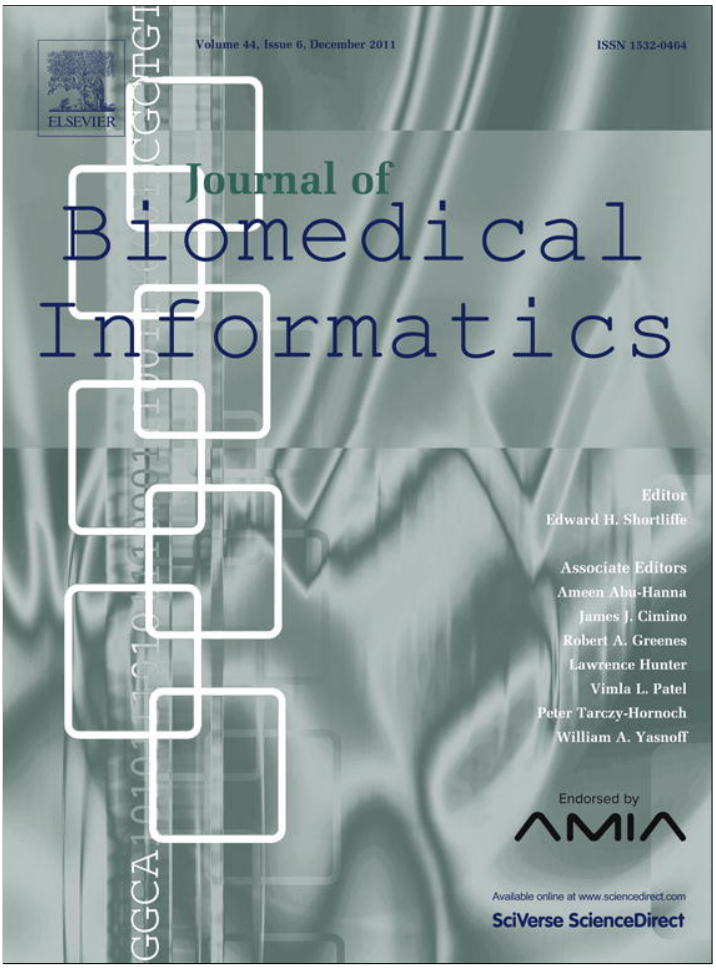

Journal of Biomedical Informatics 44 (2011) 948–957

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Classification and prioritization of usability problems using an augmented classification scheme

R. Khajouei [a,b,*,1], L.W.P. Peute [a,1], A. Hasman [a], M.W.M. Jaspers [a,*]

[a] Department of Medical Informatics, Academic Medical Center, University of Amsterdam, PO Box 22700, 1105 AZ Amsterdam, The Netherlands
[b] Kerman University of Medical Sciences, Kerman, Iran

ARTICLE INFO

Article history:
Received 30 November 2010
Accepted 4 July 2011
Available online 12 July 2011

Keywords:
Classification
User–computer interface
Usability
Medical order entry systems

ABSTRACT

Various methods exist for conducting usability evaluation studies in health care. But although the methodology is clear, no usability evaluation method provides a framework by which the usability reporting activities are fully standardized. Despite the frequent use of forms to report the usability problems and their context-information, this reporting is often hindered by information losses. This is due to the fact that evaluators' problem descriptions are based on individual judgments of what they find salient about a usability problem at a certain moment in time. Moreover, usability problems are typically classified in terms of their type, number, and severity. These classes are usually devised by the evaluator for the purpose at hand and the used problem types often are not mutually exclusive, complete and distinct. Also the impact of usability problems on the task outcome is usually not taken into account. Consequently, problem descriptions are often vague and even when combined with their classification in type or severity leave room for multiple interpretations when discussed with system designers afterwards. Correct interpretation of these problem descriptions is then highly dependent upon the extent to which the evaluators can retrieve relevant details from memory.

To remedy this situation a framework is needed guiding usability evaluators in high quality reporting and unique classification of usability problems. Such a framework should allow the disclosure of the underlying essence of problem causes, the severity rating and the classification of the impact of usability problems on the task outcome. The User Action Framework (UAF) is an existing validated classification framework that allows the unique classification of usability problems, but it does not include a severity rating nor does it contain an assessment of the potential impact of usability flaws on the final task outcomes. We therefore augmented the UAF with a severity rating based on Nielsen's classification and added a classification for expressing the potential impact of usability problems on final task outcomes. Such an augmented scheme will provide the necessary information to system developers to understand the essence of usability problems, to prioritize problems and to tackle them in a system redesign. To investigate the feasibility of such an augmented scheme, it was applied to the results of usability studies of a computerized physician order entry system (CPOE).

The evaluators classified the majority of the usability problems identically by use of the augmented UAF. In addition it helped in differentiating problems that looked similar but yet affect the user–system interaction and the task results differently and vice versa. This work is of value not only for system developers but also for researchers who want to study the results of other usability evaluation studies, because this scheme makes the results of usability studies comparable and easily retrievable.

1. Introduction

Health care information systems have the potential to support clinicians in making clinical decisions and hence to improve patient safety [1] but clinicians' reluctance to use these systems in daily clinical care has become a well known dilemma [2]. Studies have shown that usability problems are among the factors negatively affecting system's acceptance [3,4] and limiting their effectiveness in supporting and streamlining clinical care [5–8]. Usability is a quality attribute that assesses how easy user interfaces can be used. Usability problems can be defined as aspects of a user interface that may cause the corresponding system to have a decreased usability for the end user.

So far, research on health care information systems' usability focused on the identification of specific problems that compromise

* Corresponding authors. Addresses: Faculty of Management and Health Information Sciences, Haft-bagh Highway, Postal Code 7616911313, Kerman, Iran, fax: +98 341 320 5221 (R. Khajouei). Department of Medical Informatics, Academic Medical Center, PO Box 22700, 1105 AZ Amsterdam, The Netherlands, fax: +31 20 691 9840 (M.W.M. Jaspers).

E-mail addresses: r.khajouei@yahoo.com (R. Khajouei), m.w.jaspers@amc.uva.nl (M.W.M. Jaspers).

[1] These authors contributed equally to the paper.

