Context matters: Design guidance for maximizing success of storytelling sessions during a user-centered design task

Kim Gausepohl, Woodrow Winchester III, James D. Arthur, Tonya Smith-Jackson and Brian Kleiner
Virginia Tech

Empirical comparisons of elicitation methods provide guidance to engineers for the selection of the best method for a given design goal. This work explores how alterations to the context (i.e, setting, prompts) of storytelling sessions impacted the types of user needs collected from 22 emergency room (ER) nurses. Six user need sets, comprised of 477 narratives detailing an experience relating to a healthcare quality aim, were quantitatively compared. The results of the study yielded a set of decision rules and decision trees to aid the selection of an appropriate storytelling protocol that facilitates formative usability design.

INTRODUCTION

The National Institute for Occupational Safety and Health's (NIOSH) "Prevention through Design" (PtD) initiative encourages the exploration of different methods to foster a collaborative dialogue between engineers and healthcare workers. Design standards (ANSI/AAMI, 2009; IEC, 2007) recommend the identification and consideration of usability needs early in the design process through a formative usability approach to design. Although engineers are encouraged to follow a user-centered design (UCD) process to identify user needs, NORA (2009) warns that engineers may "fail to get the full range of healthcare worker input on the usability of a device".

Previous work (Gausepohl, Winchester, Arthur, & Smith-Jackson, 2011) explored storytelling as an elicitation method for medical device usability requirements. The authors accomplished this through comparison of information elicited from registered nurses from interviews and storytelling sessions. Storytelling participants contributed significantly more context-of-use information and showed an emphasis on social context. These findings suggest that storytelling aids engineers' understanding of usability requirements.

This work aims to provide engineers with practical design guidance through exploration of the research question, "How can engineers alter the storytelling context (i.e., setting, prompts) to elicit the best compiled user need set for a formative usability approach?" As engineers are not typically assigned tasks such as, "Collect the best compiled list of efficiency needs", we limited our exploration to the specific goal of incorporating usability into design, which is one of several potential design goals within a UCD process (Table 1).

Table 1. The collection of functional, usability, and context-of-use needs supports several UCD design goals.

Goal	Task	Needs of interest
Identify a design problem	Explore the problem space	Functional Context-of-use
Understand an unfamiliar domain	Specify context-of-use	Context-of-use
Incorporate usability into the design process	Formative usability testing	Functional Usability
Validate design	Summative usability testing	Functional Usability

 DOI 10.1177/1071181312561160

The selection of formative usability testing as the task of interest narrowed the focus of comparisons to functional and usability needs elicited from each context combination. For example, the collection of functional needs (i.e., "what" the system needs to do) and usability needs (i.e., "how" the system needs to operate) supports the creation of benchmarks used during formative and summative usability testing. In contrast, problem space exploration requires the collection of functional and contextual information. Successful elicitation of functional needs supports the identification of what the system and/or person needs to accomplish through the design. The collection of contextual information further elucidates the differing users and environments for which the design must function. The collection of contextual information is also valued when the goal is to understand an unfamiliar domain through the specification of the context-of-use.

We collected ER nurses' stories under varying contexts and quantitatively explored the differences in user needs between six compiled sets. PtD contributions include a decision aid for the selection of an appropriate storytelling context that facilitates formative usability informed by the empirical findings of this work.

METHOD

Data collection and sample

Twenty-eight registered nurses were recruited from four emergency room (ER) departments within one corporate healthcare organization. Two hospital locations were self-described as "mid-size city hospitals" and two locations were considered "small-size rural hospitals". Institutional Review Board (IRB) approval was obtained at the researchers' university and at each hospital. All nurses provided written informed consent prior to the start of study procedures.

