

Institution: University of Sunderland		
Unit of Assessment: 11 – Computer science and informatics		
Title of case study: Improving the user experience of GOV.UK and Wellcome websites		
Period when the underpinning research was undertaken: 2012-2016		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Sharon McDonald Helen Edwards	Professor Professor	1998-present 1993-2015
Period when the claimed impact occurred: 2015-December 2020		
Is this case study continued from a case study submitted in 2014? N		
<p>1. Summary of the impact (indicative maximum 100 words) Reliable user research is key to the development of effective websites and online services. Since 2015, Sunderland's research on think-aloud testing has underpinned training that has changed the way government and commercial user researchers conduct testing. This has improved data quality and enhanced the design (and consequently the user experience) of four cornerstone GOV.UK systems:</p> <ul style="list-style-type: none"> • GOV.UK Notify (2 billion messages sent since 2016) • GOV.UK Verify (7.5 million users registered since 2016) • GOV.UK Pay (£11.m paid via the system since 2016) • The Government Design System, the official repository of design tools for developers across Government <p>Since 2018, it has been used to test and develop all Wellcome Trust websites.</p>		
<p>2. Underpinning research (indicative maximum 500 words) Until 2010, usability testing research was dominated by global method comparison studies in which researchers compared two or more usability evaluation techniques to find a single best approach. However, the methods used could not be tightly controlled, confounding results with extensive evaluator effects. Previous work on usability evaluation methods had made great strides in examining process-based improvements in analytical inspection methods. Research led by Sunderland in collaboration with the University of York extended this to empirical user testing. A further important innovation focused on a cluster of resources that support think-aloud, widely used in user testing. In think-aloud, facilitators ask usability participants to say out loud what they are thinking as they complete tasks, revealing solving strategies, expectations, likes and dislikes; data that can be used to improve the interface. The resources supporting think-aloud include instructions given to participants, probes used by facilitators to elicit verbalisations, and the type of protocols (concurrent, retrospective, or both).</p> <p>There are both resource and evaluator effects in user testing, but the former are more amenable to control. Asking users to think aloud can change their strategies, often improving their performance on a task. However, when thinking aloud competes with the task by creating an increased demand on mental resources, performance can worsen. Changes in performance here are due to reactivity, a psychological phenomenon, which can result in task performance that would not occur in normal use, leading to usability problems being missed or misidentified, resulting in inappropriate design changes to address them. There is thus a trade-off between utility of feedback and its validity.</p> <p>The research programme began with a global survey of 207 practitioners [R1], which aimed to establish whether and how utility is traded-off against validity when planning think-aloud for user testing. Most reported that they would modify how they elicit think-aloud verbalisations in order</p>		

to solicit data that (1) they believed was easier to analyse in terms of identifying usability problems, and (2) would provide explanations or comments about users' experiences. However, these practices can lead to reactivity confounds. Consequently, the remainder of the research focused on enhancing utility and ease of analysis with reduced impact on validity when using think-aloud approaches.

R2 compared a neutral instruction (participants were simply asked to think aloud) with a focused set of utterance types that evaluators sought (e.g., explanations, likes and dislikes, usability issues). Task performance was the same in both conditions, but focused instruction yielded more useful utterances about usability issues, and identified more problems, albeit of low severity.

R3 added silent working as a control condition. There was no overall difference in successful task completions, but the instructions did change users' behaviours. Participants given explicit instructions engaged in more extended navigation behaviours than those working silently. A relatively small gain in the number of problems identified was thus achieved at the cost of distorted test results.

R4 examined facilitators' prompts, such as "you said X ... could you clarify?", instead of initial test instructions in R2 and R3. Common probes were gathered from two sources: key practitioner texts and observations of usability tests. Their impact on the quality of the verbal data produced and the problems identified were compared to un-prompted think-aloud. This study identified a range of both useful and unhelpful probes that practitioners can use, but with risks to validity that were not outweighed by larger, more significant problem yields.

