

Usability evaluation of a co-created big data analytics platform for health policy-making

Brian Cleland¹, Jonathan Wallace¹, Raymond Bond¹, Salla Muuraiskangas², Juha Pajula³, Gorka Epelde^{4,5}, Mónica Arrúe^{4,5}, Roberto Álvarez^{4,5}, Michaela Black⁶, Maurice D. Mulvenna¹, Deborah Rankin⁶, Paul Carlin⁷

¹ School of Computing, Ulster University, Northern Ireland

{b.cleland, jg.wallace, rb.bond, md.mulvenna}@ulster.ac.uk

² VTT Technical Research Centre of Finland Ltd, Wellness and Living, Oulu, Finland
salla.muuraiskangas@vtt.fi

³ VTT Technical Research Centre of Finland Ltd, Smart Health, Tampere, Finland
juha.pajula@vtt.fi

⁴ eHealth and Biomedical Applications, Vicomtech, Donostia/San Sebastián, Spain

⁵ IIS Biodonostia, Donostia/San Sebastián, Spain

{gepelde, marrue, ralvarez}@vicomtech.org

⁶ School of Computing, Engineering & Intelligent Systems, Ulster University, Northern Ireland
{mm.black, d.rankin1}@ulster.ac.uk

⁷ South Eastern Health & Social Care Trust, Ulster Hospital, Dundonald, Northern Ireland
paul.carlin@setrust.hscni.net

Abstract. The increasingly important role of big data in organisational decision-making brings with it significant challenges in terms of designing usable software interfaces. Specifically, such interfaces must allow users to explore, analyse, and visualise complex data from heterogeneous sources and derive insights to support management decisions. This paper describes a usability evaluation of the MIDAS Project, a big data platform for health policy-making, developed by an EU-funded Horizon 2020 project involving a number of international partners and pilot sites. We describe how a combination of heuristic and formative user-centred evaluation methods were employed, and give a summary of the key findings. We discuss key insights from the evaluation, including the importance of having diverse users, the role played by users' prior expectations, and the logistical challenge of coordinating user testing across multiple sites. Finally, we explore the relative value of each of the evaluation methods, and outline how our approach to usability testing will evolve for future iterations of the MIDAS platform.

Keywords: Big Data, Health, Usability, Evaluation, Policy-making, User-centred.

1 Introduction

The impact of big data has been increasingly felt across many industry sectors in recent years. Its specific impact on the health sector was examined in 2013 in a report by McKinsey [1], which outlined several key factors that were likely to drive changes in

healthcare policy and delivery. The first insight was that demand was growing for greater economic value to be generated from data, including improvements to policy-making processes, which should be linked to existing stored datasets that had not been re-used (that is, used for secondary purposes). The second factor was the expanding volumes of data being collected from organisational sources such as clinical devices and processes, as well as personal sources such as social media and mobile devices. The third point was the steady growth of open data being published by governments in order to catalyse private sector innovation. Finally, it was noted that technological change was driving predictable improvements in terms of the capture, storage and processing of data from diverse sources.

It was in this context that the MIDAS Project (Meaningful Integration of Data Analytics and Services) was funded under the EU's Horizon 2020 programme in order to capitalise on big data trends and drive policy improvements in the European health sector. The project is intended to meet the needs of citizens and policy-makers through the provision of a state-of-the-art big data platform. In this paper we present a usability study of the first iteration of the MIDAS platform, focusing specifically on the dashboard layer, which is designed to facilitate the analysis and visualisation of big data for policy-making. We begin by providing a brief overview of the project and the technical architecture of the platform, including the data integration, analytics and visualisation layers. We then explain the usability evaluation methods that were used and present a summary of the main findings from the study. Finally, we discuss the most important implications of the results, review which parts of the methodology were most useful, and outline how the usability analysis of future software development iterations will be carried out in light of the findings.

2 Background

2.1 MIDAS Project Overview

The MIDAS Project consortium consists of a variety of academic, industrial and policy experts from six European countries and the United States. A key focus of the project is ensuring that the information delivered by the MIDAS platform is relevant and actionable, and to this end the consortium has adopted a user-centred design ethos. A broad selection of experts from a diverse spectrum of backgrounds were chosen in order to effectively capture user needs and provide realistic engagement in the co-design of the system [2]. This approach has been driven by a realisation that ultimate users of the system are likely to be very diverse, and will include stakeholders with very different backgrounds and levels of technical knowledge - from statisticians and data scientists to senior civil servants and politicians.