Participants were assigned to one of four storytelling contexts based on demographic information (e.g., gender, years-of-experience, workplace location). The four storytelling contexts were:

- IH Individual storytelling with habitual prompts
- IHH Individual storytelling with habitual and hypothetical prompts
- GH Group storytelling with habitual prompts
- GHH Group storytelling with habitual and hypothetical prompts

Demographics for the recruited sample were similar to those reported in a national sampling of the registered nurse population (HRSA, 2008). The majority of participants (n = 22) were female with nursing experience ranging from 3 months to 38 years (*M* = 14.7 years).

Six group storytelling sessions (n=12) were conducted. Three pairs of nurses contributed stories in response to the GH context (n = 6) and the remaining pairs shared work experiences within the GHH context (n = 6). Sixteen individual storytelling sessions were conducted with an equal number of nurses within both the IHH (n = 8) and IH (n = 8) contexts. Unexpected complexity in scheduling group sessions, such as coordination of nurses working multiple 12-hour shifts that varied weekly, contributed to the lower participation in group sessions.

Scripts for all storytelling contexts contained habitual prompts, which asked participants to share experiences relating to an Institute of Medicine (IOM) quality aim (e.g., safe, timely, equitable, patient-centered, effective, efficient care) (IOM, 2001). The use of IOM prompts aided the engineers' overall goal of problem space exploration while facilitating discussion about topics important to stakeholders. Scripts for the IHH and GHH sessions contained hypothetical questions, which asked the participant to retell how the recently related experience would have been in an ideal world (Table 2).

Table 2. Habitual and hypothetical prompts for safe care

IOM aim	Habitual prompt	Hypothetical prompt
Safe care	Please tell me a story about an experience that you had where patient safety was an issue.	How would you re-write this story? For example, walk me through how this experience would have been in an ideal world.
	Please tell me a story about an experience that you had where you felt your own safety was an issue.	How would you re-write this story? For example, walk me through how this experience would have been in an ideal world.

Each script contained a total of eight story prompts. Since safe care and timely care affect both nurse and patient satisfaction, additional prompts were created to elicit stories from both perspectives.

Data analysis

The study employed a mixed methods approach to data analysis. Narrative and thematic analysis methods were used to facilitate the identification of user needs within participant responses. User needs were defined as a "precursor to a requirement, which contains a description of general or specific characteristics of the design problem that must be accounted for in the development of the design solution". Compiled sets of user needs across participants were quantitatively compared.

Narrative analysis. Structural analysis was used to transform participant responses (i.e., stories) into concise narratives. Participant responses, which were typically free-form recollections of events and experiences, were structured in a consistent format following Labov & Waletzky's (1967) narrative components (e.g., abstract, orientation, complicating action, resolution, evaluation). Structural analysis facilitated data consistency. For example, participant responses that did not adhere to our operational definition of a narrative were identified during structural analysis and excluded from the data set. The use of narrative components provided consistency in format and structure across narratives, which facilitated later thematic analysis.

Three teams, each consisting of three undergraduate research assistants acting as two independent coders and one coding judge, performed structural analysis of nurses' stories. These teams were trained on narrative analysis methods without knowledge of the study's research question. Following Capra's (2006) procedure, the team's coding judge recorded all agreements and disagreements during weekly reconciliation meetings. Percent agreement was calculated for a random 10% sampling of the narratives as a measure of inter-coder reliability. Percent agreement before reconciliation ranged from 0.43 to 0.71; scores after reconciliation meetings ranged from 0.99 to 1.0.

Thematic analysis. The ER nurses' narratives, which were collected via storytelling sessions of differing contexts (i.e., IHH, IH, GHH, GH), were the primary data source for thematic analysis.Thematic analysis was used to identify and categorize user needs within narratives towards the creation of a master list of user needs and categorizations. One thematic analysis team, comprised of three graduate research assistants, used ATLAS.ti software to manage the coding, categorization, and reconciliation tasks. Similar to the narrative teams, the thematic team consisted of two coders and one judge who were not told the study's research question. After training on thematic analysis, user needs specification, and ATLAS.ti software functionality, coders specified and categorized user needs following a detailed protocol and ontology based on usability definitions in medical device standards (ANSI/AAMI, 2009; IEC, 2007) (Figure 1).