1532-0464/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2011.07.002

effective, efficient and safe use of these systems [7,9–11]. Evaluators use usability evaluation methods to detect usability problems. In doing so, evaluators first observe the potential problem and then record the problem for further analysis. Despite the frequent use of forms to report the usability problems and their context-information, problems are formulated in the evaluator's own terms and problem reports are composed of only that what he perceives as salient about the problems at that moment in time. Hence, evaluators' descriptions of detected usability problems differ in specificity and completeness and are often inconsistent among evaluators. Correct interpretation of these problem descriptions is then highly dependent upon the extent to which the evaluators can retrieve relevant details from memory when they discuss these reports with system developers before planning (re)design efforts [12]. Given that between the evaluation and the redesign efforts often a considerable time lapse occurs, preventing ambiguity and information losses in reporting on usability data is relevant. To communicate the identified usability problems, a framework is needed to support consistent, accurate and complete reporting of results.

Usually evaluators cluster problem descriptions into unique problem types (bottom-up analysis). The determination whether different usability problem descriptions are referring to the same or different underlying usability problem is usually done later in time by expert judgment. Hence, bottom-up clustering of usability problems afterwards is subject to variability due to inconsistency in interpretation and grouping, and is deficient in providing adequate information for finding solutions for these problems.

An alternative way is to classify problems in terms of recognized usability principles, the so-called heuristics [13] (top-down approach). However, it has been shown that various difficulties arise when these heuristics are used as a classification scheme due to their incompleteness, lack of mutual exclusiveness, and lack of specificity [14]. Heuristics are therefore not sufficiently comprehensive to be used for complete and accurate reporting and systematic classification of usability problems. Moreover, a heuristic classification does not provide insight in the underlying causes of the usability problem types reported. As a consequence, detected usability problems might be typified as similar at a high aggregation level, although the underlying causes could be fundamentally different.

Some studies use severity ratings to classify usability problems. The severity rating is commonly based on Nielsen's classification [13] grounded on the proportion of users who (will) experience a specific problem, the impact it (will) have on them, and whether the usability problem will be a problem only the first time the users encounter it, or whether it will persistently bother them.

Usability problems not only bother users during interaction with the health care information system but also have an impact on their task performance. Therefore, these problems can be a source of errors [8], potentially compromising patient safety. Hence, in prioritizing redesign efforts the potential of usability problems to evoke user errors, for example leading to wrong medication orders, should be taken into account. Although the severity rating prioritizes the problems in terms of their effect on the user interaction, it does not consider the impact of these problems on the task outcomes.

It is clear that there is a need for a framework with which usability problem descriptions can be classified to ensure accurate, complete and consistent problem reporting. Consistent and detailed reporting of not only the usability problems but also their underlying causes, their impact on the user and their potential effect on the final outcomes are needed for guiding and prioritizing system redesign efforts. In addition, the availability of such a consistent and detailed reporting is a prerequisite for the development of a knowledge base concerning usability problem characteristics and their impact on users and task outcomes. Such a knowledge base allows researchers and system designers to learn from the errors made by their colleagues, because they are able to retrieve the relevant information using the classification of the problem type they are interested in. Such an encompassing framework is not yet available.

In this paper we introduce a classification scheme for usability problems in the domain of health care that will significantly reduce the problems mentioned above. It should satisfy the following requirements:

- Rely on theory from the domain of usability engineering: a classification scheme should include a robust taxonomy that allows the assignment of the problems to the different stages of user interaction and describe the path via which a design flaw of a system (cause of the problem) results in a usability problem. This scheme should allow consistent problem reporting and tracking of the usability problems through successive stages of iterative (re)design.
- Rank the effects of the problems on the physical and cognitive effort required from the user: Users in the health care domain are very busy and often have to take care of different issues simultaneously. Problems that severely hinder users during interaction should be distinguished from problems that have a small effect on the interaction.
- Address the potential effect of the problem on task outcome: As the impact of usability problems may threaten the health of people the classification scheme should be customized to the health care setting and address the impact of the problems on clinical task outcome.

In this paper we introduce a scheme consisting of three independent views on usability problems using (a) the User Action Framework (UAF); a validated classification framework [12], (b) the validated severity rating based on Nielsen's classification [13] and (c) a classification of the impact of the usability problems on the task outcome. We applied this scheme to classify and prioritize usability problems identified in a usability evaluation of a working CPOE system [10,15].

The paper is organized as follows. Section 2 provides an overview of the methodology that we followed to develop the augmented classification scheme. Section 3 presents the results of applying this scheme to a CPOE system. Section 4 discusses the findings and the strong and weak points of the study. Section 5 concludes the paper.

2. Methods

2.1. Augmented classification scheme

Fig. 1 shows the proposed augmented scheme for classifying and prioritizing usability problems concerning interactive health care information systems.

2.1.1. The UAF classification

In the field of usability engineering a standardized classification scheme offering various dimensions to classify usability problems is considered essential for accurate, complete and consistent problem reporting and for identifying the underlying cause of the problem. One classification scheme which is developed, validated and applied in usability research is the UAF classification scheme [12]. The UAF was built by adapting and extending Norman's theory of action model [16], a model that highlights issues about the way people interact with machines in terms of cognitive and physical user actions. The UAF classification provides insight into the users' planning and (physically) performing of activities and

