Human-computer interaction in radiology

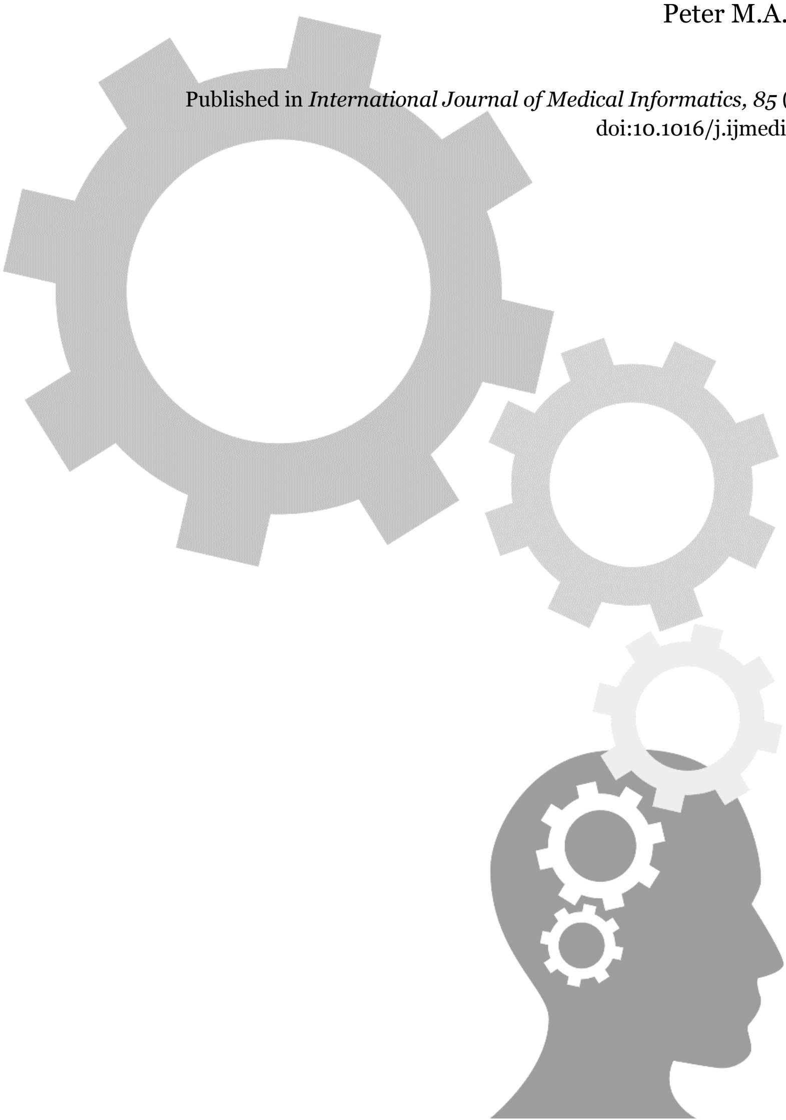
Jorritsma, Wiard

3

Post-deployment usability evaluation of a radiology workstation

Wiard Jorritsma
Fokie Cnossen
Rudi A. Dierckx
Matthijs Oudkerk
Peter M.A. van Ooijen

Published in *International Journal of Medical Informatics, 85* (2016), 28–35.
doi:10.1016/j.ijmedinf.2015.11.001



Abstract

Objectives: To determine the number, nature and severity of usability issues radiologists encounter while using a commercially available radiology workstation in clinical practice, and to assess how well the results of a pre-deployment usability evaluation of this workstation generalize to clinical practice.

Methods: The usability evaluation consisted of semi-structured interviews and observations of twelve users using the workstation during their daily work. Usability issues and positive usability findings were documented. Each issue was given a severity rating and its root cause was determined. Results were compared to the results of a pre-deployment usability evaluation of the same workstation.

Results: Ninety-two usability issues were identified, ranging from issues that cause minor frustration or delay, to issues that cause significant delays, prevent users from completing tasks, or even pose a potential threat to patient safety. The results of the pre-deployment usability evaluation had limited generalizability to clinical practice.

Conclusions: This study showed that radiologists encountered a large number and a wide variety of usability issues when using a commercially available radiology workstation in clinical practice. This underlines the need for effective usability engineering in radiology. Given the limitations of pre-deployment usability evaluation in radiology, which were confirmed by our finding that the results of a pre-deployment usability evaluation of this workstation had limited generalizability to clinical practice, it is vital that radiology workstation vendors devote significant resources to usability engineering efforts before deployment of their workstation, and to continue these efforts after the workstation is deployed in a hospital.