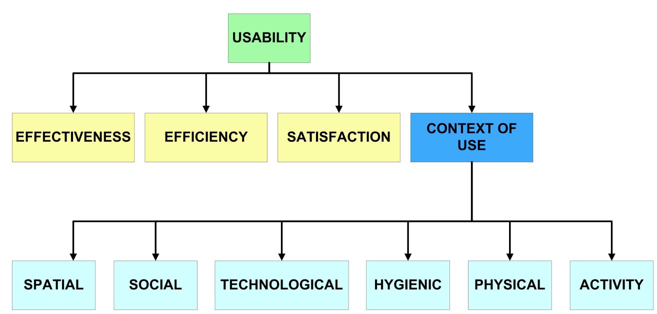

Figure 1. Usability ontology used during thematic analysis

ATLAS.ti tools facilitated calculation of the percent agreement for the entire thematic analysis data set. Prior to reconciliation, agreement averaged 0.55; agreement after reconciliation was 1.0.

Comparisons of compiled user needs sets. Triangulation of methods is frequently recommended in design literature as a means to leverage the benefits of multiple methods in an effort to elicit the most comprehensive set of user needs from stakeholders as possible (Garmer, Liljegren, Osvalder, & Dahlman, 2002; Martin, Murphy, Crowe, & Norris, 2006). To ensure that our recommendations consider opportunities for triangulation, we compared compiled user need sets comprised of two distinct method types (Table 3).

Table 3, Methods used within each complied set

Compiled User Needs Set	Methods (N = 2)	Sessions (min = 6, max = 12)	Participants (N = 12)
G	GHH GH	(n = 3) (n = 3) *N = 6*	(n = 6) (n = 6) *N = 12*
I	IHH IH	(n = 6) (n = 6) *N = 12*	(n = 6) (n = 6) *N = 12*
HH	GHH IHH	(n = 3) (n = 6) *N = 9*	(n = 6) (n = 6) *N = 12*
H	GH IH	(n = 3) (n = 6) *N = 9*	(n = 6) (n = 6) *N = 12*
Mixed 1	GHH IH	(n = 3) (n = 6) *N = 9*	(n = 6) (n = 6) *N = 12*
Mixed 2	GH IHH	(n = 3) (n = 6) *N = 9*	(n = 6) (n = 6) *N = 12*

Each set represents the compiled needs contributed by 12 participants (i.e, six nurses from each storytelling context). Since our original data collection resulted in unbalanced groups between individual (n = 6) and group (n = 8) storytelling sessions, two participants were randomly removed from both the GH and GHH groups to balance the number of participants.

To provide recommendations to engineers for how to achieve the "best" user need set for formative usability through alterations of the storytelling context, we explored the quality of the compiled set in terms of breadth and depth. Breadth, denoted as a proportion of the categories discussed to the total possible categories, represented eachthe set's topic coverage of categories within the user needs ontology. Depth, a measure of how much information was elicited, was represented by the quantity of user needs elicited within each category. Comparisons of depths were calculated as percent ratios; any ratio greater than 10% was considered to be a significant difference, as in previous work (Gausepohl et al., 2011) . Venn diagrams were used to visually discern differences in depth, as statistical methods were not appropriate.

RESULTS

A total of 477 narratives were elicited from the 24 ER nurses included in analyses. Of these narratives, 369 were habitual and 108 were hypothetical. Thematic analysis of the entire narrative set resulted in the identification of 126 functional and 289 usability needs. A total of 369 user needs were identified as 46 needs were classified within both functional and usability categories.

Breadth

No significant differences in breadth were found as all groups contributed at least one user need for each category, resulting in a 1.0 breadth score for all groups.

Depth

Depth of total needs within each compiled set ranged from 262 to 323 identified needs (Table 4). The range of usability needs within each set (Min = 206, Max = 257) exceeded the quantity of functional needs identified (Min = 93, Max = 103).

