Consigliere Evaluation: Evaluating Complex Interactive Systems with Users with Disabilities

Helen Petrie\textsuperscript{1(\$)}, Sanjit Samaddar\textsuperscript{1}, Christopher Power\textsuperscript{2}, and Burak Merdenyan\textsuperscript{1}

\textsuperscript{1} Department of Computer Science, University of York, York YO10 5GH, UK
{helen.petrie,sanjit.samaddar,burak.merdenyan}@york.ac.uk

\textsuperscript{2} School of Mathematical and Computational Science, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada
cdspower@upei.ca

Abstract. Conducting accessibility evaluations with users with disabilities is an important part of developing accessible interactive systems. Conducting such evaluations of systems which require complex domain knowledge is often impossible, as disabled users do not exist or are very rare. This paper presents a user evaluation method to address this problem, consigliere evaluation. A consigliere evaluation has the disabled user as the main participant, but they are assisted by an advisor, or consigliere, who understands the complex domain; the consigliere is in turn, monitored by an accessibility expert, who acts as an enforcer. As in all user evaluations, the disabled participant undertakes a series of tasks. But in a consigliere evaluation, if the task requires some particular domain knowledge or skill, the role of the consigliere is to provide appropriate information. However, it is easy for the consigliere, who usually does not have knowledge of the accessibility domain, to provide information not specifically about the domain, but about how to do the task in general. So the role of the enforcer, who is an accessibility expert, is to ensure this does not happen, and also to provide assistance and explanation if accessibility issues arise that the disabled participant cannot solve. The paper illustrates the consigliere method with a case study, the evaluation of Skillsforge, an online system used by a number of universities to manage progress of postgraduate students. This system requires considerable domain knowledge of the terminology and progression requirements for these students. It is used by university administrative staff, academic staff who supervise postgraduate students or are involved in monitoring students and the students themselves. The case study illustrates how the consigliere evaluation method works and some of the things which need to be considered in conducting the evaluation appropriately.

Keywords: Evaluation methods · Disabled users · Complex systems · User evaluations
1 Introduction

Conducting accessibility evaluations with users with disabilities is an important part of developing accessible interactive systems. Conducting such evaluations when the use of the system requires specialist knowledge poses particular problems. For example, to evaluate the accessibility of banking systems which are only used by highly skilled bank employees requires the user to have an in-depth knowledge of banking procedures and terminology. Finding sufficient numbers of users with disabilities for a reliable user evaluation is often difficult, and finding users with disabilities who have very specific and specialised domain knowledge in such situations is often impossible.

In response to this problem, we have developed a user evaluation method, which we have called the consigliere evaluation method, after the Godfather films. This paper will present the method and then discuss a case study of a recent evaluation conducted using the method in the evaluation of a new version of an interactive system to monitor the progress of research students at universities, to be used by university administrators, faculty members who supervise such students and the students themselves. The different views on the system used by the different user groups require different kinds of specialist knowledge about procedures and terminology within the institution in relation to research degrees and the students undertaking them.

2 Background

The most common approach to evaluating accessibility of interactive systems is to assess conformance to design guidelines. For web-based systems, the Web Content Accessibility Guidelines (WCAG) are available. There are also accessibility guidelines and design patterns for other kinds of interactive systems, for example smartphone apps, public access terminal such as automatic teller machines for banking and ticket machines for public transport, interactive voice systems and games consoles. Nonetheless, it is clear that developing interactive systems solely by following guidelines, while they are an important source of information and guidance, does not necessarily produce systems that are highly usable for disabled or older people.

Evaluation by target users of a system remains the “gold standard” of evaluation techniques. This is even more so in the case of the development of interactive systems for people with disabilities or older people, for a number of reasons, although see discussion in. Finding people with disabilities who can take part in user evaluations can often be difficult, however can be achieved with determination and persistence. However, if one is evaluating a system that requires specialist domain skills or knowledge, it can become impossible. It can be a “chicken and egg” problem: people with disabilities may not be able to develop the specialist skills if they
cannot use the system, but if there are no disabled people with the necessary skills, how can the specialist system be developed accessibly?