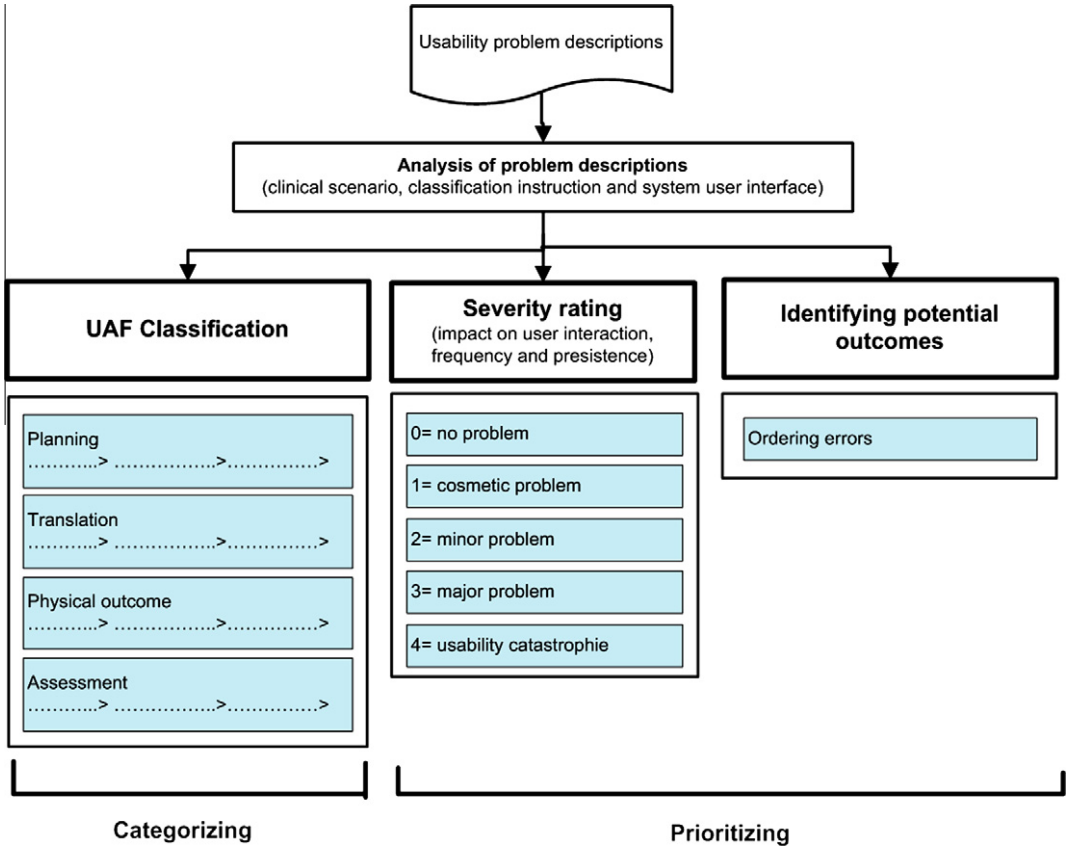

Fig. 1. The augmented scheme for classifying and prioritizing usability problems.

assessing the results of these activities throughout each cycle of interaction with a computer system. The purpose of the interaction cycle is to model the flow of user interaction in any interactive system. The UAF interaction cycle thus contains four phases; planning (high level), planning (translation), physical actions and assessment. These phases form the four elements of the first level of the classification. Each of these phases can be expressed in more detail using the standard terminology provided by UAF for another three levels. The ultimate usability problem typification can then be regarded as the fifth level.

Planning is the part of the interaction cycle that contains all cognitive actions by users to determine *what* to do. Supporting users in planning involves helping them understand the system model and helping them keep track of where they are within a task. High-level planning focuses on the system model and metaphors and the user's knowledge of the system states and modalities. Planning includes user work goal decomposition across a hierarchy of plan entities: the user establishes a goal, decomposes the goal into tasks to be performed on computer and establishes an intention of what to do to accomplish the task.

Translation is the part of the interaction cycle that contains all cognitive actions by users to determine *how* to carry out the intentions that arise during planning. In the translation phase, a user specifies the action sequence and determines which physical actions have to be executed in order to accomplish an intention, translating intentions into plans for physical actions. Cognitive affordances (e.g. visual clues) support the users' ability to plan physical actions. Therefore, usability issues concerning translation include those that pertain to the system's cognitive affordance presentation, and their content or meaning.

The physical action part of the User Action Framework is about executing the actions by manipulating user interface objects, and includes issues of interaction complexity and styles, manual dexterity, and layout. In graphical user interfaces, physical action mainly involves clicking, dragging, and selecting.

The assessment part includes a user perceiving, interpreting and evaluating the resulting system state. It concerns issues about the existence, presentation, contents and understandability of feedback and how it supports the user's ability to assess whether the outcome of physical action was desirable or effective [12,17].

The UAF provides a quasi-hierarchical tree of usability concepts and issues organized around the users' cognitive and physical actions, structured into four levels of abstraction, making it possible to classify and report usability problems from general towards more specific perspectives [12]. To each phase of the UAF interaction cycle one or more usability issues are associated with mutually exclusive standardized usability attributes or sub-categories below each issue. A usability problem can be classified using two or more levels of the UAF hierarchy. The UAF is viewed as a standard way to normalize usability problem descriptions [18] and is known to provide a highly reliable means for a detailed classification of usability problems.

As explained earlier, usability problem descriptions may be inconsistent between different evaluators, as well as ambiguous and incomplete, even when evaluators used the same usability evaluation method and even if they use a standard way of reporting. To communicate the usability problems unequivocally to system designers, it is important that evaluators match their problem descriptions on basis of similarity and/or distinctiveness. UAF classification can support the matching of usability problem descriptions; e.g. determining whether the different evaluators' problem descriptions are about the same underlying design flaw or not. The basic idea is that by using the UAF usability problems placed in different nodes of the UAF are by definition dissimilar.

UAF classification therefore supports distinction of problem types and agreement between evaluators concerning the nature of usability problems detected.

For example, a certain problem (presented by Fig. 3) may have been described by one evaluator in terms of user's actions as: 'users do not use the buttons ("m2" or "kg") on the main screen, provided for automatic calculation of the medication dosage'. Another usability evaluator may have described this problem as: 'users do not notice the buttons provided for calculating the medication dosage ("m2" or "kg")'. From the second evaluator's description it is clear that users did not notice the buttons due to poor noticeability. The first evaluator's description, however, can either refer to users not noticing the buttons due to their poor noticeability or to users not understanding the button labels as they do not guide the user in understanding the functionality they provide.

By applying the UAF, when both evaluators meant to explain that users did not notice the buttons due to their poor visibility, each of the problem descriptions would lead to an identical classification. When the first evaluator yet meant to explain that users did not use the buttons because of wrong text labeling, this problem description reflects a different type of usability problem and would result in a different classification of the problem, with different redesign recommendations for each of these problems respectively.

To elucidate how UAF classification supports matching of problem descriptions, we will illustrate the possible UAF classification paths as examples. First, presented by Fig. 2, the UAF classification path is described in which both evaluators would end up with a similar classification of the problem.

With UAF, usability evaluators first investigate in which phase of the interaction the problem occurred. The usability problem occurs after the user planned to calculate the cytostatics dosage that has to be entered in the dosage entry field. Therefore, the evaluators conclude that the problem does not occur in the planning phase but in the translation phase of the interaction. To translate his intention of calculating the dosage into a physical action, the user needs physical cues. In this example the users do not notice the existing buttons for this purpose (and in fact calculated the dosage on paper). Next, the evaluators try to map the problem to the different subcategories of the translation phase, first for level 2, then for level 3 and 4. The UAF framework provides the standardized (sub)categories that the problem could belong to. For instance, one of the subcategories of translation phase at the second level is "existence" (Fig. 2). The evaluators check whether the buttons needed to perform the physical actions exist on the screen. Since the buttons exist on the screen, the problem does not belong to this subcategory. Evaluators continue comparing the problem with the other attributes at this level. The problem is also not concerning the "meaning" of the buttons because as soon as a user would see the buttons he might understand the meaning of the labels. The problem is that the buttons are presented in such a way that they may be overlooked by a user. After finding the relevant category at level 2 (presentation) the evaluators try to map the problem to the subcategories of this category. This process continues until the problem cannot be mapped any further. The analysis of the problem described shows that it is difficult to observe the buttons, which is a "perceptual issue". And finally the reason that the user may overlook the buttons (underlying cause of the problem) is that the user may not notice them because of their small size, color close to the background, and the relatively 'large' distance from the dosage entry field. Note that dependent on the usability problem two or more levels are needed to describe the problem.