The core elements of the MIDAS platform will include technical components to enable the collection and preparation of heterogeneous data, and an agreed architecture for data storage, data integration, data virtualization, data cleansing, deployment and management. The secure management of personal data is a key deliverable, and will be addressed through the development of a 'privacy by design' approach. This approach will enable the anonymisation and aggregation of data, while simultaneously making it

interoperable to the analytics layer. The platform will also incorporate a variety of tools and algorithms that will allow meaning to be extracted from data, building on existing machine learning and other analytics methods. From the user's perspective, a key part of the platform will be the user interface (UI) layer that will draw aggregated outputs from the analytics layer for display via management dashboards and easy-to-use interactive tools.

2.2 MIDAS Platform Architecture

As stated above, a key goal of the MIDAS project is to access and integrate a range of heterogeneous datasets, located in many separate systems. A common metadata model will enable these datasets to be harmonised so that common analytics and processing can occur, but a flexible underlying data architecture is required to support the storage and processing of these disparate datasets, as well as integration of that data from diverse sources. The MIDAS platform will be based on Analytics Engines XDPTM, which is a proprietary, scalable and modular data analytics platform and Analytics Engines Unified Data View (UDV), which enables a single, unified view of data being analysed. This forms the core data storage/processing portion of the platform, upon which the analytics and visualisation layers are built.

The architecture of the MIDAS Dashboard is shown in Fig. 1. It is composed of three main layers: the User Interface, a Middleware layer and a Closed Intranet layer. The User Interface is developed with standard HTML, Bootstrap, JQuery and JavaScript technologies. It has a modular design, which will enable any third party to develop their own visualization with JavaScript. OpenVA is a Java framework developed by VTT, which is used to handle the results of the analytics that are received from the analytics layer. A Flask-based REST server manages communications between the user interface and the analytics system, allowing the frontend to retrieve specific parameters or analytics results.

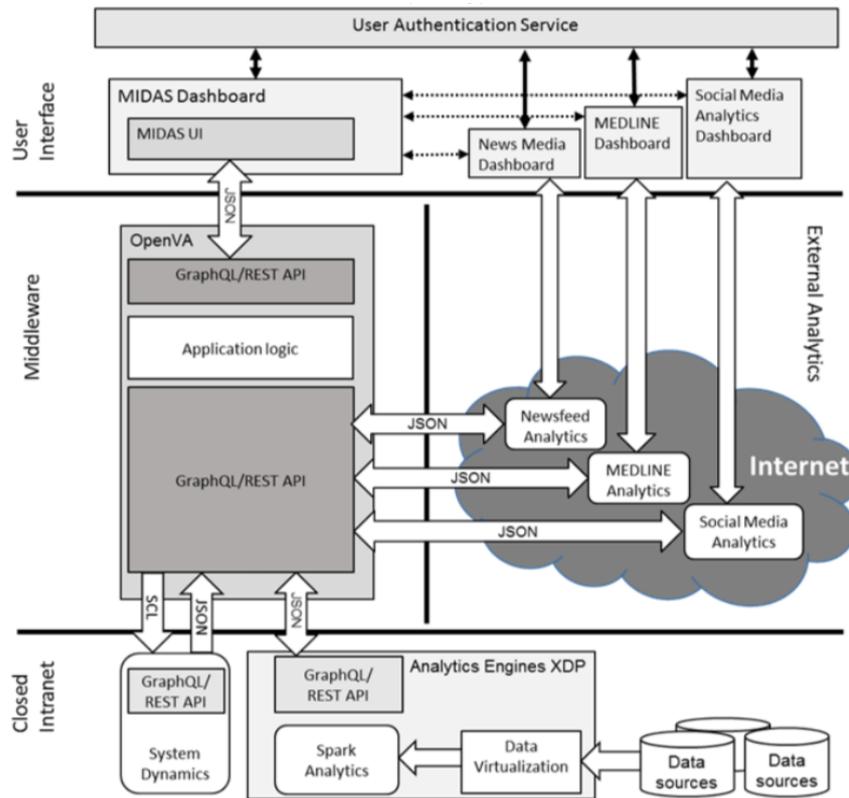


Fig. 1. MIDAS Dashboard architecture.

