most important part of planning so you get good results.

2. **Create task-based scenarios -** you need to give your user specific tasks to perform. These tasks should be embedded with scenarios, which are realistic descriptions framed around users' goals. You can observe their methods for achieving this goal. Without a common goal the user will go their own way and making it difficult to see usage patterns and recurrence of problems.

3. **Use a think-aloud process -** this is where you encourage the participant to share his or her thoughts whilst working on the product. This adds dimension of having the user share their thoughts, reactions, pains and pleasures. You don't have to guess what they are thinking. They tell you.

4. **Make changes & test again -** small studies usually show ***where*** the problems are, but not necessarily what the solutions are. A follow-up study can test the solutions and see if they work. This process is called iterative testing. The advantages are that you can learn from users, make changes based on what you learned, then test again.

Summative testing *(large studies)*

Usually done after the product is fully developed, with a goal establishing a baseline of metrics or validating that the product meets requirements; generally requires larger numbers for statistical validity.

Validity such as:

- failure on tasks
- average time on tasks
- completion rates
- error rates
- optimal navigation
- search results
- other measures

When planning a larger study, you can increase the number of participants from each subgroup, as there will likely be some overlap in the findings from different subgroups. For instance, of you have a budget for a study with 10 participants can can create three profiles for the three groups of three with a extra person in one of the groups. More budget, more profiles.

Not all large studies are with a finished product. In some cases, large studies may be needed whilst a product is in development, such as when testing large complex systems that need you to understand the user experience for many different subgroups.

Equipment for Testing

If you do not currently have your own usability lab, you may set up your own. You just need a few things -

- a ***dedicated room*** that can comfortably for the user, the moderator and maybe a few observers
- a ***desk or table and two (or more) chairs*** for the participants, moderator and observers.
- a ***computer or laptop*** (or what ever equipment is needed to support the product) and internet access.

- a ***camera*** to record the session, which can be a webcam or a mounted camera. Session recordings can be used to make highlights to include in reports and presentations.
- a ***microphone*** that projects everything the participants says. This may not be needed if using a webcam or a built in micro-phone.
- a ***screen recorder*** to capture whats on the device screen. This may be a app that can be downloaded or already built-in the testing device.
- a ***logging computer***, frequently a laptop which can be used to take notes during the session.

What to measure - The 5E's - Whitney Quesenbery

Dimension	Definition	Example
Effectiveness	How completely and accurately the work or experience is completed or goals reached	Can the users find the information they ned to complete tasks without assistance? Can the user perform a process within a predetermined timeframe?
Efficient	How quickly this work can be completed	Can the user successfully sign up for a service?
Engaging	How well the interface draws the user into the interaction and how pleasant and satisfying it is to use	Does the user rate their experience as satisfying or enjoyable? Do their comments (& body language) suggest that they are having a positive experience?
Error tolerant	How well the product prevents errors and can help the user recover from mistakes that do occur	Does the user experience errors? If so, how many? & when they experience errors, do they recover successfully? If they receive error messages, do they understand them?
Easy to learn	How well the product supports both the initial orientation and continued learning throughout the complete lifetime of use	Can the user get started straight away? Does their ability to do tasks improve as they become familiar with the system? Does the system architecture match their mental model for the way they expect the system to work?

Accessibility Goals

Accessibility should certainly be a goal - for legal and practical reasons. Knowing how to make the product accessible for people with disabilities or limitations such as diminished eyesight or mobility brought by age or infinity - extends the protection to all people with disabilities to provide them with equal access.

When improvements are made for people with disabilities, studies have shown that the user experience also improves for people without disabilities.

Accessibility benefits people without disabilities, including -

- older people
- people with low literacy
- people without native language fluency
- pole with lower bandwidth connections or older technology
- people with low web literacy skills

Legal reasons to address accessibility -

The Web Content Accessibility Guidelines (WCAG) are recommendation of the Word Wide Web Consortium (W3C). This was made to ensure support and access for all people.

Usability metrics

The ISO 9241-11 standard defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effect iveness, efficiency and satisfaction in a specified context of use”

The following metrics focus on measuring Effectiveness and Efficiency (Satisfaction is another metric which is slightly more nebulous and is typically captured via user feedback sessions).