The big difference between good and bad designers is how they handle people struggling with their design.

– Joshua Porter

Introduction

To ensure that software has high usability (i.e. allows users to perform their tasks effectively, efficiently and enjoyably), usability engineering methods should be applied throughout the software development cycle: before design, during design and implementation, and after deployment of the software [1]. However, a recent survey conducted among usability practitioners showed that usability activities are performed less frequently during the post-deployment phase than in other phases of development [2]. While pre-deployment usability activities are of course important, and can prevent usability issues in the post-deployment phase, learning from real users using the software in the real world is also valuable [1–5]. Since iteration and user feedback are core components of usability engineering [1,6], it is unfortunate that these data are being underused.

Post-deployment usability evaluation can be especially useful in domains such as radiology, where it is difficult to perform valid pre-deployment usability evaluations for two reasons. First, it is difficult to obtain a representative user group, because radiologists with different specializations work in very different ways, radiologists are relatively expensive and they often have limited time to spare for usability activities. Second, it is difficult to construct a representative testing environment, because the radiological workflow is complex and differs between hospitals and radiology software needs to communicate with multiple interdependent systems (e.g. the hospital information system, imaging modalities), which also differ between hospitals. This means that even when radiology software vendors apply usability engineering methods before and during design and implementation of their software, it is still likely that usability issues occur when the software is used in clinical practice.

Comparative usability evaluations of commercial radiology workstations indicate that there is indeed room for usability improvement [7–9]. However, because the main goal of these studies was to compare the workstations based on a quantitative representation of usability, they provide limited insight into the nature and severity of usability issues radiologists encounter with the workstations. These studies also face the same limitations as the workstation vendors (the difficulty to obtain a representative user group and a representative testing environment for usability evaluation), which raises questions about the generalizability of their results.

Other usability studies in radiology do provide more detailed qualitative usability information [10–13]. However, three of these studies [10,12,13] evaluated systems that do not play a critical role in the radiological workflow and one [11] only performed a heuristic evaluation, which means that no actual users took part in the evaluation. Also, all of these studies were conducted in laboratory settings, so they do not provide information about usability issues encountered in clinical practice.

In this study, we performed a post-deployment usability evaluation of the radiology workstation that received the highest usability rating in the comparative pre-deployment evaluation of Jorritsma et al. [7]. We aimed to determine the number, nature and severity of usability issues radiologists encounter while using this workstation in clinical practice, and to assess how well the results of the pre-deployment usability evaluation of this workstation generalize to clinical practice.

Methods

Apparatus

In this paper, the term 'radiology workstation' refers to the software radiologists use for viewing medical images and reporting their diagnosis. The radiology workstation evaluated in this study consisted of three components: an image viewer (the client for the Picture Archiving and Communication System (PACS)), which included standard post-processing capabilities, a workflow manager (the client for a rudimentary version of the Radiology Information System (RIS)) and a report editor with speech recognition. Technically, these are separate applications, but since they form one integrated whole sold by one vendor as a single package, we consider them here as one system.

Fig. 1 shows the workstation setup used in our hospital. A 30.4″ diagnostic

Figure 1. The radiology workstation setup used in our hospital. A 30.4″ diagnostic monitor (left) displays the image viewer and a 20.1″ monitor (right) displays the workflow manager and the report editor.

monitor displays the image viewer and a 20.1″ monitor displays the workflow manager and the report editor. Some radiologists have a third monitor that they use to display an image archive of the patient (an overview of all studies of the patient). The input devices are a standard mouse, a keyboard and a handheld speech microphone.

The workstation was deployed in our hospital fourteen months before the start of this study, replacing a workstation of a different vendor that had been in place for about thirteen years.

The workstation vendor applied usability engineering methods throughout the development of the workstation. These methods included pre-design user research, empirical evaluation of prototypes, and iterative design.

The workstation was included in the pre-deployment comparative usability evaluation of Jorritsma et al. [7], in which it was referred to as PACS B. Of the four workstations evaluated in this study, this workstation was found to have the best usability.

Participants

Twelve users participated in the usability evaluation. Ten were radiologists (mean years of experience: 6.1, range: 0.6 – 27; specializations: thoracic/cardiac; mammo/abdominal/paediatric; paediatric; thoracic/abdominal; musculoskeletal; abdominal; IC/mammo/thoracic; paediatric/oncologic; 2 neurological) and two were radiology residents (a first- and a third-year resident).

