

A participant-based evaluation of the UI of My5

Q Misell

April 20, 2022

Contents

1 Abstract	2
2 Introduction	2
3 Evaluation strategy	2
3.1 System Usability Scale	2
3.2 A/B testing	2
4 Research methodology	3
4.1 A/B testing design	3
4.2 SUS	3
4.3 User tasks	4
4.4 Participants	4
4.5 Data analysis	4
4.6 Pilot testing	5
5 Results	5
6 Conclusions	6
Bibliography	7
A Questionnaire	7
A.1 Basic questions	7
A.2 BBC iPlayer	7
A.3 My5	8

1 Abstract

2 Introduction

Many TV broadcasters offer video on demand services, which contribute to a sizeable percentage of the TV market globally[1]. As more and more people move towards on demand services over linear TV it is vital that broadcasters remain competitive with their VoD offering.

I have been asked to conduct an evaluation of Channel 5 TV company's My5 video on demand streaming service, to assess the usability of their service. The evaluation will be conducted in the form of a comparison with competitor video on demand services to gauge My5's usability in relation with the market, and to note where others have excelled My5, to provide opportunities for learning. The evaluation will be used to ensure Channel 5's position in the video on demand market remains competitive.

3 Evaluation strategy

The primary purpose of this evaluation is to consider how Channel 5's offering compares to the rest of the VoD market. The BBC iPlayer was chosen as a competitor with a similar offering to compare against. The BBC has the highest market VoD penetration of any traditional broadcaster in the UK - 28% - although this is far behind some competitors such as Netflix at 66%[2]. It was decided to go with the BBC over Netflix, as Netflix's offering is too different from that of Channel 5's My5 and the BBC's iPlayer to be able to effectively conduct an A/B test. The BBC's online offering is also rated highly by audiences in comparison to alternative services[3], so will offer a good benchmark for a market leading VoD service. The comparison between the two services was conducted with a System Usability Scale and an A/B test.

3.1 System Usability Scale

One of the best ways to identify pain points with a UI is by interviewing people, we can follow up on unexpected responses and dig deeper into problems identified. However, due to time constraints and the current difficulty of organising the type of in person interaction required for a good interview under Covid-19, a remote questionnaire was chosen. A questionnaire streamlines the analysis process, especially if numerical responses are given to questions, however constructing an effective but simple questionnaire is a surprisingly difficult task[4].

A Likert scale[5] is often used to collect perceptions of a system through rating scales, usually utilising a five point scale from 'strongly agree' to 'strongly disagree'. The System Usability Scale is a "robust and reliable"[6] which is particularly suited to the task at hand, both with its ability to be automatically scored (using a spread-sheeting programme) and its simplicity for participants to use. The SUS correlates well with other, more complicated, usability scales. Additionally the SUS is also available to use free of charge, unlike some other usability questionnaires, with only a reference to the original source.

3.2 A/B testing

A/B testing is representative of real world usage, as compared to lab based inspections which can result in artificial or contrived results not representative of how the user may have acted unguided in the real world[7]. As the objective here is to analyse the real world usability of a VoD platform, and users are not expected to read a manual etc. before starting to use a VoD platform entirely new to them an unsupervised A/B test is the most suitable for this.

4 Research methodology

Google Forms was used to collect responses from participants, it is a service participants will likely already be familiar with how to use, and has an intuitive and no nonsense form interface. It also allows providing blocks of text in-between questions, which were used to provide instructions about tasks to complete on the websites under evaluation before continuing with responding to the survey. *Google Forms* additionally provides some useful advanced features, such as randomisation of page order, which was used to have 50% of participants evaluate Channel 5's service first, and the other evaluate the competitor first. Google is a market leader in accessible web technology, so the form will be accessible to those who use screen readers, or other common technology to access their computer in a way that works for them.