In order to address this methodological gap, we have developed the consigliere evaluation method, in which disabled people are the main participants in the user evaluation, but are assisted by an advisor, or consigliere [13], who in turn, is monitored by an accessibility expert, who acts as an enforcer. The next section explains how the method works, and the following section presents a case study of how we have used it for the evaluation of a system that required specialist knowledge of university procedures for dealing with the administrative needs of students on research degrees.

3 The Consigliere Evaluation Method

A consigliere evaluation is similar to other user-based accessibility evaluations, in that participants are asked to do a number of tasks which are designed to assess the accessibility of the system. They may be asked to do this with a concurrent verbal protocol, this works well in consigliere evaluations. They may be asked to complete standardized or bespoke questions about the accessibility and their user experience after each task and at the end of a set of tasks. The difference in consigliere evaluation is that each evaluation session requires three people, who play different roles:

- The participant with disabilities – this will be someone from the target user group, who may be a user of assistive technology, if so a competent user of that technology. However, they will not have any of the specialist domain knowledge required to use the system under evaluation
- The consigliere – this will be someone who has the specialist domain knowledge, most likely a non-disabled user of the technology or someone from the specialist domain of interest. They will probably not know anything about disability, accessibility or assistive technologies.
- The enforcer – this will be someone who is an expert in the accessibility of the kind of technology under evaluation, for example web accessibility, accessibility of Windows applications or mobile accessibility.

Before the session, the enforcer needs to brief the consigliere about how the session will work. The consigliere should provide the participant with the necessary specialist knowledge they need to interact with the system, this might be in terms of terminology and domain procedures. But they must not give any help in how to actually interact with the system to undertake the procedures and tasks. This may be hard for the consigliere because they have no knowledge of accessibility issues and they may inadvertently give messages that provide information to the participant about how to execute actions. So rather than saying “you now need to find function x” they might say “the link for function X is in the list headed Y”. The enforcer may need to take the consigliere through some examples of things to do and not to do and how to word advice.

At the beginning of the session the participant also needs to be briefed about the nature of the evaluation. They should be reassured that they are not expected to understand the domain of the system, and that they can ask the consigliere for any help,
although this will be moderated by the enforcer. During the evaluation session, it should be the participant who leads the dialogue – they should ask for advice whenever they feel they need it. The consigliere should only proactively offer advice when they think the participant has made a mistake due to a mis-understanding of domain knowledge.

During the session, particularly in the initial sessions, the consigliere should not offer advice without first checking with the enforcer. This may require a means of communication which the participant is not party to, for example hand gestures and messages on a notebook that the participant cannot see. This ensures that the enforcer can be sure that too much information is not communicated to the participant. The enforcer may also initiate support for the participant. Because the consigliere does not necessarily understand about accessibility issues, they may not realize whether the participant is having a problem due to a domain issue or an accessibility issue, so the enforcer may ask the consigliere to provide some domain knowledge. In addition, the enforcer can help when the participant gets stuck due to accessibility problems. Once a problem has been identified, there is no point leaving the participant to struggle with it, the enforcer can explain what the problem is and give assistance in getting over it.

This sounds very complicated, but with a little practice and cooperation, it becomes relatively easy and very interesting. It does mean that evaluation sessions, particularly the first ones, can take longer than usual, one needs to allow 25% to 50% longer. However, we have found that all parties involved find it a very rewarding and interesting experience. Participants learn something about a new domain, a new evaluation method and know they have contributed to the accessibility evaluation of an important system. Consiglieres (or consiglieri) learn a lot about accessibility and feel valued for their domain expertise. Enforcers learn about a new domain and often acquire more detailed accessibility information from the discussions with participants and consiglieres.