Three users did not work with the workstation that was previously in place in our hospital, but did have experience with workstations from different vendors.

Users were randomly selected from the total group of radiologists and radiology residents in our hospital until a test group of twelve users was formed. During this process one radiologist and one radiology resident refused to participate because they were too busy at the time.

Design and procedure

Sessions were held with individual users at their own workplace. A session lasted approximately 45 to 60 minutes and consisted of a brief semi-structured interview and an observation of the users during their daily work. The interview aimed to assess the users' general opinions about the workstation and to identify the main usability issues encountered. The following questions were used to guide the interview: "how satisfied are you with the workstation (on a scale of 1–10)?", "what is better/worse in this workstation compared to the previous one (or in case the user did not work with the previous workstation: "other workstations you have worked with)?" and "what problems do you encounter?".

After the interview, users were asked to review a number of studies from their current worklist that were as representative of their daily work as possible. They were instructed to vocalize the steps they were taking (think-aloud protocol) and to illustrate usability issues as they were encountered. If users' comments or behavior required clarification, the session observer would prompt them to explain themselves in more detail. A video recording of the screens was made during the observation. Audio was recorded throughout the session.

Informed consent was obtained. Users could give separate permission for the use of their data for this study (inclusion criterion), audio recording during the session, and video recording during the observation. All users gave permission for all three items.

All data collection sessions were conducted by one usability expert (the first author). Because users performed naturally occurring rather than controlled, pre-defined tasks, we did not obtain quantitative measures of effectiveness and efficiency. Instead these usability aspects were evaluated based on the qualitative data.

Data analysis

The audio and video recordings were synchronized. The audio recordings were transcribed and the fragments describing usability issues were marked and given an issue number (i.e. each issue was identified based on users' verbal data). For each issue, a time stamp was added to the transcript so that it could easily be found later in the video recordings. Excerpts were taken from the video recordings to illustrate the issues that were difficult to understand by description alone. A general description of all unique issues was put in a spreadsheet. Positive usability findings were also documented.

Each usability issue was given a severity rating according to a rating system based on those described by Rubin and Chisnell [14] and Tullis and Albert [15]: severity = impact on the user experience (1 = low, 2 = medium, 3 = high) + predicted frequency of occurrence (1 = low, 2 = medium, 3 = high). The issues were rated by one usability expert (author 1). We used the ratings of one expert rather than a consensus rating of multiple experts because we could not find other usability experts with sufficient domain knowledge. There are not a lot of usability experts with experience in radiology, let alone with PACS/RIS.

After the study, the list of usability issues was reviewed together with application specialists of the workstation vendor and the workstation's key users (users that received extensive training on the workstation in order to be able to support the general users). No changes to the severity ratings were suggested and no changes were deemed necessary based on the reviewing sessions.

The identified issues were divided into five categories: image arrangement/display, image interaction, workflow, report dictation, and miscellaneous. These categories were determined post hoc based on clusters in the data.

The issues were also categorized according to their root cause. The possible causes were interface design, configuration (i.e. how the workstation is set up in our hospital), technical (i.e. bugs or performance issues), functional (i.e. missing functionality), and external (i.e. issues that users experience in the workstation, but are caused by an external process (e.g. the workstation's display protocols fail with images from other hospitals, because these images are not correctly archived by the hospital's PACS administrators)). The root causes were determined post hoc together with the workstation vendor and the workstation administrators in our hospital.

In order to assess how well a pre-deployment usability evaluation of a radiology workstation generalizes to clinical practice, we compared the results of Jorritsma et al.'s pre-deployment evaluation [7] to our post-deployment results. Although this pre-deployment evaluation was summative (i.e. it was designed to quantitatively evaluate the workstation's usability and not to identify specific usability issues), it did make two specific claims about the workstation's usability that we could compare against the results of our post-deployment evaluation: (1) users can effectively and efficiently perform image retrieval tasks due to the workstation's effective display protocols[1], and (2) users can efficiently perform measurement tasks because they can measure in images without having to click on them first, and because measurement tools stay active until another tool is selected. The generalizability of these claims was evaluated against the related usability issues and positive usability findings identified in the post-deployment evaluation.

These claims were based on several image retrieval and measurement tasks users performed as part of different clinical scenarios. These scenarios covered a range of diagnostic questions, body parts and imaging modalities.

Results

Overall, users were moderately satisfied with the workstation, giving it a mean rating of 6.7 out of 10 (range: 6 – 8). Fifteen positive usability findings were identified, ranging from useful functionalities to interface elements that enable fast workflow patterns.