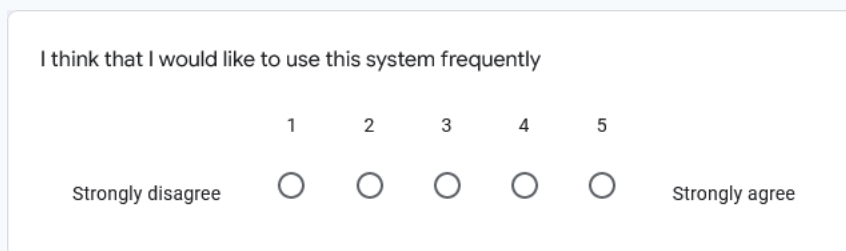
The lack of control in a remote unsupervised test was seen as a benefit to this study, as it provides opportunity for more natural behaviour when interacting with the systems under test, as opposed to the artificial results that can result from lab-based studies. The tests were conducted against the live versions of the services available at <https://my5.tv> and <https://www.bbc.co.uk/iplayer>. This ensures the test data and interaction fidelity is representative of the actual product, and not some stripped down version contrived exclusively for this test. As users are being left unsupervised when interacting with the services there is a risk that users may come across content upsetting to them. There will be nothing illegal or extremely explicit on these services, as they're regulated by the Ofcom broadcasting code[8]. Both services also provide warnings before watching any content that may be unsuitable for some audiences, regardless all participants were over 16 and voluntarily agreed to take part.

4.1 A/B testing design

The interface of the two VoD services was evaluated through a completion of the same set of tasks on both systems. This ensures the results are comparing the usability of the two systems, and not on the relative difficulty of different tasks. The testing was completed using a *within-groups* experiment. This is suited best for simple tasks with a low learning curve, such as the systems under test[9]. It has certain disadvantages with using the same users to test both systems, in that there might be advantages or disadvantages to having completed a similar task before. As stated above the *Google Forms* page randomisation feature will be used to show some users My5 first, and others the iPlayer first.

4.2 SUS

The participants were presented with the SUS questions and a linear scale 1-5 from strongly disagree to strongly agree, as shown below.



I think that I would like to use this system frequently

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

None of the SUS questions were required, and this was communicated to the participants in the notes accompanying the questionnaire. If a participant felt they couldn't answer a question a score of 3 was given for the final score calculation[6].

The questions in the SUS alternate between questions designed to elicit a positive response, with those designed to elicit a negative response. This engages the participant and attempts to eliminate inadvertent bias; e.g. in always selecting responses closer to 5.

For odd numbered questions their score is the scale position minus 1, whereas for even questions it is 5 minus the scale position. These individual question scores can then be summed and multiplied by 2.5 to give a final SUS score from 0 to 100.

4.3 User tasks

Four tasks were selected to be given to users to complete, representing the primary features common to both platforms. The tasks chosen were:

1. Watch a specific programme; on My5 this was chosen to be 'London's Greatest Bridges' episode 4, on the iPlayer the chosen programme was 'The Graham Norton Show' series 29, episode 20.
2. Watch live TV; on My5 participants were asked to watch 5Select, on the iPlayer it was BBC Parliament.
3. Find out when a programme is on; on My5 find what is on at 7pm tomorrow on 5Action, on the iPlayer find what is on at 7pm tomorrow on BBC Scotland.
4. Find a programme of your choosing and add it to 'My List' on My5 and 'My Programmes' on the iPlayer.

Tasks will be completed in the same order with all users, and are roughly in order of increasing difficulty[9]. Users will not be provided with instructions on how to complete the tasks, as they would be hitting the systems cold in a real life scenario. Users were asked if they managed to complete each task, as a rough gauge of both task difficulty and the usability of both systems in respect of individual tasks.

4.4 Participants

We were strictly prohibited from conducting this study with the general public. The survey was sent out to friends and acquaintances in various society chat groups (e.g. Discord). There is a good mix of people and backgrounds in University societies. The sample was however self-selecting in some ways, as it was sent out as an open invitation with no obligation to complete, it was however expected that a mostly uniform distribution of characteristics would appear, with the exception of computer proficiency, which would tend towards the more proficient end of the spectrum due to how the invitation to participate was given.

Some basic questions were placed at the start of the questionnaire to gather background information about the participants, some objective such as age, native English speaker status, etc., some more subjective such as to rate their perceived proficiency with computers.

4.5 Data analysis

The responses from *Google Forms* are provided in the form of a *Google Sheets* spreadsheet, which can easily be exported to a CSV file for analysis. The data was processed, analysed, and rendered in Python; using Pandas for processing and statistical analysis and Matplotlib for graph rendering.

4.6 Pilot testing

Whilst conducting pilot testing the questionnaire was overall rated as easy to understand and follow by pilot participants. A few issues where however raised, primarily that it wasn't clear that a My5 and BBC account where required to complete some of the tasks, and that it was not clear to participants that such accounts where free to create. A notice about this was added to the questionnaire, drawing attention to the free nature of these accounts. It was also raised by one pilot participant that not everyone will have the TV license required to access the iPlayer, so some advice was added regarding this, suggesting that if one does not have a TV license they may wish to complete this questionnaire at a friend's house who does have a TV license, to be in compliance with the law.