R5 focused on the overall protocol as an evaluation. A classic concurrent think-aloud with neutral instructions and no interactive prompts (to safeguard validity) was followed by retrospective discussion to elicit explanations about the user experience. This exposed more usability problems, and reduced analysis time, but with no concomitant risk to validity, and for only a small increase (25%) in test session length.

3. References to the research (indicative maximum of six references)

R1 McDonald, Sharon, Edwards, Helen and Zhao, Tingting (2012) [Exploring Think-Alouds in Usability Testing: the Findings of an International Survey](#). *IEEE Transactions on Professional Communication*, 55 (1). pp. 2-19. **Quality indicators:** Peer reviewed article in Q2-ranked journal. 92 citations.

R2 Zhao, T. T., McDonald, Sharon and Edwards, Helen (2012) [The Impact of two different think-aloud instructions in a usability test: a case of just following orders?](#) *Behaviour and Information Technology*. pp. 1-21. **Quality indicators:** Peer reviewed article in Q2-ranked journal. 40 citations.

R3 McDonald, Sharon and Petrie, H (2013) [The Effect of Global Instructions on Think-aloud Testing](#). In: *ACM SIGCHI Conference on Human Factors in Computing Systems 2013*, April 27 – May 2, 2013, Paris, France. **Quality indicators:** Peer reviewed paper presented at the most prestigious conference in the field. 21 citations.

R4 McDonald, Sharon, Zhao, T and Edwards, Helen (2013) [Dual Verbal Elicitation: The Complementary use of Concurrent and Retrospective Reporting Within a Usability Test](#). *International Journal of Human-Computer Interaction*. 29(10), 647-660. **Quality indicators:** Peer reviewed article in Q2-ranked journal. 20 citations.

R5 McDonald, Sharon, Zhao, Tingting and Edwards, Helen (2016) [Look Who's Talking: Evaluating the Utility of Interventions During an Interactive Think-Aloud](#). *Interacting with Computers*. pp. 387-403. ISSN 0953-5438 **Quality indicators:** Peer reviewed article in Q2-ranked journal. 9 citations.

4. Details of the impact (indicative maximum 750 words)

This body of research was translated into a practical professional training course on think-aloud methods in usability testing, designed and delivered by Zhao (a former Sunderland PhD student who contributed to the underpinning research) during her employment in the Government Digital

Service (GDS). Between 2015-2020 the training ran within the GDS (approximately 25 attendees) and the Home Office (approximately 45 attendees). She then moved to the online mortgage broker Habito, where she delivered the training to colleagues, and since 2020 she has delivered the training at public events (74 attendees) [S1].

Impact on Government web services

The GDS user research team is responsible for ensuring a high-quality user experience for visitors to the 3,000 busiest GOV.UK webpages (receiving 80% of the site's traffic, or between **11-14 million unique users each week**) and is also responsible for the quality of the site's search functionality and the GOV.UK Design System digital service design tool.

In 2015, a former User Research Lead within the GDS attended Zhao's training. She recalls: "*The training was important because (1) the majority of our work uses think-aloud protocols, and (2) the method is not covered by the most widely-used resources on usability testing.*" Seeing the need for the training in other parts of the organisation, **she incorporated it into internal training given to other Government departments**. The training "*improved my and my team's understanding of think-aloud protocols, particularly around what researcher can and cannot say during a think-aloud study.*" [S2] Its quality was summarised by the Home Office Digital (HOD) Head of User Research, who said "***This is the single best course on think-aloud usability testing you will ever attend. [It] informed the training I now run myself. Even if you think you know all there is to know about think-aloud, this course is worth attending***" [S3].

The research-based training has **changed the way both GDS and HOD user research teams conduct think-aloud protocols**. The former GDS User Research Lead said "*This better understanding was applied immediately, and the team and I changed how we conduct think-aloud protocols to incorporate this best practice. Previously when asking participants to talk through their thinking during a test, user researchers would use imprecise language, and wouldn't fully consider how they should instruct a participant to complete the task. This meant that delivery could be unstructured and inconsistent across the team. Now the team uses specific language and considered instructions to encourage participants to vocalise what they are thinking.*" [S2]. Likewise, the HOD Head of User Research says the work "***transformed my approach to think-aloud***" [S3].