Effectiveness

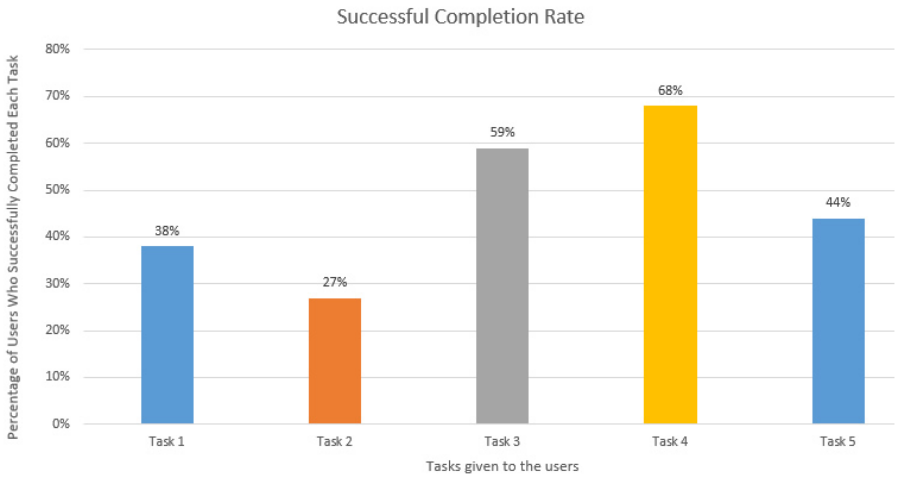
The accuracy and completeness with which users achieve specified goals

$$Effectiveness = \frac{\text{Number of tasks completed successfully}}{\text{Total number of tasks undertaken}} \times 100\%$$

For example 5 users perform a task using the same system. At the end of the test session, 3 users manage to achieve the goal of the task while the other 2 do not.

$$Effectiveness = \frac{3}{5} \times 100\% = 60\%$$

Each task can be categorised and understood as follows:



Errors

Another measurement involves counting the number of errors the participant makes when attempting to complete a task. Errors can be unintended actions, slips, mistakes or omissions that a user makes while attempting a task. Consider assigning a short description, a severity rating and classify each error under the respective category.

It is possible to discover how many errors there are per task and also ask the question as to what is an acceptable level.

The resources expended in relation to the accuracy and completeness with which users achieve goals.

Efficiency is measured in terms of task time, that is, the time (in seconds and/or minutes) the participant takes to successfully complete a task. The time taken to complete a task can then be calculated by simply subtracting the start time from the end time.

A more complex equation to get some stats on efficiency is as follow:

$$Time\ Based\ Efficiency = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N}\frac{n_{ij}}{t_{ij}}}{NR}$$

N = Number of Tasks
R = Number of Users
j = Success of task completion (1 or 0)
i = Time taken to complete a task

Example: 4 users who use the same product attempt to perform the same task. 3 users manage to successfully complete it – taking 1, 2 and 3 seconds respectively. The fourth user takes 6 seconds and then gives up without completing the task. Which gives the following sum:

$$Time\ Based\ Efficiency = \frac{\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{0}{6}\right)}{1 \times 4} = 0.46\ goals/sec$$

Readability statistics

Selecting text from a Reform service and pasting into a Word Doc the tester can check for readability.

Test Your Documents Readability

How to test with Microsoft Word and Outlook

Understand how languages affect readability scores

The languages that you use in a document can affect how your Office program checks and presents readability scores.

- If you set up Word to check the spelling and grammar of text in other languages, and a document contains text in multiple languages, Word displays readability statistics for text in the last language that was checked. For example, if a document contains three paragraphs — the first in English, the second in French, and the third in English — Word displays readability statistics for the English text only.
- For some European languages within an English document, Word displays only information about counts and averages, not readability.

Flesch Reading Ease test

This test rates text on a 100-point scale. The ***higher the score, the easier it is to understand*** the document. For most standard files, you want the score to be ***between 60 and 70***.

The formula for the Flesch Reading Ease score is:

206.835 – (1.015 x ASL) – (84.6 x ASW)

where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

Flesch-Kincaid Grade Level test

This test rates text on a **U.S. school grade level**. For example, a score of 8.0 means that an eighth grader can understand the document. For most documents, aim for a score of approximately ***7.0 to 8.0.***

The formula for the Flesch-Kincaid Grade Level score is:

(.39 x ASL) + (11.8 x ASW) – 15.59

where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

GDS specific checks

The following is a list of usability related GDS requirements extrapolated from the Digital Service Standard and supporting pages.

https://www.gov.uk/service-manual/technology/designing-for-different-browsers-and-devices#test-for-compatibility

"You must also make sure there's an obvious way for users to report problems so you can carry out additional testing and make adjustments to your service."

References

User experience (UX) team space

https://dojo.ministryoftesting.com/lessons/a-software-tester-s-guide-to-usability?utm_source=Newsletter194&utm_medium=Email&utm_campaign=Newsletter194&utm_source=Ministry+of+Testing&utm_campaign=9a259c12fe-EMAIL_CAMPAIGN_193&utm_medium=email&utm_term=0_51c24cf39c-9a259c12fe-289342369

https://www.nngroup.com/articles/usability-metrics/

https://www.techsmith.com/morae.html