2.3 MIDAS Interface

In this section we describe the main components in the MIDAS Platform User Interface which were the primary focus for the evaluation.

MIDAS Dashboard. The MIDAS Dashboard has two views for the end user: the sign in page and main page. The main page is only shown to the user after they have successfully signed in. The three primary components of the main page are the menu bar at the top of the page, the name bar below the menu bar, and the workspace for the widgets. The workspace can extend outside the displayed area, and when this happens scrollbars appear on bottom and right border. This is normally triggered when there are more widgets than will fit on the current screen. An example of the MIDAS Dashboard is shown in Fig. 2.

Social Media Widget. Trying to get a sense of how the public feel about a new policy can be a challenge. The MIDAS social media platform is designed to allow policy-makers to create chatbots to ask questions on their behalf. Responses are analysed using natural language processing techniques and the resulting insights can be displayed via a social media widget. This widget will enable policy-makers to get a sense of how the public feel about specific policy matters by visualising the sentiment detected in the public's responses. This will indicate, for example, whether people are responding positively or negatively to individual policy proposals.

MEDLINE Widget. The MEDLINE custom widget is an interactive visual tool that helps to surface information by re-indexing search results based on user input. More specifically, the user drags a cursor over a graph showing precomputed clusters of search terms, which triggers updates to the list of relevant documents. For example, when we enter a search term 'diabetes type 2', the system performs an elasticSearch over the MEDLINE dataset and extracts groups of keywords that best describe different subgroups of results. By moving the cursor over word-groups, the user provides the relevance criteria, thereby bringing to the top of the list the most relevant articles. The user can read its title and first lines of each abstract, and by clicking on it, open that article in the browser at its PubMed location.

External Platforms. A number of linked external analytics resources will be available to later versions of the MIDAS Dashboard. These were not included in the current evaluation study but will be included in the future UX testing iterations when the connectivity with external components has been improved.

3 Methodology

3.1 Heuristic Evaluation

This first stage of the protocol was to carry out a heuristic evaluation. This involves an expert using a series of established usability and design principles to audit the user interface. In this case, Ulster University UX lab experts used well-established usability principles such as Jakob Nielsen's 10 heuristics [3], laws from Gestalt psychology [4] and the 8 golden rules by Ben Shneiderman [5]. The test team included three specialists with extensive experience of UX testing. Each of the experts applied the whole protocol to the current system, and after these tests the severity of the problems were evaluated. Following the evaluation, usability issues were set out in a series of slides for sharing with the development team. These slides also included suggestions to improve the system. Most findings were addressed according to the suggestions prior to the formative evaluation taking place.

3.2 Formative Evaluation

For the formative evaluation a number of subjects were recruited to attempt a series of tasks using the MIDAS prototypes. It had been agreed at the outset of the MIDAS Project that the platform would be tested and validated with real data and representative users of the system from across the partner countries. Twelve users in total were selected from the Basque Country, Finland, Northern Ireland and Republic of Ireland. These users were from a diverse range of technical and policy backgrounds. Although the final test protocol (see below) focused primarily on two main user personas - a data scientist and a policy-maker, it was decided that each user should perform the tasks associated with both personas in order to maximise data collection. The specific tasks for each persona are presented in Table 1.

Table 1. Tasks for each persona for usability testing protocol.

Persona	Task No.	Task Description
Data Scientist	1	Create new dashboard and give it a name
Data Scientist	2	Add an analytics widget using a specific dataset
Data Scientist	3	Add a MEDLINE search widget
Data Scientist	4	Add a social media widget
Data Scientist	5	Rearrange the dashboard layout and save your changes
Data Scientist	6	Share the dashboard with another user
Policy Maker	1	Make sense of the dashboard whilst thinking aloud
Policy Maker	2	Add 3rd widget to the dashboard, i.e. social media widget

A rigorous test protocol was jointly developed by consortium members, led by usability testers from Ulster University's UX Lab. The usability testing protocol was informed by Ulster's UX-Lab having carried out a range of usability tests on medical devices, software and data visualisations [6-16] (see Fig. 3). Each subject attempted the task whilst 'thinking-aloud'. The Think-Aloud Protocol (TAP) [17] allows the assessor to understand the user's cognitive processes hence eliciting usability issues and cognitive errors. Before and after each task, the user was asked how difficult they expected the task to be and how difficult the task was, thus measuring whether the system met the user's expectation. This is known as the Single Ease Question (SEQ) [18]. A questionnaire based on the System Usability Scale (SUS) [19] was also given to each user in order to measure perceived usability. A so-called SUS score was computed and benchmarked using a SUS distribution. In addition, after the usability test, we measured the frequency of usability errors/issues, task completion rates, and task completion times.