Table 4. Quantity of functional and usability needs per compiled set

Set	Total	Functional	Usability
G	323	108	253
I	284	94	227
H	262	93	206
HH	325	107	257
Mixed1	313	103	246
Mixed2	284	96	226

Sets with the highest number of total needs (i.e., G, HH, Mixed1) also contained the highest number of functional and usability needs. This was expected as the specification protocol used in this study required initial identification of functional needs followed by usability needs that described the functional need.

Depth of functional needs. The G, HH, and Mixed1 sets contained a significantly greater quantity of functional needs than some other sets (Table 5). The compiled G set contained significantly greater depth of functional needs than the H, I, and Mixed2 sets. The G set contained 16.1% more functional needs than the H set, 14.9% more than the I set and 12.5% more than the Mixed2 set.

Table 5. Percent differences in functional needs per compiled set

	G	I	H	HH	Mixed1	Mixed2
G*	X	***14.9****	***16.1****	1.0	4.9	***12.5****
I	-13.0	X	1.1	-12.2	-8.7	-2.1
H	-13.9	-1.1	X	-13.1	-9.7	-3.2
HH*	-1.0	***13.9****	**15.1***	X	3.9	**11.5***
Mixed1*	-4.5	9.6	**10.8***	-3.7	X	7.3
Mixed2	-11.1	2.1	3.2	-10.3	-6.8	X

***significant difference (i.e., >10%)**

The compiled HH set contained 15.1% more distinct functional needs than Set H, 13.9% more than Set I 12.5% more than Set Mixed2 set. The H set also contained significantly less functional needs than the Mixed1 set, which represented a 10.8% increase of the H set.

Depth of usability needs. Significant differences between sets for functional depth were also found for usability depth (Table 6). The I set and Mixed2 sets contained significantly greater usability needs than the H set, resulting in the identification of the H set as containing the smallest usability depth than all other sets.

Table 6. Percent differences in usability needs per compiled set

	G	I	H	HH	Mixed1	Mixed2
G*	X	***11.5****	***22.8****	1.6	2.9	***12.0****
I*	-10.3	X	***10.2****	-11.7	-9.2	0
H	-18.6	-9.2	X	-19.8	-16.3	-13.3
HH*	1.6	***13.2****	***24.8****	X	4.5	***13.7****
Mixed1*	-2.8	8.3	**19.5***	-4.3	X	8.9
Mixed2*	-10.7	-0.4	**15.3***	-12.1	-8.1	X

***significant difference (i.e., >10%)**

Although percent comparisons do not identify potential differences in usability depth between the G, I and Mixed1 sets, the Venn diagram (Figure 2) reveals that the Mixed1 set did not contribute any unique information. For example, the G set identified 16 usability needs not identified by either Mixed1 or I, while the I set identified 22 unique usability needs.

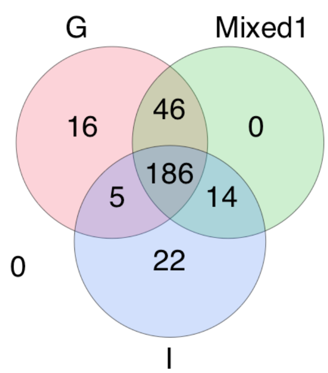

Figure 2. Comparison of usability depth for G, HH, and Mixed1 sets

DISCUSSION

Decision rules

The depth findings provide decision rules that can be used in conjunction with a baseline decision tree to provide guidance to engineers working on one of the four UCD goals identified in Table 3. As the formative usability task involves the collection of functional and usability needs, the rules provide decision guidance for the choosing the best storytelling context for eliciting this type of information (Table 7).