Second, if both evaluators would refer to two different usability problems, the classification path of the usability problem descriptions would then diverge starting from level 2 of UAF. Now if users do not perceive buttons they will not use it but if the buttons would be made more visible, users may use them. But if users do perceive these buttons but still do not understand their meaning, we deal with another issue, namely how to redesign these buttons and their labeling or even find another graphical way to ensure affordance directness.

Two usability evaluators with more than 5 years of usability expertise classified all usability problem descriptions from two usability evaluation studies of a CPOE system by traversing the UAF decision structure and selecting the most appropriate classification category and sub-category at each level of the hierarchy. One of these evaluators had not been involved in the CPOE usability evaluation studies and mainly received the problem descriptions produced. Each usability evaluator analyzed the problem description using the classification instruction [19] and the clinical scenario used for conducting the usability evaluations. Usability problems were then coded via the UAF hierarchy to the most detailed level. The resulting sets of usability problem classifications were reviewed in a meeting revealing disagreements which were

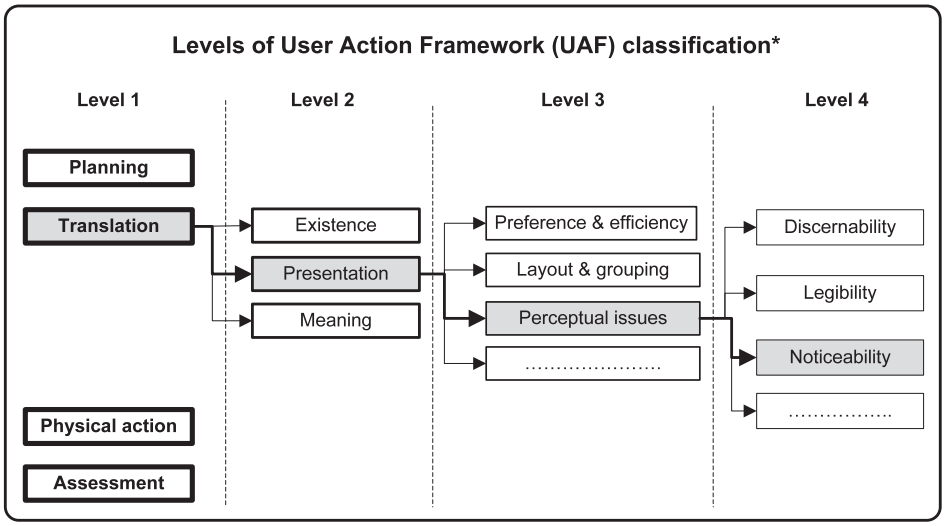

Fig. 2. The path for classifying the usability problem presented by Fig. 3 (the UAF (sub)categories belonging to this path are shown in gray with bold arrows). *Only categories belonging to the "translation" phase of the UAF are presented.

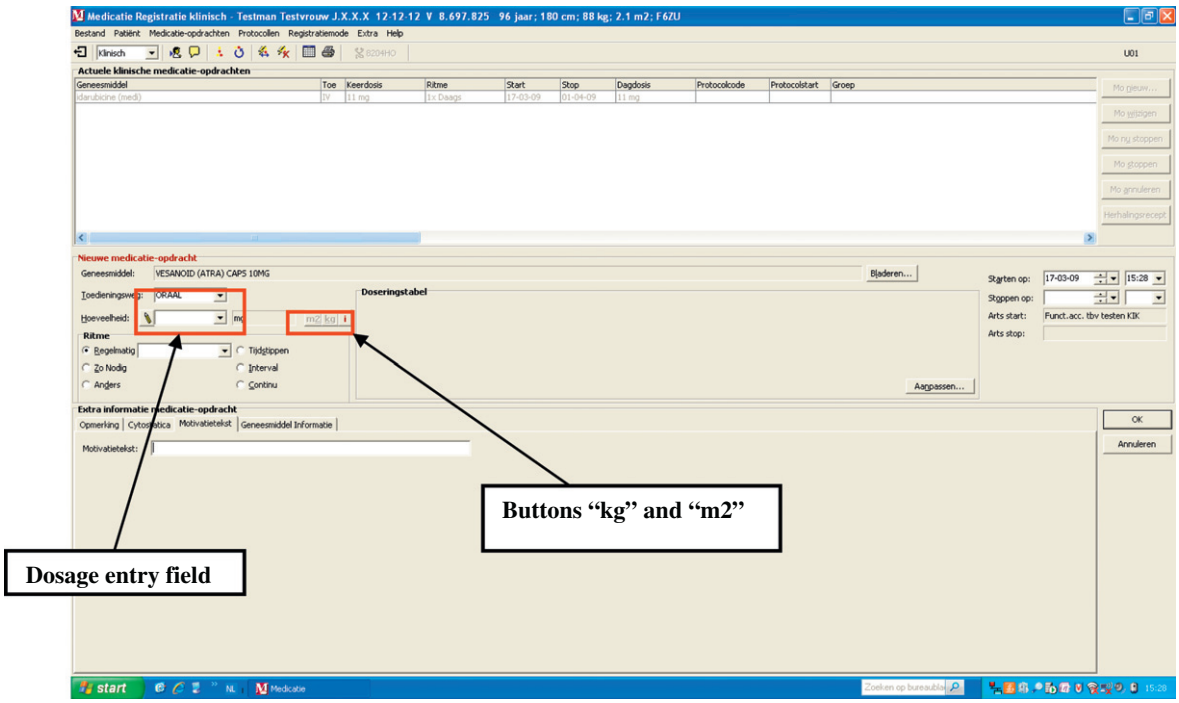

Fig. 3. Screenshot of a CPOE showing hardly noticeable buttons for the calculation of the medication dosage.

discussed. Any remaining disagreements were resolved through discussion with a third evaluator.