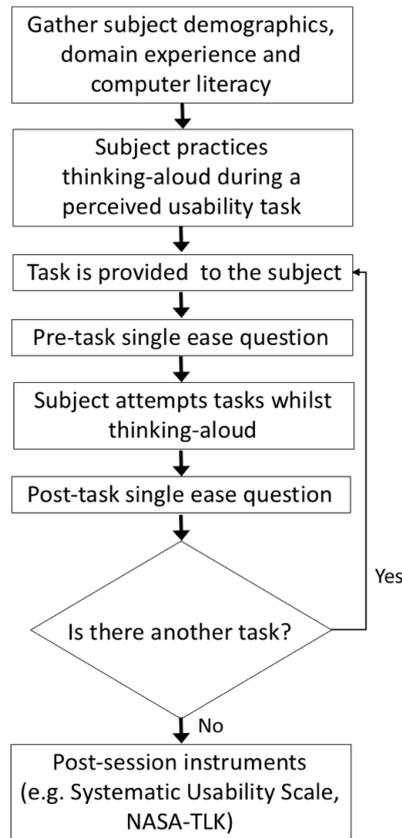


Fig. 3. Typical usability testing protocol.

Testing sessions using the shared protocol took place in Northern Ireland (including Northern Ireland and Republic of Ireland users), Finland and the Basque Country. Notes were taken by moderators or a separate note taker during the sessions and video was recorded for subsequent detailed analysis. The testing data was gathered and shared on a Google Doc, and the Ulster University team carried out an analysis of the aggregated results. This analysis includes summaries of user demographics, task completion rates and times, SEQ (Single Ease Question) scores for each task, and SUS (System Usability Scale) scores for each persona. It also included a summary and analysis of the qualitative comments and insights gathered during the testing sessions. The findings of the UX Evaluation Report were shared with developers to facilitate bugs fixes and enhance features, and to drive usability improvements as the project progressed.

4 Results

4.1 Heuristic Evaluation Results

The primary aim of heuristic evaluation in this study was to ensure that the system was ready for formative testing. Experts evaluated the MIDAS Dashboard interfaces to find the most critical problems and root causes for issues which might confuse end users. Most issues were corrected before end user representatives began the formative testing phase. Based on the heuristic evaluation the technical team fixed a number of issues, including visual fixes on boxes and text alignment, adding better search functionality, including a help menu on the menu bar, reordering menu items, and adding a “save as” option. Some problems and suggested fixes were not corrected before the formative evaluation due to time constraints, including a “forgot password” button for sign-in, better registration information on the sign-in page, an incorrect password lock on sign-in, and coloured sharing symbols for the dashboard menu.

4.2 Formative Evaluation Results

In this section we summarise the key findings and recommendations from the UX Evaluation Report.

Task Completion Times. An analysis of the task completion rates and times showed that the majority of users were able to complete the majority of tasks without assistance. With the exception of Data Scientist Tasks 2 and 5, and Policy Maker Task 1, the majority of tasks were completed in less than a minute. Data Scientist Tasks 2 and 5 raised a number of specific issues that are explored in more detail in the Qualitative Analysis (below), and Policy Maker Task 1 was an exploratory task, for which the time taken was not a particularly relevant metric.

SEQ Results Summary. An analysis of the Single Ease Question (SEQ) scores showed that in general, the platform was slightly easier to use than anticipated. One exception to this was Data Scientist Task 5, which users found more difficult than they expected. This finding is reflected in the longer task completion times, and also in the issues raised in the Qualitative Analysis. For the Policy Maker persona, the SEQ scoring indicated that user felt that Task 1 - the interpretation of shared dashboards - was not at all straightforward (scoring on average only 3/7). On the other hand, Task 2 - modifying an existing dashboard - was perceived to be relatively easy (scoring on average 6/7).