Table 7. Decision rules for method selection

Category	Decision rules	Notation
Functional	Choose G or HH over I, H, Mixed 2 methods	(G | HH) > (I | H | Mixed2)
	Choose Mixed 1 over H methods	Mixed1 > H
Usability	Choose G or HH over I or Mixed2 methods	(G | HH) > (I | Mixed2)
	Choose all over H	(G | HH | I | Mixed1 | Mixed2) > H

Decision trees

Baseline decision tree. A baseline tree was created to determine which methods are available to engineers within a variety of potential design constraints (Figure 3). As engineers do not work within ideal conditions, our decision tree provides recommendations for real-world situations in which the available methods are constrained by time and/or scheduling issues. Our decision tree accounts for two issues frequently cited in design research: limited time to conduct user research and potential issues with group sessions (e.g., scheduling issues, hierarchical issues between stakeholders). To allow for trade-offs between engineers' goals, our decision tree is not restricted to binary decisions. For example, in response to the decision "Groups problematic?", engineers may select "Maybe". In contrast to the "No" response, the "Maybe" path for this decision only excludes method combinations that include the sole use of group storytelling sessions. This provides more flexibility for engineers as the remaining available methods include combinations of group and individual sessions.

Altering storytelling context for use during a formative usability design problem. The combination of the decision rules for method selection (Table 7) with the baseline decision tree (Figure 4) yielded a decision aid for engineers interested in using storytelling to elicit functional and usability needs during formative usability design. For example, the decision tree guides engineers towards the use of individual sessions (i.e., an I set comprised of IHH and IH sessions) in situations where group sessions are not possible. However, if some group sessions are possible, the decision tree recommends the use of HH methods over I and Mixed2 methods.

Limitations

As all storytelling sessions were conducted with ER nurses during the concept stage of design, future work is needed to determine the generalizability of the results to other domains and design stages.

Future work

In future work we plan to create additional decision trees for other UCD tasks and to further refine the current decision rules to consider additional qualities, such as the uniqueness of needs identified.

Conclusion

Design recommendations based on empirical findings are needed as part of the PtD initiative. This study contributes design guidance based on empirical comparisons of storytelling contexts to the PtD effort.

Acknowledgements

This paper was supported by Grant number T01OH008613 from CDC NIOSH. Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of NIOSH.

REFERENCES

ANSI/AAMI. (2009). Human factors engineering - Design of medical devices.

Capra, M. G. (2006). Usability Problem Description and the Evaluator Effect in Usability Testing. (Doctor of Philosophy), Virginia Polytechnic Institute and State University, Blacksburg, VA.

Garmer, K., Liljegren, E., Osvalder, A.-L., & Dahlman, S. (2002). Arguing for the Need of Triangulation and Iteration when Designing Medical Equipment. Journal of Clinical Monitoring and Computing, 17(22), 105-114.

Gausepohl, K., Winchester, W. W., Arthur, J. D., & Smith-Jackson, T. (2011). Using Storytelling to Elicit Design Guidance for Medical Devices. Ergonomics in Design, 19(2), 19-24.

HRSA. (2008). The Registered Nurse Population: Findings from the 2008 National Sample Survey of Registered Nurses: U.S. Department of Health and Human Services.

IEC. (2007). Medical devices - Application of Usability engineering to medical devices.

IOM. (2001). Crossing the quality chasm: A new health system for the 21st century: Institute of Medicine.

Martin, J. L., Murphy, E., Crowe, J. A., & Norris, B. J. (2006). Capturing user requirements in medical device development: the role of ergonomics. Physiol Meas, 27(8), R49-62.

NORA. (2009). State of the Sector | Healthcare and Social Assistance.