4 The Case Study: Consigliere Evaluation of Skillsforge

We recently used the consigliere evaluation method to assess the accessibility of a system used in a number of universities for the management of the progression and assessment of students doing postgraduate degrees, the Skillsforge system [13]. This system requires considerable domain knowledge of the terminology and progression requirements for these students. There are numerous processes which need to be logged in Skillsforge with different forms. This includes:

- Reporting on supervision meetings between students and supervisors (which need to be approved and signed by both student and supervisor, and if either edits the text, both need to sign off again);
- Scheduling and writing reports on Thesis Advisory Panel (TAP) meetings which consist of the student, supervisor and an independent Assessor (who also needs to be appointed and logged in the system);
- Logging that students have given seminars and submitted papers for publication (which need to be signed off by different people);
• Logging that students have submitted annual progress reports;
• Scheduling and writing reports on Progression meeting which consist of the student, the independent Assessor and a Progress Panel Chair (who also needs to be appointed and logged in the system);
• Appointing External Examiners for the examination of the thesis;
• Logging submission of the thesis.

As can be seen there are many different processes, with different people involved in particular processes. There are also different processes in each of the three or four years that a student is enrolled in their PhD so they cannot develop a rhythm of the processes, and students enroll at different times of the academic year, so supervisors find it difficult to develop a rhythm as well.

There are three separate main roles of users of the Skillsforge system:

• University administrators responsible for a group of students, in a department or unit. Typically, they need to ensure that all students are meeting the requirements for progression towards their degrees. They may need to monitor between 10 to 100 students.
• Supervisors, independent assessors and Progress Panel chairs. These are academic members of staff, they are responsible for one or more students they are supervising, plus one or more students for whom they are acting as independent assessor and Progress Panel chair. Thus, they may be responsible for one to several dozen students (the first author of this paper is responsible for 17 students as an academic)
• Research students. They are responsible to log their own progress and may use Skillsforge as a repository for documents for their PhD.

8 blind participants took part in the consigliere evaluation. They included 2 women and 6 men, aged 19 to 64. Some participants had been blind since birth, some had lost their sight at some later point in life. There was a mixture of users of different screenreaders, JAWS, NVDA and VoiceOver. All the screenreaders were used in the evaluation, as well as different browsers, including Chrome and Safari. Each participant evaluated all three roles in Skillsforge: university administrator, supervisor/independent assessor and PhD student. They completed 9 tasks, 3 tasks per role. Each evaluation session took between 90 min and 2 h.

The first, second and fourth author acted as the consigliere and enforcer in different parts of the evaluation. The first author is very experienced in the role of being a supervisor and independent assessor and the second and fourth authors were experienced at being PhD students. A member of the development team at Skillsforge also acted as consigliere, particularly for the university administrator role, which none of the authors were familiar with.

Here we illustrate some of the interesting situations which arose in these evaluations which highlight issues that need to be considered when conducting a consigliere evaluation:

Participant is confused about deadline numbers and why there is no deadline. The consigliere provides expert knowledge on the domain space without providing extra navigation help for the task. The consigliere provides domain expertise, and in turn the participant makes clear that even with the domain knowledge they would need a better heading.
P1: Deadline 2? Where is deadline 1?
Consigliere: Deadline 1 may have passed in this case [for this student] as of today’s date. Deadline 1 has already passed, so will not be available in the menu.
P1: Okay but I would expect a better heading then to explain what the deadline is for.

The *consigliere* should not give specific information that allows the participant to easily search for the solution to a task, sometimes this made things difficult. The *enforcer* decides when to help the participant. Here the *enforcer* has the knowledge to hint towards using different screenreader keywords, whereas the *consigliere* may have been tempted to ask the participant to search for “milestones”.

Participant was told to find “upcoming deadlines” although in Skillsforge they are called “milestones”. The phrasing of this task was intentionally chosen to avoid participants immediately using the search function and encourage them to navigate around the screen. In this instance the participant could not identify what was required:

P1: Ah yes, if I had known it was milestones I would have just searched for that after not finding anything on the page by tabbing through it.

Enforcer: Okay how about you look through the drop downs or headings?

In completing forms, the *consigliere* has to give detailed domain information about how fields should be completed in a form. The *consigliere* is explaining the process for a PhD student to request a leave of absence. However, he gave the information one form field at a time, means that participant could guess what type of form field to expect next. We then realised that we should give all the information at the beginning of the form, not necessarily in the order the participant will find them in the form, and tell the participant they can ask for the information to be repeated:

JAWS: Change summary stat, type edit text.
Consigliere: So here you can just say “requesting a leave of absence”
P1: Okay. (types into text field)
Consigliere: And I can tell you that this is a “long term temporary” type of absence
P1: So that’s the next field then! (Participant managed to fill this field in very quickly because he had been cued as to what came next, this was too much support).