SUS Results Summary. In the context of the SUS (System Usability Scale) scores, statements 8 and 10 were outliers for both the Data Scientist and Policy Maker persona tasks. Statement 8 relates to how “cumbersome” users find the interface, although it is important to note that some studies have suggested that this score can be affected by language factors with non-native English speakers [19]. The Statement 10 score suggests that users found the system easy-to-learn. Another interesting finding from the SUS analysis is that the platform scored better for the Data Scientist persona than for the

Policy Maker persona (75.0 vs. 56.7), implying that users perceived that policy makers might find the current prototype more challenging to use, and in particular to interpret visualisations. Figs. 5 and 6 show the distribution of the aggregated results for the Data Scientist and Policy Maker tasks.

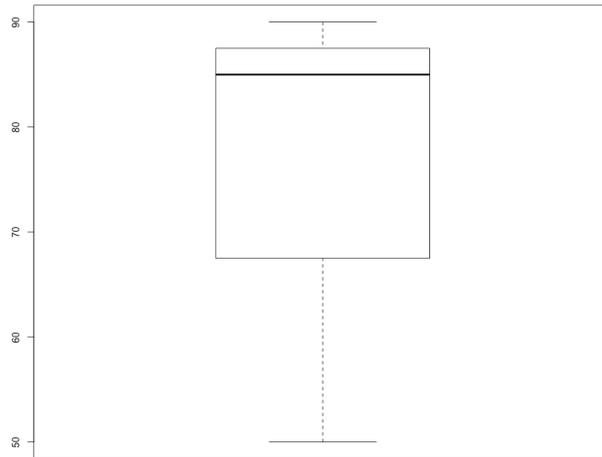


Fig. 4. Data Scientist SUS results (box plot showing overall scores for all tasks).

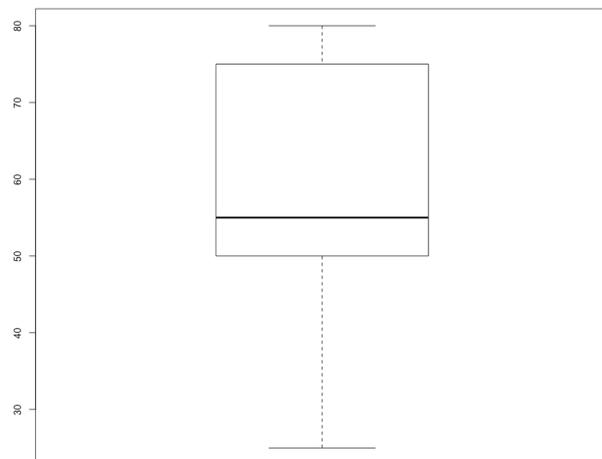


Fig. 5. Policy Maker SUS results (box plot overall scores for both tasks).

Qualitative Results Summary. The qualitative data gives perhaps the richest understanding of the issues and insight that emerged during testing. Despite the diverse nature of the users involved in the usability testing, there were a number of clear recurring themes in the user feedback across all sites. Perhaps the most commonly identified issue was the need for more information to guide users through the system. Some options identified included more descriptive text in the menus and wizards, more hover text and pop-up hints and tips, making videos available on how to create, share and interpret dashboards, and rich user documentation. Dashboard sharing was generally seen to be a key feature for development. Many users suggested that the platform should support annotations and comments for widgets and dashboards, as well as chat functionality for dashboard creators to assist policy makers in interpreting the data. Sharing dashboards would also require some consideration of the need for effective version control, as well as user search functionality.

Alongside these and other high-level feature requests, there were numerous observations on how the interaction with individual widgets and menu items might be improved. Additionally, one user who had spent a lot of time training clinicians on how to use dashboards felt that the model used by the system conflicted with her expectations. For example, this user expected the system to provide the ability to have a number of different tabs or dashboards open simultaneously. While this user's expectations were tied to a specific product that she had personally spent a number of years working with, it raised an important question about the extent to which the MIDAS design paradigm would align with user expectations based on established products and services.