2.1.2. Severity rating of the usability problems

The severity of usability problems was rated based on Nielsen's validated classification [13] (Fig. 1). For determining the severity rating, the frequency with which a problem (might) occur(red), the (potential) impact of the problem on the users and the (potential) persistence of the problem are taken into account [20]. In one case study, Nielsen showed that the probability of getting a severity rating within ±0.5 rating unit from the true severity of a problem on a 5-point rating scale was only 55% with a single usability evaluator, but 95% for the mean of ratings of four independent evaluators [21]. He recommended collecting ratings from at least three evaluators. In this study the problem descriptions coming from two earlier studies [10,15] were used. For severity rating the availability of the problem descriptions alone is not always enough: the evaluators also need to have access to the system or to recordings of the think aloud user test sessions. Since only one of the authors of the current paper was involved in the previous CPOE usability evaluations, the severity of problems was determined by consensus of three usability evaluators. The evaluator involved in the previous studies could show the other evaluators relevant parts of the system or of the user tests' recordings when needed. In this way possible bias was eliminated. Severity classification of usability problems was done after UAF classification.

2.1.3. Determining the potential effect of the problems on task outcome

The two usability evaluators reviewed all usability problem descriptions and the corresponding system states at each moment in the user system interaction to determine what the potential effect of a usability problem could be on the ordering outcome. The potential effect of usability problems on ordering outcome could be classified (for this computer application) as wrong medication name, dosage, frequency, duration, and route of administration. The classification can be extended to be suitable for information systems in other domains with other types of effect.

3. Results

In total 57 usability problems identified in the CPOE usability evaluation studies were classified by the evaluators. The UAF classification of the usability problems together with each problem's severity and the potential effect of each problem on the ordering outcome were determined (Tables 1–4). Two (4%) of these problems needed classification up to the second, 16 (28%) up to the third and 39 (68%) up to the fourth level of the UAF hierarchy. At the first level of the UAF, 55 out of 57 usability problems were classified identically by the two evaluators.

One of the two incorrectly classified problem descriptions concerned the fact that users were looking for a way to undo a previous action in the CPOE system. Since this usability problem occurred when a user wanted to translate her intention of undoing a step into a physical action (and an undo button or an active link was missing), this problem should be classified under translation (as lack of existence of a way). The evaluator who had not been involved in the usability studies interpreted the problem description as if, the CPOE system did not provide system feedback to the user after its action and classified this problem under the assessment phase (as (lack of) "existence" of information).

The other problem description differently classified by the two evaluators concerned the situation where a user could not infer the functions of fields from their labels for altering the dosage in the drop down-menu of an alert screen. This problem also occurred when the user searched for a way to change the dosage without being alerted by the system. It should therefore be classified under translation (naming of labels > cognitive directness). Based on the problem description, the evaluator not involved in the usability evaluations interpreted this problem description wrongly as if users did not understand the system feedback provided by the alert and classified the problem as a faulty presentation of information related to the assessment phase of the UAF. In addition to the

Table 1
Example of a usability problem classification in the UAF[a] planning phase with its severity and potential effect on task outcome.

UAF categories				Example	No.[c]	Severity	Potential effects on outcome[b] (no. problems)
Level 1	Level 2	Level 3	Level 4				
Planning	User's ability to keep track of how much is done			*Users could not infer from the system state how much of the ordering task was achieved*	1	*2*	
	Users knowledge of system state, modalities			*User did not know that the system does not allow the opening of "Patient" menu during ordering and tried to open it*	1	*1*	
	Goal decomposition	Users ability to determine what to do next		*Users could not immediately infer from the screen 'Order set' that they had to enter a start date for a medication in order to proceed with the ordering process*	1	*1*	
		Supporting human memory limitations		*Users forgot to enter start and stop times. The layout of the start and stop time entry field did not attract the attention of the users*	1	*3*	***Wrong medication duration (1)***
	Users model of the system	User ability to determine what to do first		*From the layout of the main screen, users cannot instantaneously find out how to start the order process*	2	*1,1*	

[a] User Action Framework.
[b] Potential effect(s) of given example on outcome(s) are shown in italic bold.
[c] The number of problems with the same classification path. The corresponding severity ratings of problems classified under the same path are given in the next column.

two problems classified differently at the first level, there were three problems that were classified differently at a lower level (two at level 3 and one at level 4). These disagreements were mostly caused by a different interpretation of the problem descriptions by the evaluator not having participated in the usability evaluations. There was no disagreement at these levels when the evaluators looked at the system interface and reviewed the user test sessions with the system for clarification.

There was 96% agreement between evaluators (kappa = 0.94) for the classification of the problems at the first level of UAF hierarchy. The 57 problems were classified into 29 different UAF classes, also called paths. Different problem descriptions can be classified identically, for example 11 different problem descriptions were classified with the path: Assessment > Feedback > Existence > Existence of a cognitive affordance. Different problem descriptions belonging to one UAF classification leaf can have different severities as is apparent from Tables 1–4.

In total five problems were classified differently by the two evaluators. The accuracy of the classification therefore was 91%. Since the confusion matrix was too sparsely filled, it was not possible to determine the corresponding kappa value for the total classification (taking all levels into account). In Tables 1–4 examples of problem classifications are provided per phase of the UAF framework.

3.1. Planning

Of the total number of identified usability problems six (10%) problems were found relating to this phase of user–system interaction. Classification of the usability problems showed that usability problems in this phase of interaction were caused by the user's difficulties to understand the overall CPOE system model, system state or modalities and the users' inability to keep track of the ordering steps that were completed and to determine the next goal to be accomplished. Several users had for example difficulties in remembering which steps they had completed, and what next steps they had to carry out to enter a medication prescription.

The severity of problems concerning the planning phase was low except for one problem that could potentially result in wrong medication durations. Table 1 shows that problems with a similar severity and description can have different underlying causes when they are categorized with the UAF. The augmented scheme is able to differentiate between problems having identical severity and similar problem descriptions. For example one of the problems was that it took a while before the users entered a start date for a medication, and another problem was that it took a while before the users started the ordering process. These two problems look similar based on their description and severity, but when they were classified using the UAF framework the underlying cause of these problems appeared to be different. In the first problem when the users decided to order medication using order sets, they could not decompose their goal into the tasks required to proceed with ordering. In the other one the users already determined the task (start ordering process) but as the model of the system differed from what they expected they could not decide which action to take to accomplish the task.

3.2. Translation

Sixteen (28%) of the usability problems concerned this phase of the interaction (Table 2). The UAF classification revealed that usability problems in this phase were mainly caused by the lack of certain functionality so that certain tasks cannot be performed or, if the functionality is available, the lack of cognitive affordances showing how to successfully perform a task. Also poor presentation and design, poor content and meaning of available cognitive affordances may cause problems. Mismatches between the users' terminology and the terminology, abbreviations and labeling of buttons used in the CPOE system also were a source of usability problems. Seventy-five percent of the problems concerning this phase of interaction had a minor severity (severity 2). Although their severity was similar, their underlying cause was different and these problems were categorized under different subcategories of the translation phase of the UAF. Classification of the problems with the augmented scheme showed that some of the

Table 2
Example of a usability problem classification in the UAF[a] translation phase with its severity and potential effect on task outcome.