Three factors indicate the reach of the impact on GOV.UK, its designers, and its users: the GDS user research team is responsible for optimising the user experience of the pages and systems that receive 80% of all GOV.UK traffic; the majority of their user research uses think-aloud protocols; and every project incorporates fortnightly user research at discovery, alpha, and beta stages, and continues once live [S5]. By informing the way GDS conduct this research, the **Unit's research contributes to the design and maintenance of critical Government web pages and systems** including:

- **GOV.UK Verify** (launched May 2016): a single, common identity assurance platform for users to verify who they are when accessing government services online [S6]. **7.95 million registered individuals use GOV.UK Verify** to e.g. apply for Universal Credit, file self-assessment tax returns, sign mortgage deeds, apply for a DBS check, or receive their state pension [S9].
- **GOV.UK Notify** (launched March 2016), which enables local and national Government departments to send text messages, emails or letters to keep users updated on the progress of their transactions with Government [S6]. 4,003 services across 920 Government organisations use Notify, including NHS Test and Trace, Tax Free Childcare, Her Majesty's Passport Office, HM Courts and Tribunals Service, the Cabinet Office Register to Vote service, and the Driver and Vehicle Standards Agency's MOT Remind Me service. **Since 2016 over 2.4 billion notifications have been sent through GOV.UK Notify** [S7].
- **GOV.UK Pay** (launched Sept 2016): an online interface for users to pay local and national Governments using a range of online methods, regardless of what service they are using [S6]. For example, users can pay their vehicle tax, apply for an emergency travel document, apply for a Blue Badge, search for local land charges and apply for

transport and travel passes. To date, over **12.7 million payments worth £687 million have been made through the system [S8]**.

GDS are also responsible for creating and maintaining **GOV.UK Design Systems [S10]**, the government's official repository of styles (such as layout and colour), components (such as forms and navigation) and patterns (such as question pages and account creation) for use by Government departments when designing their own digital service. **The tools in this repository are optimised and user-tested using the enhanced think-aloud protocol developed by Sunderland.** The toolkit is used across national and local government departments and agencies, including by HMRC, Ofsted, and the Department for Business, Energy and Industrial Strategy **[S4]**.

Impact on the Wellcome Trust

The former Lead User Researcher responsible for putting the research into practice at GDS moved to become the Head of User Experience at the Wellcome Trust in 2018. Since then, **she has introduced the “best practice” developed at Sunderland within her team of 11 user researchers “to enhance the rigour of their work” [S2]**. The team is responsible for the user experience of all the Trust's websites, including those for the Wellcome Collection (245,000 visits per month), and the Sanger Institute (273,000 visits per month). She concludes: *“Adopting this best practice means that participants provide better feedback; they provide the right amount, and they are less inclined to provide responses that they think will make the researcher happy. This improves the quality of data [which] improved the quality of the GOV.UK and Wellcome websites, [which] in turn ... improved users' experiences.” [S2]*

5. Sources to corroborate the impact (indicative maximum of 10 references)

- S1** [Registration page for public training](#) showing number of attendees
- S2** Written testimonial, Head of User Experience, Wellcome Trust; formerly User Research Lead, Government Digital Service
- S3** [Tweet by Head of User Research, Home Office Digital](#)
- S4** [Discussion board on GDS Github specifying how particular Design System components are being used across Government](#)
- S5** [User research service manual](#)
- S6** [Digital transformation in government. Report by the Comptroller and Auditor General. 2017](#)
- S7** [Government services using GOV.UK Notify](#)
- S8** [GOV.UK Pay performance dashboard](#)
- S9** [Government services using GOV.UK Verify](#)
- S10** [GOV.UK Design System](#)