5 Discussion

As stated earlier, the usability evaluation was comprised of two main parts - a heuristic analysis and a formative analysis. The advantages of heuristic evaluations are that they are relatively fast and inexpensive to deploy [20]. They also allow problems to be identified early in the design cycle, when they are less costly to fix. One drawback is that usability experts can be hard to come by [20], although this did not apply in the case of the MIDAS Project. Most scholars agree that heuristic evaluations are best seen as a complement rather than an alternative to user testing [21]. For this reason, we decided to employ both heuristic and formative methods to the evaluation. By employing the heuristic method first, and fixing as many of the identified problems before the formative evaluation, the intention was to ensure that the maximum value was derived from the user testing. This was particularly important given the logistical challenges of engaging with relevant users from the various MIDAS sites, which in many cases required relatively senior staff giving up significant amounts of time to assist with the testing protocol.

The user-centred formative evaluation phase involved a variety of users from different backgrounds carrying out task associated with two key user personas - the Data Scientist and Policy Maker. The findings showed that there was very little overlap in results compared to the heuristic evaluation, which was perhaps not surprising given the fact that developers had already addressed most of the major concerns that had been

raised by the usability experts. Task completion times were generally satisfactory, and the Single Ease Questions (SEQ) for the tasks showed that overall the system was slightly easier to use than most users had anticipated. Both the SEQ and the System Usability Scale (SUS) results indicated that the application was harder to use from a Policy Maker perspective, compared to a Data Scientist perspective.

Arguably the most valuable feedback came from the qualitative results, which were enabled by adopting a Think Aloud Protocol approach to the testing. The qualitative feedback enabled the testers to identify key issues which were common to all testing sites. These included the need for more metadata and guidance via the user interface, which many users initially found somewhat confusing. In addition to the need to make functionality more discoverable, the complexity of the underlying datasets meant that users often required help in understanding the information they were attempting to use. Another challenge was raised by one user who had already built up significant experience of an alternative dashboard-based analytics product. In their case, the UX paradigm adopted by the MIDAS platform did not align with prior expectations. Ultimately, the researchers were able to achieve a compromise proposal in terms of how dashboards should be presented to users. This does raise an interesting question for developers, however: how can they balance the need to follow established industry practices and user expectations, with the desire to pursue more innovative and perhaps “superior” interface solutions?

6 Conclusions

Although research indicates that heuristic evaluation results tend to uncover similar problems to more user-centred methods, it is generally accepted that they should be seen as a complement to usability testing [21]. In order to reduce the potential for duplicated effort in this study, the heuristic evaluation was completed and the software was updated before the formative user-centred evaluation was carried out. This approach meant that the majority of concerns raised during the formative testing were newly identified. Testing with a variety of users across multiple sites allowed the researchers to identify recurring issues that should be prioritised for future development. While the task completion times, SEQ scores and SUS scores were useful for generating quantitative usability measures, it is worth noting that the most valuable feedback for the development team came from the qualitative feedback. This qualitative approach and analysis of user comments as they spoke about their experience using the platform provided the most valuable insights in terms of determining how ongoing development work should be targeted. This is perhaps reflective of the fact that software usability is by its nature a complex construct, requiring a depth of analysis that can only come from a qualitative approach.

It should be borne in mind that this evaluation study of the MIDAS platform was focused on an early prototype, which entailed incomplete functionality and limited access to client data. For future iterations of the platform and usability testing, it is anticipated that the functionality of the application will be significantly extended, and a richer ecosystem of datasets will be available for analysis. It is likely that large parts of

the existing test protocols will be re-used for subsequent evaluation exercises. One important lesson learned through this process was the scale of the challenge in recruiting suitable users and coordinating times to attend lab-based tests. It was clear that future test events should be scheduled well in advance, preferably two to three months beforehand. Future test protocols will be adapted to focus on the specific needs of each MIDAS pilot site, and are expected to involve a richer user interface that will include external data sources and visualisation tools.

Acknowledgements

The MIDAS Consortium gratefully acknowledge the support to this project from the European Union research fund 'Big Data Supporting Public Health Policies', under Grant Agreement No. 727721 (H2020-SC1-2016-CNECT SC1-PM-18-2016).