UAF categories				Example of classified problems	No.[c]	Severity	Potential effects on outcome[b] (no. problems)
Level 1	Level 2	Level 3	Level 4				
Translation	Content and meaning	Clarity, precision and predictability	Completeness and sufficiency of meaning	*Users did not understand the meaning of some abbreviations used in the medication list*	2	*2,2*	***Wrong medication selection (2)***
		Naming of labels	Cognitive directness	*Users could not infer the functions of two buttons "change" and "record" based on their labels*	2	*2,4*	***Wrong medication dosage (1)***
	Existence	Existence of a way		*Users could not undo an action in the system and change a previously made selection in the ordering process*	1	*2*	
		Cognitive affordances		*Users expressed that they needed help information to know which consolidation (1, 2 and 3) is suitable for the patient*	3	*2,2,2*	
	Presentation	Preferences and efficiency issues		*Users preferred to select from an alphabetically organized list of order sets rather than checking all items in a non-organized list*	2	*2,3*	
		Perceptual issues	Noticeability	*Users did not notice the buttons provided for calculating medication dosages on the main screen*	4	2,2,3,3	***Wrong medication dosage (2)***
			Discernability	*Users cannot comprehend the function of the button "new order" from its shape*	1	*2*	
	Task structure and interaction control	Consistency and compliance of task structure		*The labeling of menus and submenus provided for initiation of an order is not consistent with the labeling of buttons provided for the same purpose*	1	*2*	

[a] User Action Framework.
[b] Potential effect(s) of given example on outcome(s) are shown in italic bold.
[c] The number of problems with the same classification path. The corresponding severity ratings of problems classified under the same path are given in the next column.

problems would get a low priority if they were classified based on their severity only while these problems will be given a high priority if also their impact on the task outcome is considered for prioritization. For example, of the five problems leading to a medication error, the severity of three problems was rated as minor.

3.3. Physical actions

Nine (16%) of the usability problems were encountered in this phase of the interaction (Table 3). The UAF classification showed that these usability problems were caused by faulty presentation and lay out of the physical objects to be manipulated on the screen, lack of user control over screen objects as they were being manipulated and failure of the system to meet specific preferences of users for doing physical actions.

Based on severity 78% of the problems in this category were similar (rated as severity 2) while the causes of these problems were different as was apparent from the assignment to different subcategories of the UAF. One of the problems, which was rated as severity 2 based on its frequency, persistence and impact on the user, could lead to a wrong medication selection and, thus, should get a high priority for fixing.

3.4. Assessment

In total 26 (47%) of the 57 identified usability problems were classified in the assessment phase (Table 4). These problems concerned the existence, presentation, content and meaning of system feedback about the course of the user-interaction and the display of information resulting from users' actions. Around half (12 out of 26) of the problems concerning the assessment phase were severe problems (severity 3 and 4) and they could influence the task outcome. Not all the problems influencing the outcome were highly severe problems since three of the problems potentially resulting in ordering errors were assigned severity 2. The UAF classification showed that 19 (73%) of the problems concerning the assessment phase of interaction were caused by a lack or suboptimal design of system feedback or by unclear feedback contents. The remaining seven (27%) problems in this phase were caused by the absence, unclear contents and meaning, and poor presentation of information displayed after the users' actions.

4. Discussion

This study introduces an augmented classification scheme for classifying and prioritizing usability problems. The scheme was applied to the problem descriptions of the CPOE usability evaluation studies. Examination of problems at the various levels of the UAF enabled us to identify meaningful problem clusters containing problems that shared common characteristics and hence revealed the nature of the usability problems at both more global and more specific levels. We could show that the use of only problem descriptions and severity ratings, as is usual in usability practice, does not suffice to distinguish different usability problem types. Classification of usability problem descriptions by use of UAF was helpful in determining whether these descriptions were about the same underlying usability problem or not, leading to similar or different redesign recommendations respectively.

Furthermore, the scheme in the hand of both evaluators led to the same classification in more than 90% of the 57 problems descriptions. Although the two usability experts showed a high inter-rater agreement on the classification of problems based on their problem descriptions, some descriptions appeared vague in describing the essence of the underlying usability problem and consequently led to a different UAF classification. One of the evaluators had not participated in the CPOE usability evaluation studies. This evaluator apparently needed detailed information to interpret certain problem descriptions correctly. This indicates that when further analysis and discussions as input to a system's redesign are performed after a delay in time, problem descriptions may be interpreted differently by various people involved in the redesign process. When usability evaluators would have used the UAF

Table 3
Example of a usability problem classification in the UAF[a] physical action phase with its severity and potential effect on task outcome.

UAF categories				Example of classified problems	No.[c]	Severity	Potential effects on outcome[b] (no. problems)
Level 1	Level 2	Level 3	Level 4				
Physical action	Perceiving physical objects	Perceiving objects as they are being manipulated		*Different behavior of the button "select" in the patient selection window than that expected by users*	1	*2*	
			Discernability	*Users cannot perceive in which format they should enter patient's date of birth*	1	*2*	
	Manipulating objects	Physical control		*User could not open tab motivation text to enter the motivation for an order*	2	*2,3*	
		Physical layout		*User needed to use scroll down bar to see the whole list of order sets*	1	*2*	
			Proximity issues	*User by accident clicked on a wrong option in the medication order set list*	2	2,2	***Wrong medication selection (1)***
			Preferences and efficiency	*User cannot review the system's calendar year by year to set date of birth*	1	*2*	
		Preferences and efficiency		*Users could not directly enter the number of days for medication duration. Instead, The CPOE system requires the entering of a start and stop date*	1	*4*	***Wrong medication duration (1)***

[a] User Action Framework.
[b] Potential effect(s) of given example on outcome(s) are shown in italic bold.
[c] The number of problems with the same classification path. The corresponding severity ratings of problems classified under the same path are given in the next column.

framework early in the analysis process of usability problems this would have supported accurateness, completeness and consistency in usability problem reporting.

With the addition of two dimensions for ratings of severity and of the potential effect of revealed usability problems on task outcomes, the augmented UAF will add to that by supporting a proper prioritization of usability problems in redesign efforts. As such, the scheme advanced our understanding of the usability problem set of the CPOE usability studies.