[1] The display protocols eliminate the need to manually select the imaging studies to retrieve for a given patient, because they automatically retrieve all relevant images.

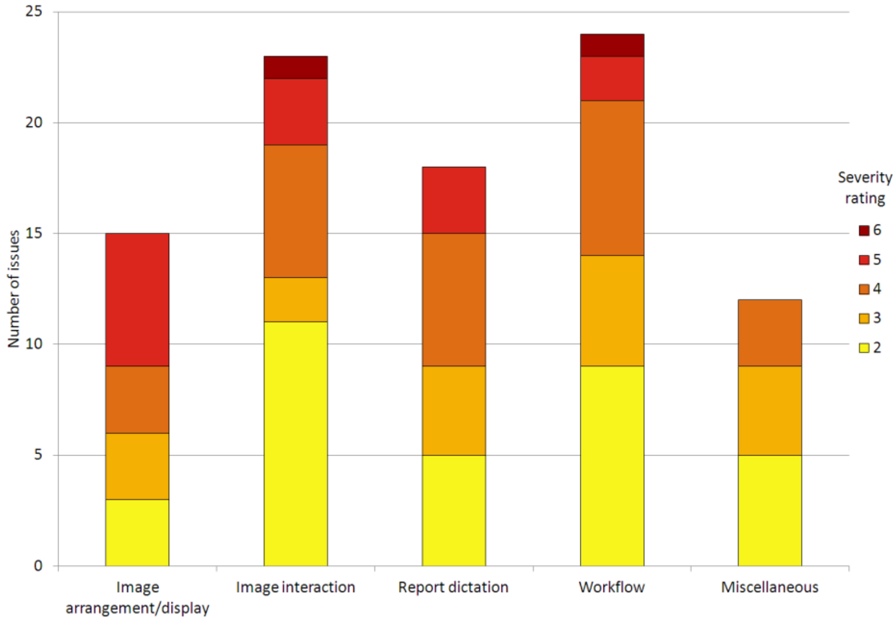


Figure 2. The number of usability issues in each category, broken down according to their severity rating.

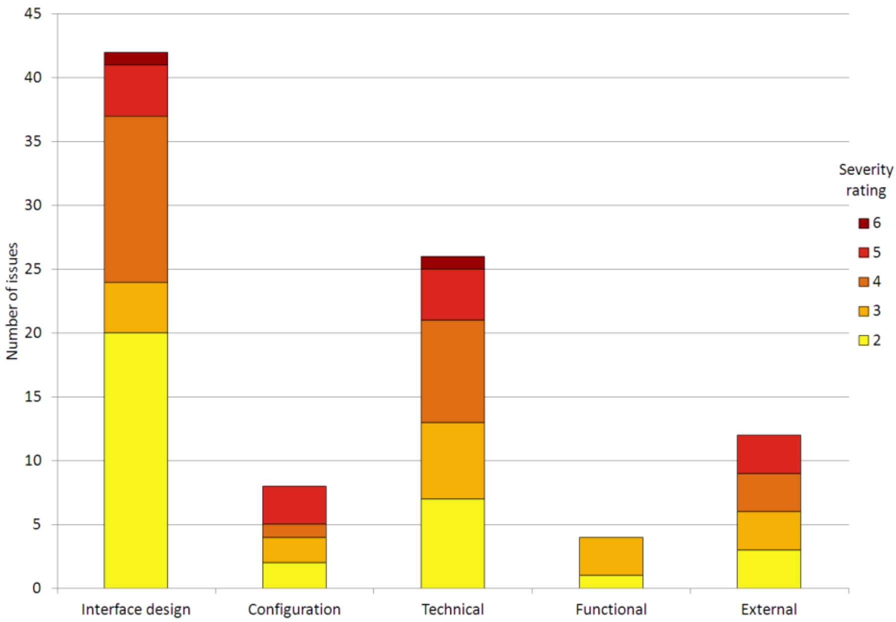


Figure 3. The number of usability issues for each root cause, broken down according to their severity rating.

Fig. 2 shows the number of usability issues in each category, broken down according to their severity. Ninety-two usability issues were identified, ranging from issues that cause minor frustration or delay, to issues that cause significant delays, prevent users from completing tasks or even pose a potential threat to patient safety. The mean number of identified issues per user was 15.4 (range 6 – 22). Examples of the usability issues are shown in Table 1. The issues were selected such that all categories and root causes were represented, the selection contained issues of different severity ratings, and the issues were understandable without in-depth knowledge of the specific workstation we evaluated. Note that a 'study' is a collection of one or more series of images obtained during one scanning procedure.

