*Nursing Informatics 2016*
*W. Sermeus et al. (Eds.)*

*doi:10.3233/978-1-61499-658-3-530*

# Heuristic Evaluation of a mHealth Diabetes Self-Management System Using Disease Specific Patient Profiles

Mattias GEORGSSON, MSc, RN[a, b, 1] and Nancy STAGGERS, PhD, RN, FAAN[a, c]
[a] *Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA*
[b] *Faculty of Computing, Blekinge Institute of Technology, Karlskrona, Sweden*
[c] *College of Nursing, University of Utah, Salt Lake City, Utah, USA*

**Abstract.** Patient-centeredness is an important concept in diabetes treatment. We modified Nielsen's expert heuristic evaluation method addressing common gaps: a patient perspective and variability in findings. Two expert, dual-domain evaluators referred to validated patient profiles (mild, moderate, severe diabetes) when conducting uniform evaluation processes on a diabetes mHealth system. Evaluators found 103 usability problems and 224 heuristic violations. For 69 % of the problems, the profiles had an effect on severity ratings. "Consistency and Standards" (n=57) and "Match between the System and Real World" (n=55) violations dominated at 50%. The overall system severity rating was major. Severity was highest for a severe diabetic profile due to likely visual issues (crowded elements), cognitive concerns (remembering many steps) and for insufficient medication information. Interrater reliability was respectable at Kappa =0.67. Our novel evaluation method represents one way of improving on a usability expert technique making it more patient-centered with less individual evaluator variability.

**Keywords.** mHealth, diabetes, user-centered design, usability, heuristic evaluation, patient-centered evaluation

## 1. Introduction

About 29.1 million people in the U.S. suffer from diabetes [1] with over 90% having Type 2 diabetes [2]. About 95% of diabetes care is by the patients themselves [3] so self-management support is paramount. More pointedly, the American Diabetes Association 2015 standards indicate patient-centeredness should permeate all areas of disease and lifestyle management including team coordination [4, 5].

Patient-centeredness, or user-centeredness in informatics, is critical for the Information and Communication Technology (ICT) tools patients use. User-centeredness is one of the main concepts of User Centered Design (UCD) that incorporates users throughout the development and evaluation process in the engineering and human-computer interaction disciplines [6]. Usability, also central to UCD, can be measured by several different methods. One of the most common is Heuristic Evaluation (HE), an expert-based inspection method developed by Nielsen and Molich [7]. HE focuses on assessments by usability experts but excludes actual users, one of its major

[1] Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden.

drawbacks [8, 9]. A second drawback is that individual evaluators' expertise can vary considerably and result in variable study outcomes [10]. Our objective was to address these two gaps by using validated patient profiles in the expert-based heuristic evaluation process, having dual domain evaluators (usability experts and registered nurses) complete assessments and applying uniform HE evaluation processes. Users (patients) are not involved in HE. The novel method modifications presented can provide improved usability evaluations for contemporary expert-based techniques.

## 2. Methods

The diabetes system we tested is a commercially available mHealth system for diabetes self-management. It consists of a web portal and mobile phone solution. Patients enter values, such as glucose and blood pressures, using their mobile phone. They can track their medical progress, exercise, and medications on the web portal. After a review of validated diabetes instruments, we selected two applicable instruments for the patient profiles. These were descriptions from Grootenhuis et al.'s (1994) [11] Diabetes Symptoms Checklist (DSC-R) validated by Arbuckle et al. [12] that include Psychological cognitive, Neuropathic sensoric, and Ophthalmologic changes [11, 12]. We based ICT proficiency categories on the Computer Proficiency Questionnaire (CPQ) developed/validated by Boot et al [13] and designed for individuals with a wide range of proficiencies [14]. Profiles were for mild, moderate and severe conditions (see Table 1).

**Table 1** Patient profiles

| |
|---|
| ***Patient Profile 1*** *(Severity level 1) mild*<br>*Diabetes related: Visual acuity and sensoric abilities: good visual and sensoric capabilities. Cognitive ability: no difficulties in understanding, concentrating, ICT proficiency: high* |
| ***Patient Profile 2*** *(Severity level 2) moderate*<br>*Diabetes related: Visual acuity and sensoric abilities: more visual and sensoric difficulties. Cognitive ability: more difficulties in understanding, concentrating. ICT proficiency: medium* |
| ***Patient Profile 3*** *(Severity level 3) severe*<br>*Diabetes related: Visual acuity and sensoric abilities: low visual capability and major sensoric difficulties. Cognitive ability: major difficulties in understanding, concentrating. ICT proficiency: low* |