5 Results

13 people participated in this study, with a 100% response rate. Participants where aged between 17 and 50, and 85% where native English speakers. Most rated their proficiency with computers as 4 (on a scale of 1-5), with only 2 choosing a lower scoring. This is likely due to the way the participants where recruited, that being using internet communications, so people would have to be fairly competent with computers in the first place to receive the invitation. 77% of participants had used the iPlayer previously, whilst only 54% had previously used My5. 46% participated using a mobile phone, 38% on a laptop, and 15% on a desktop computer.

Figure 1: Self rated computer proficiency of the participants

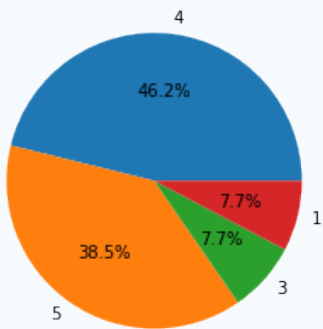
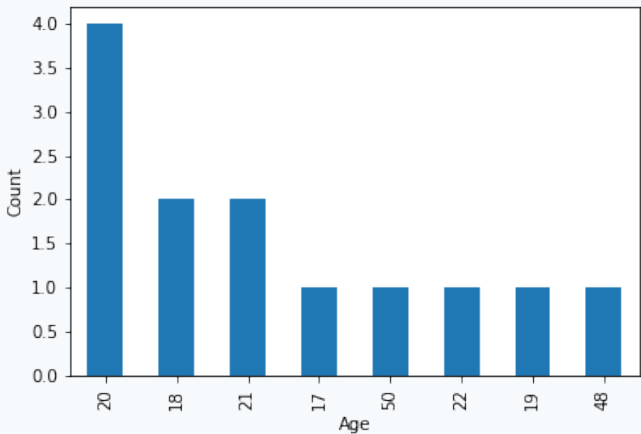
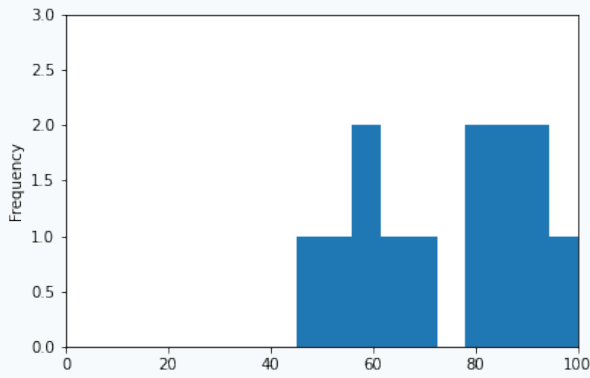


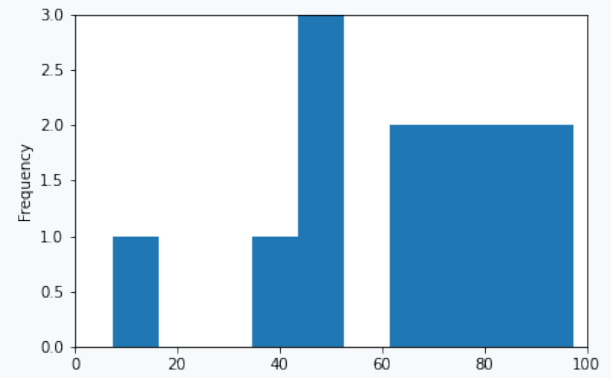
Figure 2: Age of the participants



The table below summaries the overall SUS scores for both platforms, as well as the SUS scores for some groups of participants. My5 is rated higher than the iPlayer, and of particular note is that those who had not used My5 before gave it a much higher rating than those who had not used the iPlayer before. There is not notable difference in ease of use between native and non-native English speakers.