Many studies have evaluated the usability of a variety of interactive health care applications and reported on the usability problems and their impact on the user interaction. Apart from the application of usability heuristic classifications in some studies, none of these studies used a systematic framework for guiding and structuring the evaluation and reporting of these usability flaws.

The augmented scheme provides better classifications than the current usability problem classification strategies such as heuristic [22] and bottom up classification [23,24] in the sense that classifications of usability problems by this new scheme based on the UAF are distinguishable, mutually exclusive, complete and specific. Heuristic taxonomies lack these characteristics leading to usability problems of a very different nature being classified to the same heuristic, a single problem classified to more than one heuristic or to one heuristic not adequately capturing its essence and to impossibilities to classify some problems at all. This leaves the designer of the system with ambiguity about the precise nature of the usability problem. Moreover, the utility of a classification that is essentially a by-product of an evaluation technique such as heuristic analysis can be limited because classification is not the intended purpose of the technique [14].

In bottom up classifications of usability problems, the frameworks of reference evolve with the analyses and as such not only become highly dependent upon the expertise of the evaluators, but more importantly lack a common foundation for future usage: the consistent, complete and accurate reporting of usability problems and understanding of their underlying causes. The need for standardized classification and reporting of usability issues is clear when the aim is to share the usability experiences by the development of a knowledge base concerning usability problem characteristics and their impact on user–system interactions.

Severity classifications of identified usability problems, providing insight in their frequency, persistence and impact on users, are commonly used in prioritizing system redesign efforts. Severity classifications applied to usability problem datasets represent somewhat isolated characteristics of these sets. Such classifications only address the criticality of the problem without providing designers of the systems with more information about what the problem is, how it occurred, and what could be its effect on task outcome. Furthermore, these classifications do not indicate the potential impact of usability problems on the final outcomes of the user–system interaction.

Our scheme reveals the core problems and hence allows prioritization of the order in which to address problems in a system redesign. Prioritization based on frequency, persistence and effect on users only does not reflect their potential effect on patient safety. The problems with a low severity could be given higher priorities if their potential effect on the task outcome is also taken into account. Therefore, for prioritizing redesign efforts severity ratings should preferably be accompanied by assessments of the possible influence of identified usability problems on final task results.

Going deeper into the hierarchy of the UAF, the number of the problems that were classified under different categories increased. This highlights the fact that problems seeming similar at the surface (the first levels of the UAF) can have different underlying causes [18]. A one level classification such as severity rating or bottom up approaches treats all problems that are clustered in the same category similarly without paying attention to the potential differences in their cause. In the two classification examples we provided, we addressed the issue of vague problem descriptions potentially leading to a different classification concerning the essence of the problem. When the low visibility ('noticeability") of buttons on the screen would be the root cause for users not using these

Table 4
Example of a usability problem classification in the UAF[a] assessment phase with its severity and potential effect on task outcome.

UAF categories				Example of classified problems	No.[c]	Severity	Potential effects on outcome[b] (no. problems)
Level 1	Level 2	Level 3	Level 4				
Assessment	Feedback	Content and meaning	Completeness and sufficiency of meaning	*Users could not understand the function of the column "dosage percentage" in the medications table*	2	2,2	
			Cognitive directness	*Users did not understand the recommendation provided by the alert "medication dosage-unit control" correctly*	4	1,2,2,4	***Wrong medication dosage (1)***
			Error avoidance	*User can enter frequency of medication in letters, but system alerts "no dosage [not frequency] is entered". User should only enter numbers*	1	2	
		Existence	Existence of a cognitive affordance	*Users do not receive a feedback or warning when they forget to enter a stop date for a medication and can proceed to the next step*	11	2,2,2,2,2,3,3,3,3,3,4	Wrong medication selection (1), ***Wrong medication duration*** (4), wrong medication dosage (3), wrong medication frequency (1)
		Presentation	Perceptual issues > timing	*Alert screen "medication dosage-unit control" shows up too late in the ordering process*	1	3	***Wrong medication dosage (1)***
	Information display	Existence	Human memory aids	*Users are not provided with the unit of the calculated dosage in "Dosage calculation" windows*	2	3,3	***Wrong medication dosage (1)***, Wrong medication selection (1)
		Content and meaning	Error avoidance	*User was confused when the system retrieved an alternative name of the medication than the one typed by the user*	1	2	
			Layout and grouping	*Independent of the administration time entered by users a different administration time is shown in a different section of main screen*	2	2,4	***Wrong medication duration (1)***
		Presentation	Perceptual issues > noticeability	*User did not notice that the administration time was already recorded and re-entered it*	2	2,3	Wrong medication duration (1)

[a] User Action Framework.
[b] Potential effect(s) of given example on outcome(s) are shown in italic bold.
[c] The number of problems with the same classification path. The corresponding severity ratings of problems classified under the same path are given in the next column.

buttons' functionalities to calculate medication dosages, redesign efforts would concentrate on improving the visibility of these buttons. When the meaning of the button would however not be clear to users, redesign efforts would focus on finding another graphical way to ensure that system users can infer the affordance provided (automatic calculation of dosage based on a patient's weight and body volume). To tackle the usability problems fundamentally a redesign effort should thus target the underlying cause of each problem. This indicates that one-level classifications such as heuristic analysis might fail to meet this requirement. In order to provide more details concerning a problem, one level classifications should be expanded horizontally. This cannot be done for classifications such as severity rating and classifications based on heuristics because the problem classes are pre-defined.

The proposed augmented classification scheme provides the rationale behind design recommendations by giving information about how a problem endangers the interaction with the system in the planning, translation, physical or assessment phase. Moreover, system redesigners are supported when prioritizing their efforts by focusing on the problems that critically affect the user-interaction and the task outcomes. Furthermore, our scheme could be of value for studies comparing the strengths and weaknesses of usability evaluation methods for user interface evaluation. Common practice is to compare the proportion of minor and major problems found by these methods. While such a criterion may be useful in examining which method has a higher detection rate of major problems, it does not provide insight in the capability of each method to detect specific classes of usability problems (e.g. problems concerning a specific phase of interaction).