### *2.1. Evaluation Procedure*

Each evaluator had the same materials to ensure consistency – a video on portal navigation, a manual of the specific tasks, Nielsen's HE scale with severity ratings and the patient profiles. The evaluators used eight validated tasks to generate a list of usability problems. Sample tasks were viewing and locating glucose values on graphs, correcting values and viewing/exporting summary statements. Following Nielsen's HE process, the evaluators assigned Nielsen's 10 heuristics [15] to the identified usability problems. Problems were consolidated and discussed. A master list of all usability problems and heuristic evaluations was compiled and sent to each evaluator to determine severity ratings using the three patient profiles. For each usability problem, evaluators had to think about how the specific problem would affect someone with mild, moderate and severe diabetes. Nielsen's severity ratings were used: (0) not a usability problem, (1) cosmetic, (2) minor, (3) major, (4) catastrophic [15]. All ratings were averaged for each usability problem and the system as a whole. Percent agreement and interrater reliability

(IRR) Kappa were measured to determine the agreement on assigned severity ratings. Both measures are particularly fitting for assessing two evaluators and categorical variables to determine consistency. Plus, Kappa takes into account the estimated agreement beyond chance.

## 3. Results

The heuristic evaluation resulted in a total of 103 usability problems and 224 heuristic violations. The usability problems by place of occurrence (view), number of heuristic violations and mean severity ratings are summarized in Figure 1.

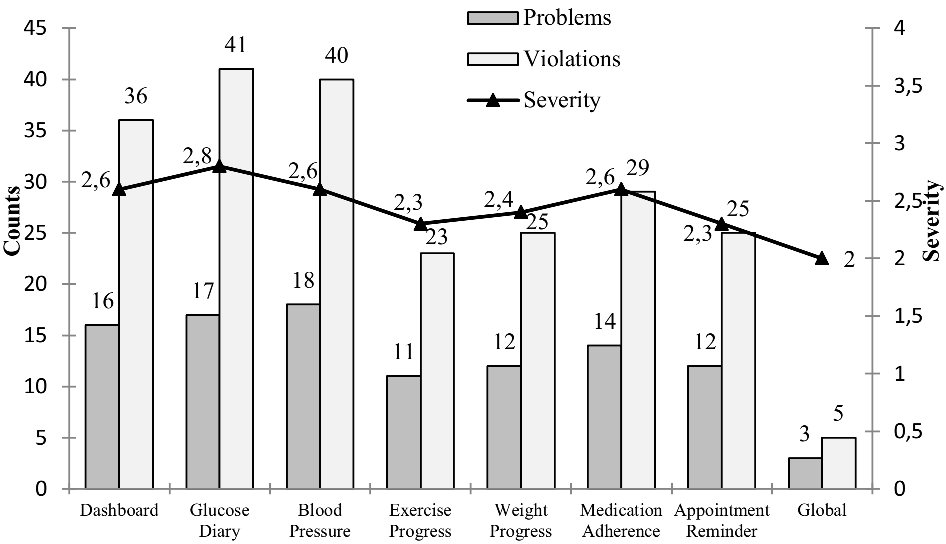


**Figure 1** Location frequency and averaged severity for usability problems and heuristic violations by place of occurrence

The Blood Pressure, Glucose Diary and Dashboard views had an almost equal number of usability problems at 18, 17 and 16 respectively and the highest number of heuristic violations at 40, 41 and 36. The severity rating for all views averaged 2.53 of 4 (or major severity). The highest severity ratings by view were the Glucose Diary at 2.8, followed by the Medication Adherence, Dashboard and Blood pressure view at 2.6.

A set of 224 heuristic violations was identified. “Consistency and Standards” violations were 57 (25.4%) followed by “Match between the System and Real World” at 55 (24.5%) or 50% combined. The highest number of violations/major issues was in the Dashboard and Blood Pressure views (11 each) followed by the Glucose Diary view (10) and Medication adherence view (7). Catastrophic severity ratings were most frequent in the Glucose Diary view and Medication Adherence with 3 and 2 problems respectively.

In 69% of the usability problems, the patient profiles had an obvious influence on severity ratings, meaning that the disease condition correlated with usability severity ratings. Sample issues in the Glucose values view included elements that were too crowded. The visual changes in patients with severe diabetes would likely result in the values being difficult to distinguish and read. Cognitive considerations meant a patient would have to remember too many steps when editing and exporting data, including memorizing system quirks such as using the non-intuitive Delete button to export data. Other problems were in the Medication adherence view due to insufficient medication information. Rather than using discrete data, the product only indicated a percent of medications taken (50%, 75% or 100%, for instance).

The percent agreement was 82% (84 problems of 103 rated the same), an excellent agreement level. Using SPSS, Kappa was 0.67 ($p < 0.001$) with a standard error (SE) of 0.067, indicating a substantial agreement across the raters [16].

## 4. Discussion

Results of this evaluation indicated a substantial need for improving the usability of this mHealth system. Usability violations in "Consistency and Standards" and "Match between System and the Real World" comprised 50% of all violations. Severity ratings indicated major issues for individual views as well as the overall system.