The participant is asked to log into the system. The *consigliere* does not say what edit fields are required or the nature of the input type. For example, he does not say “Could you please type in your email and password”, as this would prompt participants to immediately use the form or edit shortcuts to find the sign in fields. However, in this case the participant explains that he would default to trying to find form fields anyway as the task is to sign in:

Consigliere: Please try logging in.
P4: One of the things with JAWS is that there is so much customization. So if I know I am looking for a form field to fill in, normally I would type ‘F’... and that seems to have worked. A lot of this is once you have got some clues, you don’t have to scramble around to find it.
5 Conclusions

We have presented a method to conduct accessibility evaluations with users of complex systems which require considerable domain knowledge. The consigliere evaluation method involves using two facilitators – a consigliere who provides domain knowledge and an enforcer who ensures that the consigliere does not provide too much information about interacting with the system and who also provides accessibility support for the participant and explanations to the consigliere. We have found this method very useful in conducting evaluations which have yielded much helpful information for developers of complex systems such as Skillsforge and an interesting and rewarding experience for the participants, the consiglieres and the enforcers.

Acknowledgements. We would like to thank the participants who took part in this evaluation, they helped us clarify the consigliere evaluation method. We would also like to thank Skillsforge for the opportunity to work with them and the Engineering and Physical Sciences Research Council (EPSRC) and the University of York for funding the research through the Innovation Voucher Programme.

References

1. World Wide Web Consortium – Web Accessibility Initiative (W3C WAI): Web Content accessibility Guidelines (WCAG) 2.1 (2018). https://www.w3.org/TR/WCAG21/. Accessed 12 Apr 2020
2. Android Developers. Make apps more accessible. https://developer.android.com/guide/topics/ui/accessibility/apps. Accessed 12 Apr 2020
3. Apple Human Interface Guidelines. Accessibility. https://developer.apple.com/design/human-interface-guidelines/accessibility/overview/introduction/. Accessed 12 Apr 2020
4. Centre for Excellence in Universal Design. Public access terminals, 1 July 2019 http://universaldesign.ie/technology-ict/irish-national-it-accessibility-guidelines/public-access-terminals/public-access-terminals.html. Accessed 12 Apr 2020
5. Centre for Excellence in Universal Design. Telecoms. http://universaldesign.ie/technology-ict/irish-national-it-accessibility-guidelines/telecoms/telecoms.html. Accessed 12 Apr 2020
6. AbleGamers. Design patterns for APX. https://accessible.games/accessible-player-experiences/design-patterns/. Accessed 12 Apr 2020
7. Sani, Z.H.A., Petrie, H.: User evaluation of an app for liquid monitoring by older adults. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2017. LNCS, vol. 10279, pp. 86–97. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58700-4_8
8. Power, C., Freire, A., Petrie, H., Swallow, D.: Guidelines are only half the story: accessibility problems encountered by blind users on the web. In: Proceedings of the 30th International Conference on Human Factors in Computing (CHI 2012). ACM Press, New York (2012)
9. Petrie, H., Bevan, N.: The evaluation of accessibility, usability and user experience. In: Stephanidis, C. (ed.) The Universal Access Handbook. Taylor and Francis, London (2009)
10. Sears, A., Hanson, V.L.: Representing users in accessibility research. ACM Trans. Access. Comput. 4(2), 7 (2012)
11. Aizpurua, A., Arrue, M., Harper, S., Vigo, M.: Are users the gold standard for accessibility evaluation? In: W4A: Proceedings of the 11th Web for All Conference (2014). Article 13
12. Power, C., Petrie, H.: Working with participants. In: Yesilada, Y., Harper, S. (eds.) Web Accessibility. HIS, pp. 153–168. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7440-0_9
13. https://skillsforge.com/