Fig. 3 shows the number of usability issues for each root cause, broken down according to their severity rating. Most issues were due to interface design or technical problems. A substantial number of issues were not due to the implementation of the workstation itself, but to poor configuration of the workstation in our hospital or external processes that negatively affected the workstation's usability.

Jorritsma et al.'s pre-deployment evaluation [7] made two specific claims about the workstation's usability that we could compare against the results of our post-deployment evaluation: (1) users can effectively and efficiently perform image retrieval tasks due to the workstation's effective display protocols, and (2) users can efficiently perform measurement tasks because they can measure in images without having to click on them first, and because measurement tools stay active until another tool is selected. Table 2 shows the usability issues and positive usability findings related to these tasks (retrieving patient images and performing measurements) for both the pre-deployment and the post-deployment evaluation. This allowed us to assess how well the pre-deployment results generalize to clinical practice. The pre-deployment evaluation revealed that users could effectively and efficiently perform the image retrieval tasks due to the workstation's effective display protocols. The post-deployment evaluation revealed that the display protocols are indeed effective in some cases, but are ineffective in others due to several severe usability issues. None of these issues were due to the implementation of the workstation itself, but were caused by poor configuration of the workstation and external processes.

In the pre-deployment evaluation, the fact that measurement tools remain active until another tool is selected was considered a positive finding, because it allowed users to quickly perform consecutive measurements. In the post-deployment evaluation on the other hand, this was considered a usability issue, because users were frustrated by the fact that they have to deactivate the measurement tool after each single measurement, and that they often perform accidental measurements when they click on an image without realizing the measurement tool was still active.

Table 1. Examples of usability issues.

Issue	Severity	Category	Root cause
The highlighted study in the patient archive strip does not necessarily correspond to the study that is in the viewport above the strip. Because a study is highlighted in the strip, users might think that the image above the strip belongs to this study, while it actually belongs to a different study. This could cause users to base their report on the wrong images.	5	Image arrangement /display	Interface design
Display protocols treat digital radiography (DX) and computed radiography (CR) studies as different studies, while they are equivalent to the user. This can cause them do display a previous study of an incorrect date (e.g. if the current study is a DX thorax, a display protocol would display a previous DX thorax next to it, while there is a more recent CR thorax available). If users are unaware of this, they compare the current study to the wrong previous study.	5	Image arrangement /display	Configuration
Scrolling through stacked images is inconvenient. Scrolling with the scroll wheel and right mouse button is too slow. Scrolling with the scroll bar is faster but awkward, because it requires users to locate and move their mouse to the scroll bar handle, which is very small for large series, each time they want to scroll.	6	Image interaction	Interface design
Registration (which allows users to scroll through multiple stacked image series in synchrony) sometimes fails and produces an error message. Sometimes two series cannot be registered in one interface tab, but can, and are even automatically registered, in another.	5	Image interaction	Technical
The clinical information has to be copied to the report manually. This cannot be done automatically because the reports are not stored properly in the system from which the workstation retrieves them.	5	Report dictation	External
The speech recognition often misses words. This can lead to serious errors in the report (e.g. "metastases" instead of "no metastases").	5	Report dictation	Technical
There is no 'undo' possibility in the report. When users want to correct a word, they sometimes accidentally select the entire sentence or paragraph. If they then dictate the word they want to correct, the	4	Report dictation	Interface design

entire sentence or paragraph is gone. This mistake cannot be undone.			
When a study's acquisition status has not been set to 'completed', users can not view the images. Although it is good to have a safeguard against reporting on incomplete studies, radiologists should at least be allowed to view the images. In the current situation, clinicians *can* view these images through another application, but radiologists, whose job it is to view images, cannot view them on the workstation.	5	Workflow	Interface design
Users can not store studies in personal lists. This would be useful to save interesting studies, or to check the results of biopsies.	3	Workflow	Functional
Some functions that users add to the customizable toolbar tab/right click menu disappear when a new study is opened.	3	Miscellane-ous	Technical

Table 2. Usability issues (-) and positive usability findings (+) related to retrieving patient images and performing measurements identified in the pre- and post-deployment usability evaluations of the workstation.