A majority of usability problems, 69%, were directly affected by the patient profiles, particularly where cognitive and visual changes might affect users' interaction performance. The new profiles assisted usability experts in thinking about and painting a clearer picture of specific users in this evaluation. Using patient profiles represents a novel application of methods for HE. In particular, it introduces a beginning patient perspective in a method that would be devoid of any patient-centered considerations.

The profiles were an important influence on how usability problems were viewed by evaluators compared to traditional HE methods which leave it up to each evaluator to imagine how users might interact. The patient profiles assisted in making the evaluation process more uniform and possibly making results more comparable and reproducible. The latter are important because mHealth technology is increasingly being implemented in health care [17]. By combining the profiles with more standardized HE methods, we attempted to decrease the variability of results across evaluators. This was reflected in the high agreement percentage at 82% and good IRR score at Kappa = 0.67 [16].

### *4.1. Limitations and Future Research*

The patient profiles described and used in this paper were based on evidence-based guidelines. However, we did make decisions about their content and delimiters, e.g., we predicted that patients with severe diabetes would likely have lower ICT capabilities due to visual and cognitive disease changes. It is possible that an individual patient with severe diabetes might instead have very good ICT skills. This is a limitation to the current work but necessary to develop a specific frame of reference for a beginning evaluation and to limit profile factors to a reasonable number. Future work in usability testing might assess the correlation between disease condition and ICT skills for more definitive data.

Future research could also focus on adding more dimensions to the patient profiles, more disease-related data, and a variety of ICT measures to obtain a wider variety of factors influencing patients in their mHealth interactions. For instance, age could be another dimension potentially making the systems even more difficult for these groups to use [18]. In this paper, we focused on disease progression but future researchers might add aging as an additional factor.

### 4.2. Conclusion

Expert usability evaluation techniques can be improved. Our results showed that patient profiles allowed usability experts to consider disease conditions during a common expert-based evaluation technique. Specific areas, such as medication views, had major and catastrophic issues for patients with severe disease conditions. Experts can anticipate that patients with more severe disease conditions may need improved support and different designs than patients with milder forms of the condition. The outlined modifications improve Nielsen's HE methods through more standardized/uniform methods, more consideration for specific users and less variability of results across evaluators.

## References

[1] Centers for Disease Control and Prevention, National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014., Department of Health and Human Services,: Atlanta, GA: U.S.
[2] World Health Organization (WHO), Diabetes Fact Sheet N°312. 2015 January 2015 [cited 2015 06 Sept]; Available from: http://www.who.int/mediacentre/factsheets/fs312/en/.
[3] M.M. Collins et al., Self-care coping strategies in people with diabetes: a qualitative exploratory study. BMC Endocr Disord, 2009. 9: p. 6.
[4] American Diabetes Association, Standards of medical care in diabetes – 2015. Diabetes Care, 2015. 38(Suppl 1): p. S1–94.
[5] M.A. Powers et al., Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. Diabetes Educ, 2015. 41(4): p. 417-30.
[6] International Organization for Standardization, Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. International standard, 9241-210 2010, Geneva: ISO.
[7] J. Nielsen, R. Molich, Heuristic evaluation of user interfaces, in Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people. 1990, ACM: Seattle, Washington, United States. p. 249-256.
[8] R. Jeffries, H. Desurvire, Usability testing vs. heuristic evaluation: was there a contest? SIGCHI Bull., 1992. 24(4): p. 39-41.
[9] T. Hollingsed, D.G. Novick, Usability inspection methods after 15 years of research and practice, in Proceedings of the 25th annual ACM international conference on Design of communication. 2007, ACM: El Paso, Texas, USA. p. 249-255.
[10] M. Hertzum, N.E. Jacobsen, The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interaction, 2003. 15(1): p. 183-204.
[11] P.A. Grootenhuis et al., Development of a type 2 diabetes symptom checklist: a measure of symptom severity. Diabet Med, 1994. 11(3): p. 253-61.
[12] R.A. Arbuckle et al., Psychometric evaluation of the Diabetes Symptom Checklist-Revised (DSC-R)--a measure of symptom distress. Value Health, 2009. 12(8): p. 1168-75.
[13] W.R. Boot et al., PRISM measure development: The design and validation of the computer proficiency questionnaire. 2014. 2014.
[14] W.R. Boot et al., Computer proficiency questionnaire: assessing low and high computer proficient seniors. Gerontologist, 2015. 55(3): p. 404-11.
[15] J. Nielsen, Heuristic evaluation, in Usability inspection methods, R.L. Mack and J. Nielsen, Editors. 1994, John Wiley & Sons inc.: New York. p. 25-62.
[16] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data. Biometrics, 1977. 33(1): p. 159-74.
[17] M. Georgsson, N. Staggers, Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc, 2015. DOI 10.1093/jamia/ocv099.
[18] J. Sharit et al., Investigating the Roles of Knowledge and Cognitive Abilities in Older Adult Information Seeking on the Web. ACM Trans Comput Hum Interact, 2008. 15(1): p. 3.