(a) My5 SUS



(b) iPlayer SUS

Figure 3: Frequency distribution of the SUS scores

Platform	Group	Count	Mean	Std. dev.	Min	Max
My5	Total	13	74.42	17.74	45.0	100.0
	Used My5 before	6	77.92	17.92	52.5	92.5
	Hadn't used My5 before	7	71.43	18.42	45.0	100.0
	Native English speaker	11	70.91	16.78	45.0	92.5
	Non-native English speaker	2	93.75	8.84	87.5	100.0
iPlayer	Total	13	63.65	25.30	7.5	97.5
	Used the iPlayer before	10	72.00	18.29	47.5	97.5
	Hadn't used the iPlayer before	3	35.83	28.75	35.0	65.0
	Native English speaker	11	65.05	27.18	7.5	97.5
	Non-native English speaker	2	72.50	10.60	65.0	80.0

A paired samples t-test was conducted with a threshold value of 0.05 to determine if any significant difference existed between the two interfaces. The t-value between the two SUS scores is 1.567 with a p-value of 0.143. This is not significant enough to reject the null hypothesis, and therefore we must conclude there is no significant difference between My5 and the iPlayer in terms of usability.

Additional ANOVA tests were performed, using previous experience with My5/iPlayer as the factor. Between those who had not used the iPlayer before and those who had the F-value is 7.119 with a p-value of 0.022. For My5 the F-value is 0.140 with a p-value of 0.716.

6 Conclusions

Bibliography

- [1] "The uk vod market, Current status and future developments," en, Ampere Analysis Ltd, 2019. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0026/149075/ampere-analysis-current-status-future-development.pdf (visited on 03/18/2022).
- [2] "Ofcoms Annual Report on the BBC, 2019/20," en, Office of Communications, Nov. 2020. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/207228/third-bbc-annual-report.pdf (visited on 03/21/2022).
- [3] "BBC Performance Tracker." en. (2021), [Online]. Available: <https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/bbc-performance-tracker> (visited on 03/21/2022).
- [4] D. Benyon, *Designing User Experience*, 4th ed. Pearson Education Limited, 2019.
- [5] R. Likert, "A technique for the measurement of attitudes," *Archives of psychology*, 1932.
- [6] J. Brooke, "SUS: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, Nov. 1995.
- [7] S. Firmenich, A. Garrido, J. Grigera, J. M. Rivero, and G. Rossi, "Usability improvement through A/B testing and refactoring," *Software Quality Journal*, vol. 27, no. 1, pp. 203–240, 2019.
- [8] "The Ofcom Broadcasting Code (with the Cross-promotion Code and the On Demand Programme Service Rules)." en, Office of Communications. (2020), [Online]. Available: <https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code> (visited on 03/24/2022).
- [9] C. Lewis and J. Rieman, "Task-centered user interface design," en, 1993. [Online]. Available: <https://web.cs.dal.ca/~jamie/TCUID/tcuid.pdf> (visited on 03/24/2022).

A Questionnaire

A.1 Basic questions

- How old are you?
- Are you a native English speaker?
- How would you rate your proficiency with computers?
- Have you used My5 before?
- Have you used the iPlayer before?
- On what device will you be partaking in this evaluation?

A.2 BBC iPlayer

Instructions given:

Please complete the following tasks on the BBC iPlayer (available at <https://www.bbc.co.uk/iplayer>), in the order that they are given.

1. Find and start watching 'The Graham Norton Show' series 29, episode 20.
2. Watch BBC Parliament live.

3. Find out what is on at 7pm tomorrow on BBC Scotland.
4. Find a programme of your choosing and add it to 'My Programmes'.

Questions:

- Please tick all tasks that you managed to complete successfully
- I think that I would like to use this system frequently
- I found the system unnecessarily complex
- I thought the system was easy to use
- I think that I would need the support of a technical person to be able to use this system
- I found the various functions in this system were well integrated
- I thought there was too much inconsistency in this system
- I would imagine that most people would learn to use this system very quickly
- I found the system cumbersome to use
- I felt very confident using the system
- I needed to learn a lot of things before I could get going with this system

A.3 My5

Instructions given:

Please complete the following tasks on My5 (available at <https://my5.tv>), in the order that they are given.

1. Find and start watching 'London's Greatest Bridges' episode 4.
2. Watch 5Select live.
3. Find what is on at 7pm tomorrow on 5Action.
4. Find a programme of your choosing and add it to 'My List'.

Questions:

- Please tick all tasks that you managed to complete successfully
- I think that I would like to use this system frequently
- I found the system unnecessarily complex
- I thought the system was easy to use
- I think that I would need the support of a technical person to be able to use this system
- I found the various functions in this system were well integrated
- I thought there was too much inconsistency in this system
- I would imagine that most people would learn to use this system very quickly
- I found the system cumbersome to use
- I felt very confident using the system
- I needed