Task	Pre-deployment evaluation	Post-deployment evaluation
Retrieving patient images	+ Effective display protocols.	+ Display protocols often correctly display the relevant previous study of the patient. + Users can easily switch between different tabs within a display protocol, each containing a different set of images (e.g. a tab with images of the left breast and a tab with images of the right breast). - There are no appropriate display protocols for some types of studies (e.g. brain MRI and musculoskeletal studies) (severity = 5, cause = configuration). - Display protocols treat DX and CR studies as different studies, while they are equivalent to the user (severity = 5, cause = configuration). - Display protocols do not recognize relevant images within a study that consists of multiple body parts (e.g. CT abdomen images within a CT total body study) (severity = 5, cause = external). - Display protocols do not recognize studies received from other hospitals (severity = 5, cause = external). - If there is a study (e.g. a lumbectomy)

				between the current and the previous mammogram, the mammography display protocol displays this study instead of the previous mammogram (severity = 3, cause = configuration).
Performing measurements	+	Users can perform measurements on images without having to click on them first.	-	Measurement tools stay active until another tool is selected (severity = 5, cause = configuration).
	+	Measurement tools stay active until another tool is selected.	-	Measurements cannot be made by click-and-drag (in addition to click-and-click) (severity = 2, cause = interface).

Discussion

In this study, we performed a post-deployment usability evaluation of a radiology workstation. We aimed to determine the number, nature and severity of usability issues radiologists encounter while using this workstation in clinical practice, and to assess how well the results of a pre-deployment usability evaluation of this workstation generalize to clinical practice.

Our usability evaluation revealed a large number and wide variety of usability issues with the workstation. These issues spanned all aspects of interaction with the workstation (image arrangement/display, image interaction, report dictation, and workflow) and ranged from issues that cause minor frustration or delay, to issues that cause significant delays, prevent users from completing tasks, or even pose a potential threat to patient safety. This underlines the need for effective usability engineering in radiology.

Most of the issues were due to interface design or technical problems. Addressing these issues requires changes to the implementation of the workstation. A substantial number of issues on the other hand were not due to the implementation of the workstation, but to poor configuration of the workstation in our hospital or external processes that negatively affected the workstation's usability. Addressing these issues would significantly improve the workstation's usability without the need to make any changes to its implementation.

In the pre-deployment usability evaluation of Jorritsma et al. [7], the workstation performed well on image retrieval tasks, which was attributed to its effective display protocols. Our post-deployment evaluation showed that the display protocols are indeed effective in some cases, but ineffective in others due to a variety of usability issues. These issues were not due to the workstation itself, but to the display protocol configuration in our hospital and external processes that negatively affect the functioning of the display protocols (e.g. incorrect archiving of images received from other hospitals).

The pre-deployment evaluation contained two tasks where the display protocols dealt with computed and digital radiography studies in a correct way. However, these types of studies caused a usability issue in clinical practice. This shows that the display protocol configuration for these types of studies was actually better during a pre-deployment test than during routine clinical use.

Due to the wide variety of radiological examinations and protocols that are used in our hospital, the correct configuration of the display protocols is far more complex than for the limited set of a pre-deployment test. This caused configuration of non-optimal or even faulty display protocols during the workstation vendor's efforts to tailor the display protocol set to the needs of the hospital's radiologists (which is done for every hospital in which the workstation is deployed).

The positive pre-deployment results on the measurement tasks were partly attributed to the fact that measurement tools stay active until another tool is selected, which allows users to quickly perform consecutive measurements. Interestingly, this feature was considered a usability issue in the post-deployment evaluation, because users were frustrated by the fact that they have to deactivate the measurement tool after each single measurement, and that they often perform accidental measurements when they click on an image without realizing the measurement tool was still active.

A reason for this discrepancy is that the pre-deployment evaluation considered the measurement tasks in isolation and did not consider the transition from a measurement task to a different task. This emphasized the positive aspect of the continued activation of the measurement tools (fast consecutive measurements), and underemphasized the negative aspects (the user action required to deactivate the measurement tool and the accidental measurements).

Also, the task set used in the pre-deployment evaluation consisted of an equal number of tasks involving consecutive measurements and tasks involving a single measurement. This might not have been representative of clinical practice (even though the tasks were developed in consultation with a group of radiologists).

The reason that user's subjective responses in the pre-deployment evaluation did not reflect this issue could be that it simply did not occur in the limited time frame of the evaluation, it did not occur frequently enough to bother them, or because they blamed accidental measurements on themselves because it was their first time using this workstation.