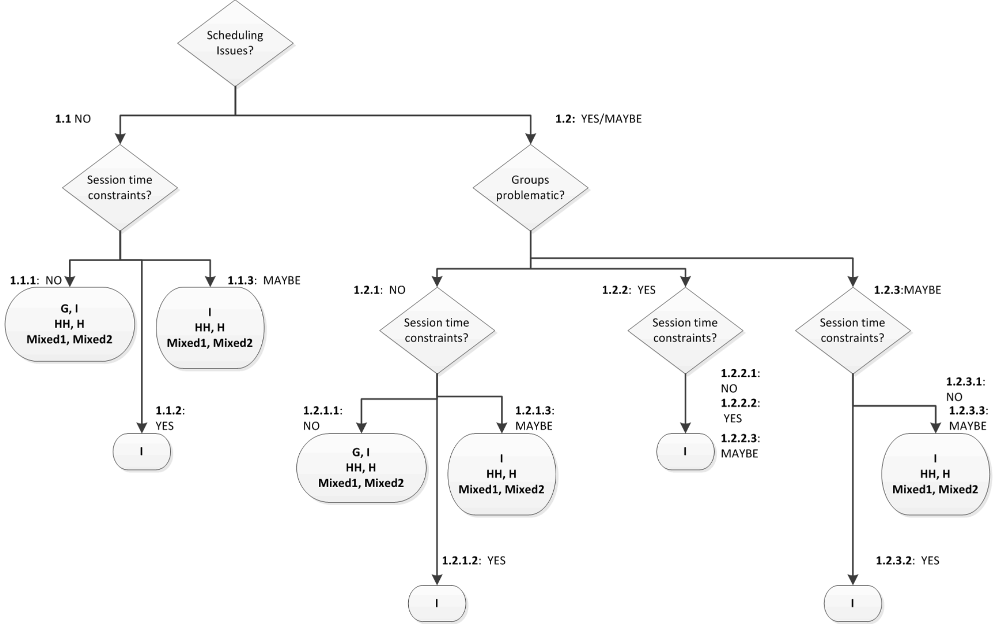

Figure 3. Baseline decision tree

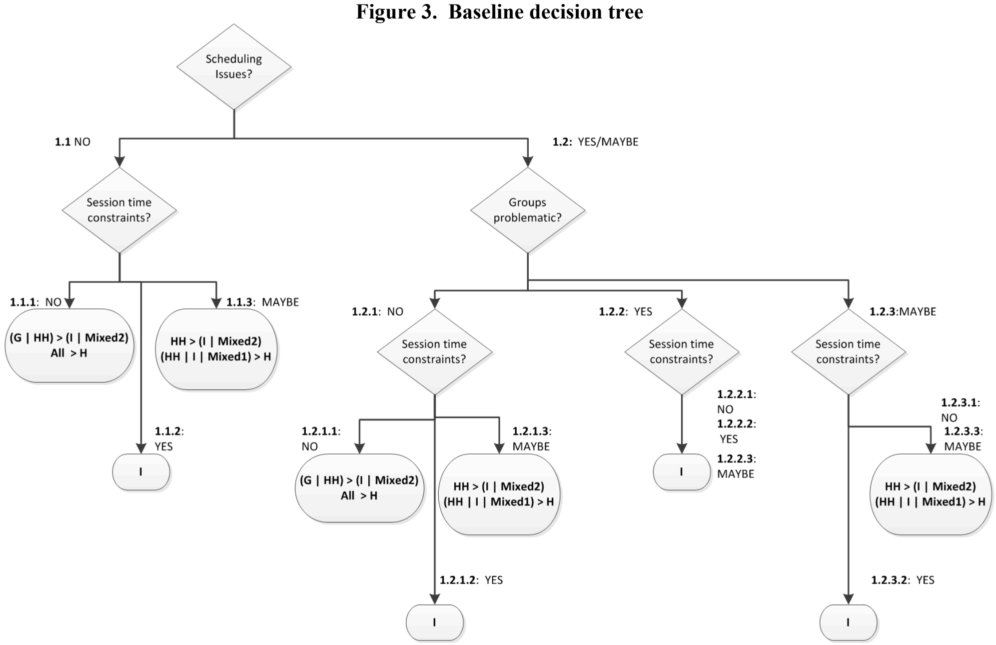

Figure 4. Decision tree for formative usability tasks