This study has some limitations. In our study the two usability experts were comfortable with classifying problems according to the UAF and showed a low number of disagreements among their ratings. These disagreements seemed to result from vague problem descriptions and not from differences in expertise level. The small number of the expert usability evaluators in this study may likewise explain the high degree of agreement. The UAF and Nielsen's classification have both been validated. Andre et al. [12] have shown that the reliability of the UAF for categorization of problems is higher than the heuristic classification and the Usability Problem Taxonomy [14], a multi level usability problem classification. These results support the notion that the UAF provides a reliable classification system that is helpful in developing a common understanding of the different usability problem attributes. Nevertheless, though UAF yields better results in terms of consistency and accuracy of usability problem reporting, applying the scheme is an elaborate and time consuming task. Future studies could address these trade-offs in order to establish its practical value in usability evaluation studies of health information systems. Another limitation of the UAF classification is that it in essence does not contain information on the context of use of interface features considered to be or lead to usability problems. Some system functionalities may only be used in a certain context by a specific group of system users. The analysis of these contexts-of-system use is likewise essential in guiding system redesign efforts. High quality

problem descriptions would contain information on the contexts-of-use but this information is not retained after standard UAF classification, even at its lowest level. We added the two dimensions, severity rating and assessment of potential impact on task outcomes, to strengthen the original UAF classification scheme. To some extent, these two dimensions provide indications of the context of use of certain system functionalities for which usability problems are revealed, namely in terms of frequency with which a problem occurred, the impact of the problem on the users, the persistence of the problem during system interaction and its impact on user task performance. A next step would be to consider the addition of other dimensions to the augmented UAF to describe contexts-of-use like what type of end-users encountered the problem and under which specific conditions of system use. We furthermore acknowledge the fact that the results of this study are based on one usability problem set concerning a single CPOE system. Moreover, the two evaluators were familiar with the CPOE system, one of them conducted the usability tests with end users and thus was aware of the context in which usability problems occurred. Evaluators not familiar with the type of system under study, not having participated in the usability end user tests may experience more difficulties in the UAF classification of usability problems. But if evaluators would use the UAF classification directly after the system usability evaluation they would have less difficulty in analyzing and classifying the revealed problems. Furthermore, the usability problem dataset used for examining the value of the augmented classification scheme was based on CPOE usability studies at one academic hospital. Despite these limitations, we believe that this study highlights an important issue and serves as a model for other researchers seeking to enhance their insights into the impact of information system designs on usability on both more global and specific levels.

In a general sense, usage of this augmented scheme will minimize subjective analysis and inconsistent classification of usability evaluation results. Widespread application could help to obtain a complete and consistent classification and description of usability problems based on their underlying causes and as a result produce problem reports of higher quality that better support redesign of the evaluated systems. Adoption of the scheme in practice would aid in more easily revealing trends and patterns across problem sets of usability evaluation case studies. This would assist in building comprehensive usability knowledge bases and advancing human computer interaction science in the health care domain. The usefulness and value of the approach exemplified in this study to classify and prioritize usability problems of a CPOE medication system should be further investigated for usability evaluations of other health care applications. Future studies could examine how the consideration of the potential task outcome of usability problems for severity rating enhances the prioritization of usability problems to be tackled. Finally, research could address the issue of which dimensions to add to the framework for the description of contexts-of-use in relation to specific user problems to even better guide redesign efforts.

5. Conclusions

The augmented classification scheme used in this study enables usability evaluators to analyze and classify usability problem sets in relation to the phase of the user–system interaction and addresses system users' cognitive, physical and assessment efforts triggered by each of these problems. This classification differentiates problems that on the basis of their description and severity seem similar but affect the user–system interaction and the task results differently. Evaluators using the scheme independently arrived at the same problem classification in more than 90% of the problems.

Competing interests

None.

References

[1] Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med 2003;163(12):1409–16.

[2] Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res 2010;10:231.

[3] Ash JS, Sittig DF, Dykstra R, Campbell E, Guappone K. The unintended consequences of computerized provider order entry: findings from a mixed methods exploration. Int J Med Inform 2009;78(Suppl. 1):S69–76.

[4] Campbell EM, Guappone KP, Sittig DF, Dykstra RH, Ash JS. Computerized provider order entry adoption: implications for clinical workflow. J Gen Intern Med 2009;24(1):21–6.

[5] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104–12.

[6] Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293(10):1197–203.

[7] Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005;74(7–8):519–26.

[8] Khajouei R, Jaspers MW. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: a systematic review. Methods Inf Med 2010;49(1):3–19.

[9] Guappone KP, Ash JS, Sittig DF. Field evaluation of commercial Computerized Provider Order Entry systems in community hospitals. AMIA Annu Symp Proc 2008:263–7.

[10] Khajouei R, de JD, Jaspers MW. Usability evaluation of a computerized physician order entry for medication ordering. Stud Health Technol Inform 2009;150:532–6.

[11] Beuscart-Zephir MC, Pelayo S, Bernonville S. Example of a Human Factors Engineering approach to a medication administration work system: potential impact on patient safety. Int J Med Inform 2010;79(4):e43–57.

[12] Andre TS, Hartson HR, Belz SM, McCreary FA. The user action framework: a reliable foundation for usability engineering support tools. Int J Human–Comput Stud 2001;54:107–36.

[13] Nielsen J. Usability engineering. Boston: Academic Press; 1993.

[14] Keenan SL, Hartson HR, Kafura DG, Schulman RS. The usability problem taxonomy: a framework for classification and analysis. Emp Softw Eng 1999;4:71–104.

[15] Khajouei R, Peek N, Wierenga PC, Kersten MJ, Jaspers MW. Effect of predefined order sets and usability problems on efficiency of computerized medication ordering. Int J Med Inform 2010;79(10):690–8.

[16] Norman DA. Cognitive engineering. In: Norman DA, Draper SW, editors. User centered system design: new perspectives on human–computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986. p. 31–61.

[17] VirginiaTech. The user action framework (UAF) and its tools; 2003. <www.cs.vt.edu/research>, <http://research.cs.vt.edu/usability/projects/uaf%20and%20tools/index.htm> [cited 10 A.D. December 18].

[18] Hartson HR, Ander TS, Wilinges CR. Criteria for evaluating usability evaluation methods. Int J Human–Comput Interact 2003;15(1):145–81.

[19] Andre TS. Determining the effectiveness of the usability problem inspector: a theory-based model and tool for finding usability problems. PhD dissertation, Faculty of the Virginia Polytechnic Institute and State University; 2000.

[20] Nielsen J. Severity ratings for usability problems; 2010. <http://www.useit.com>, <http://www.useit.com/papers/heuristic/severityrating.html> [cited 10 A.D. December 18].

[21] Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL, editors. Usability inspection methods. New York: John Wiley and Sons, Inc.; 1994. p. 25–62.

[22] Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1–2):23–30.

[23] Horsky J, Kuperman GJ, Patel VL. Comprehensive analysis of a medication dosing error related to CPOE. J Am Med Inform Assoc 2005;12(4):377–82.

[24] Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004;37(1):56–76.