Since the activation behavior of the measurement tools can be customized, users can decide for themselves whether the positive consequences of the continued activation outweigh the negative. However, since users tend to underuse interface customization facilities [16] (none of the users in our study were even aware that they could change the activation behavior of the measurement tools), it is important to choose a default setting that is optimal for the majority of users. Log data of users interacting with the workstation in our hospital [17] show that single measurements are more common than consecutive measurements, which means that the optimal setting would be to deactivate the measurement tools after one use.

These findings indicate that the results of a pre-deployment usability evaluation of a radiology workstation have limited generalizability to clinical practice. This does not mean that they do not provide meaningful insights into the workstation's usability, but they paint an incomplete picture that has to be supplemented with data from actual use in a real-world environment. It is therefore important that radiology workstation vendors devote resources to post-deployment usability evaluations, in addition to pre-deployment evaluations, in order to identify usability issues that slip through the pre-deployment phase.

This is in line with the recommendations of Chilana et al. [2], who found that the frequency of usability activities in current practice decreases after product deployment and proposed a focus on 'usability maintenance', in parallel to general software maintenance, in order to enhance the post-deployment user experience based on the actual use of a product and to provide ongoing support for usability issues that arise in practice.

A limitation of our study is that we compared the results of a formative post-deployment usability evaluation to the results of a summative pre-deployment evaluation, and not also to the results of a formative pre-deployment evaluation. One could therefore argue that the discrepancy between the pre- and post-deployment results was due to the methodological differences between the evaluations alone. However, if the pre-deployment evaluation had been formative, this would not have changed its fidelity (i.e. it would not have been more representative of clinical practice). Therefore, all issues caused specifically by the workstation's integration in a clinical environment (i.e. issues caused by configuration and external processes), which account for six of the seven post-deployment issues that undermined the validity of the claims made in the pre-deployment evaluation, could still not have been identified.

Another consequence of this limitation is that it means that our study did not allow us to quantitatively compare the pre- and post-deployment evaluations. It would be interesting to compare our results to the results of a formative pre-deployment evaluation in order to compare the number of issues identified in each evaluation.

It would also be interesting to compare our post-deployment results to the results of the pre-deployment evaluations carried out by the workstation vendor. This would allow us to determine why the vendor did not find the usability issues identified in our study. It could also be the case that the vendor did find the issues, but did not think it was worth the effort to solve them. Unfortunately we could not gain access to the data of the vendor's usability evaluations, or to the details of their methods.

A limitation of our study is that it only included twelve users. Although the number of unique usability issues identified tends to decrease with each additional user that participates in the evaluation [18], we did not experience such diminishing returns, probably due to the heterogeneity of the user group. We did experience diminishing returns for the issues with a severity rating higher than 4. Although more issues could have been identified if more users participated in the evaluation, we believe that the size of our user group was sufficient to demonstrate the volume and variety of usability issues radiologists encounter in clinical practice.

While there are differences in usability requirements between radiologists with different specialties (e.g. an abdominal radiologist cares about navigating through stacks of images, while a mammography radiologist does not) and it would be interesting to study these differences in more detail, we did not have enough users from each specialty to present any meaningful data on these differences.

Another limitation is that we used the severity ratings of a single usability expert, rather than a consensus rating of multiple experts. However, the fact that no changes to the ratings were proposed or deemed necessary in the review sessions with the application specialists of the workstation vendor and the workstation's key users indicates that the ratings were reasonable.

Another limitation is that we only evaluated one workstation. It could be argued that our results are unique to this specific workstation and that other workstation would produce more positive results. While we do not deny that there could be workstations with better usability than this one, we doubt that any of them has perfect usability. Comparative usability evaluations of radiology workstations indeed showed that all workstations had room for usability improvement [7–9]. The fact that this workstation received the highest usability rating in the comparative evaluation of Jorritsma et al. [7] also shows that it is definitely not an inferior system (on the contrary). We therefore believe that our usability evaluation of this workstation successfully demonstrates the importance of usability engineering, both before and after deployment, of radiology workstations in general.

We invite other researchers to perform similar usability evaluations of different workstations in order to determine 'universal' issues and best practices that can inform the design of any workstation. It would also be interesting to compare the results of usability evaluations of the same workstation at different hospitals in order to determine how contextual factors (e.g. radiological workflow, IT infrastructure, demographics of the radiologists) affect the usability issues encountered with the workstation.

We believe that our results are not limited to the radiology workstation, but can be generalized to any system that is used by a heterogeneous user group in a complex environment. Our study supports results from previous studies that demonstrated the added value of post-deployment usability evaluation [1,4,5].