References

1. Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S.: The 'big data' revolution in healthcare. *McKinsey Quarterly*, 2(3). (2013)
2. Cleland, B., Wallace, J., Bond, R., Campus, J., Ireland, N., Black, M., ... & Connolly, P.: Meaningful Integration of Data Analytics and Services in MIDAS Project: Engaging Users in the Co-Design of a Health Analytics Platform.. 10.14236/ewic/HCI2018.169. (2018)
3. Nielsen, J.: 10 usability heuristics for user interface design. Nielsen Norman Group, 1(1). (1995)
4. Mazumder, F. K., & Das, U. K.: Usability guidelines for usable user interface. *International Journal of Research in Engineering and Technology*, 3(9), 79-82. (2014)
5. Shneiderman, B.: Designing for fun: how can we design user interfaces to be more fun?. *interactions*, 11(5), 48-50. (2004)
6. Bond RR, Finlay DD, Nugent CD, Moore G, Guldenring D. A usability evaluation of medical software at an expert conference setting. *Computer methods and programs in biomedicine*. (2014).
7. Bond R, Mulvenna M, Wallace JG, Finlay DD, Boyd K, Patterson M. UX-handle: a cloud-based system to streamline usability testing analytics. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference* (p. 2). BCS Learning & Development Ltd. (2017)
8. Boyd K, Nugent C, Donnelly M, Sterritt R, Bond R. A usability protocol for evaluating online social networks. In *International Conference on Smart Homes and Health Telematics* (pp. 222-225). Springer, Berlin, Heidelberg. (2012)
9. Torney H, O'Hare P, Davis L, Delafont B, Bond R, McReynolds H, McLister A, McCartney B, Di Maio R, McEneaney D. A usability study of a critical man-machine interface: Can layperson responders perform optimal compression rates when using a public access defibrillator with automated real-time feedback during cardiopulmonary resuscitation? *IEEE Transactions on Human-Machine Systems*. 46(5):749-54. (2016)
10. Boyd K, Nugent C, Donnelly M, Bond R, Sterritt R, Hartin P. An investigation into the usability of the STAR training and re-skilling website for carers of persons with dementia. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 4139-4142). IEEE. (2014)

11. Gibson A, McCauley C, Mulvenna M, Ryan A, Laird L, Curran K, Bunting B, Ferry F, Bond R. Assessing usability testing for people living with dementia. In Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques (pp. 25-31). ACM. (2016)
12. Bond R, O'Hare P, Di Maio R. Usability testing of a novel automated external defibrillator user interface: A pilot study. In Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference (pp. 1486-1488). IEEE. (2015)
13. Boyd K, Bond R, Gallagher S, Moore G, O'Kane E. Usability and behaviour analysis of prisoners using an interactive technology to manage daily living. In Proceedings of the 31st British Computer Society Human Computer Interaction Conference (p. 80). BCS Learning & Development Ltd. (2017)
14. Bond RR, Finlay DD, McLaughlin J, Guldenring D, Cairns A, Kennedy A, Peace A. Human factors analysis of a novel engineering solution to the problem of electrode misplacement during 12-lead electrocardiogram acquisition. *Journal of electrocardiology*. 49(6):933. (2016)
15. Fidler R, Bond R, Finlay D, Guldenring D, Gallagher A, Pelter M, Drew B, Hu X. Human factors approach to evaluate the user interface of physiologic monitoring. *Journal of electrocardiology*. 48(6):982-7. (2015)
16. Bond RR, Van Dam E, Van Dam P, Finlay DD, Guldenring D. Evaluating the human-computer interaction of 'ECGSim': A virtual simulator to aid learning in electrocardiology. In Computing in Cardiology Conference (CinC), (pp. 409-412). IEEE. (2015)
17. Peute, L. W., de Keizer, N. F., & Jaspers, M. W.: The value of Retrospective and Concurrent Think Aloud in formative usability testing of a physician data query tool. *Journal of biomedical informatics*, 55, 1-10. (2015)
18. Sauro, J.: 10 things to know about the Single Ease Question (SEQ). *Measuring U*, (2012)
19. Finstad, K.: The system usability scale and non-native english speakers. *Journal of usability studies*, 1(4), 185-188. (2006)
20. Nielsen, J. Chapter 2: Heuristic Evaluation. In *Usability Inspection Methods*, Nielsen, J. and Mack, R.L., Eds. John Wiley and Sons, New York, pp. 25-62. (1994)
21. Paz, F., Paz, F. A., Villanueva, D., & Pow-Sang, J. A.: Heuristic Evaluation as a Complement to Usability Testing: A Case Study in Web Domain. In 2015 12th International Conference on Information Technology - New Generations (pp. 546-551). Las Vegas, NV, USA: IEEE. <https://doi.org/10.1109/ITNG.2015.92> (2015)