Conclusion

This study showed that radiologists encounter a large number and a wide variety of usability issues when using a commercially available radiology workstation in clinical practice. This underlines the need for effective usability engineering in radiology. Given the limitations of pre-deployment usability evaluation in radiology, which were confirmed by our finding that the results of a pre-deployment usability evaluation of this workstation had limited generalizability to clinical practice, it is vital that radiology workstation vendors devote significant resources to usability engineering efforts before deployment of their workstation, and to continue these efforts after the workstation is deployed in a hospital.

References

[1] J. Nielsen, The usability engineering life cycle, Computer 25 (1992) 12–22.

[2] P.K. Chilana, A.J. Ko, J.O. Wobbrock, T. Grossman, G. Fitzmaurice, Post-deployment usability: a survey of current practices, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst., ACM Press, Montreal, Canada, 2011: pp. 2243–2246.

[3] D.M. Nichols, D. Mckay, M.B. Twidale, Participatory Usability: supporting proactive users, in: Proc. 4th Annu. Conf. ACM Spec. Interes. Gr. Comput. Interact., Dunedin, New Zealand, 2003: pp. 63–68.

[4] C.M. Nielsen, M. Overgaard, M.B. Pedersen, J. Stage, S. Stenild, It's worth the hassle! The added value of evaluating the usability of mobile systems in the field, in: Proc. 4th Nord. Conf. Human-Computer Interact., ACM Press, Oslo, Norway, 2006: pp. 272–280.

[5] H.B.-L. Duh, G.C.B. Tan, V.H. Chen, Usability evaluation for mobile device: a comparison of laboratory and field tests, in: Proc. 8th Conf. Human-Computer Interact. with Mob. Devices Serv., ACM Press, Espoo, Finland, 2006: pp. 181–186.

[6] J.D. Gould, C. Lewis, Designing for usability: key principles and what designers think, Commun. ACM. 28 (1985) 300–311.

[7] W. Jorritsma, F. Cnossen, P.M.A. van Ooijen, Merits of usability testing for PACS selection, Int. J. Med. Inform. 83 (2014) 27–36.

[8] T. Boehm, O. Handgraetinger, J. Link, R. Ploner, D.R. Voellmy, B. Marincek, et al., Evaluation of radiological workstations and web-browser-based image distribution clients for a PACS project in hands-on workshops, Eur Radiol. 14 (2004) 908–14.

[9] N. Bazak, G. Stamm, F. Caldarone, J. Lotz, A. Leppert, M. Galanski, PACS workstations 2000: evaluation, usability and performance, in: 18th Int. EuroPACS Conf., Graz, Austria, 2000: pp. 133–142.

[10] D. Markonis, M. Holzer, F. Baroz, R.L.R. De Castaneda, C. Boyer, G. Langs, et al., User-oriented evaluation of a medical image retrieval system for radiologists, Int. J. Med. Inform. (2015). doi:10.1016/j.ijmedinf.2015.04.003.

[11] E. Nabovati, H. Vakili-Arki, S. Eslami, R. Khajouei, Usability evaluation of Laboratory and Radiology Information Systems integrated into a hospital information system., J. Med. Syst. 38 (2014) 35.

[12] I. Geldermann, C. Grouls, C. Kuhl, T.M. Deserno, C. Spreckelsen, Black box integration of computer-aided diagnosis into PACS deserves a second chance: results of a usability study concerning bone age assessment., J. Digit. Imaging. 26 (2013) 698–708.

[13] W. Jorritsma, J.T. Prins, P.M.A. van Ooijen, Comparing four touch-based interaction techniques for an image-based audience response system, Int. J. Hum. Comput. Interact. 31 (2015) 440–450.

[14] J. Rubin, D. Chisnell, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, Wiley Publishing, Inc., New York, NY, 1994.

[15] T. Tullis, B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, Morgan Kaufmann, San Fransisco, CA, 2008.

[16] W. Jorritsma, F. Cnossen, P.M.A. van Ooijen, Adaptive support for user interface customization: a study in radiology, Int. J. Hum. Comput. Stud. 77 (2015) 1–9.

[17] W. Jorritsma, F. Cnossen, R. Dierckx, M. Oudkerk, P.M.A. van Ooijen, Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client, Int. J. Med. Inform. 85 (2016) 26–42.

[18] R.A. Virzi, Refining the test phase of usability evaluation: how many subjects is enough?, Hum. Factors. 34 (1992) 457–468.