Medical Informatics Europe '97
C. Pappas et al. (Eds.)
IOS Press, 1997

An Example of Usability Measurement in Clinical Software Procedures

Adriana TERAZZI, Andrea GIORDANO, Giuseppe MINUCO
"Salvatore Maugeri" Foundation, IRCCS, Rehabilitation Institute of Veruno (Italy)
Service of Bio-Engineering

Abstract. As a consequence of the dramatic improvements achieved in information technology standards in terms of single hardware and software components, efforts in the evaluation processes have been focused on the assessment of critical human factors, such as work-flow organisation, man-machine interaction and, in general, quality of use, or usability.
This trend is particularly valid when applied to medical informatics, since the human component is the basis of the information processing system in health care context.
With the aim to establish an action-research project on the evaluation and assessment of clinical software procedures which constitute an integrated Hospital Information System, the authors adopted this strategy and considered the measurement of perceived usability as one of the main goals of the project itself: the paper reports the results of this experience.

1. Introduction

Technological evolution and organisational changes negatively affect the perceived quality of software procedures in use for a long time. Such was the case with the integrated Hospital Information System of Veruno Rehabilitation Institute [1,2]: after five years from its launching, most clinical software procedures were considered unsuitable to meet new requirements and the introduction of up-to-date solutions was urgently called for.
In order to plan the re-engineering process in an effective way, it was decided to carry out a 'Constructive Technology Assessment' project [3] to provide practical guidelines for the subsequent software design-development phase and to redefine product requisites starting from evolved users needs [4].
As a unique standard evaluation methodology for health care information technology has yet to be defined, the criteria put forward by EU-sponsored projects in the AIM area [3,6] were adopted. These involved the identification of the validation objectives [5] followed by a multidisciplinary intervention using appropriate available techniques and know-how from various R&D domains (e.g. social sciences, informatics).
Although a comprehensive evaluation would have been advisable for decision-making [7], the lack of skilled professionals as well as the complexity and high cost of that approach forced us to delimit the scope [5]. In our project, three different feasible and meaningful levels of analysis were singled out [6] and performed as routine iterative steps of the software life-cycle [8]:

- *Verification*
  Automated log procedures were introduced at module level to error-handling purposes [9]. In addition, the set-up of controlled functional scenarios was planned in order to guide system check and de-bugging operations during development and field-tests.

- *Validation*
  Integrity rules were constructed and applied to periodically verify consistency and completeness [10] of the complex database (287 relational tables, 4,308 administrative and clinical variables); the use ratio and type of single variables were investigated to optimise data structure and improve storage accuracy; the re-design of the user-interface was managed in close co-operation with end-users in an alternation of prototyping and control.
- *Human Factors Assessment*
  With medical informatics, it is mandatory to consider human concerns as peculiar issues [11]. In this context, user acceptance of information technology is fundamental to enhance performance and to hit the target of efficiency. It has been demonstrated that perceived usefulness and ease of use (subjective measures of quality of use) are valid indicators of user behaviour, more relevant to evaluate staff attitudes to automatic systems than the corresponding objective parameters [12,13,14]. Consequently, we chose perceived usability as a determining factor to be assessed.

In the next chapters the applied methodology and results of the last-mentioned analysis will be presented and discussed, leaving the other topics to future works.

2. Measuring Perceived Usability

2.1. Materials and Methods

Measurement Tool

The Software Usability Measurements Inventory, a world-wide validated questionnaire method adopted by several commercial and academic organisations, was chosen as a qualified tool, suitable to provide reliable measures of user perceived software quality (SUMI, © 1993 Human Factors Research Group, University College Cork, Ireland; ESPRIT project 5429, Measuring Usability of Systems in Context, MUSiC) [13,14,15,16,17].

Study Design

According to SUMI directives, only health care staff fully acquainted with our Hospital Information System applications [1,2] were enrolled in the research, i.e. all physicians, technicians and nurses working at Veruno Rehabilitation Institute with either one year of experience or who had been personally involved in the automated management of a significant amount of cases.
Since 10 users is the minimum recommended sample size to gather accurate measures, clinical software procedures had to be grouped into homogeneous packages before investigation, with the exception of Cardiology Ward application. In particular, two macro-procedures were respectively derived from in-house developed laboratory applications (Echocardiography, Dynamic ECG, Nuclear Medicine) and from those (Ergometry, Haemodynamics) produced by the same external partner; in each group, the procedures shared the same user-interface and functionality and differed merely in forms layout for minor operational details. Finally, three products were selected:

a. *In-house Cardiology Ward*
17 users (9,559 patients, 11,983 admissions)

b. *In-house Laboratory*
13 users (Echocardiography: 7,696 patients, 11,198 exams; Dynamic ECG: 2,652 patients, 3306 exams; Nuclear Medicine: 15,677 patients, 22,402 exams)

c. Third-Party Laboratory

14 users (Ergometry: 3220 patients, 4495 exams; Haemodynamics: 286 patients, 384 exams)

User anonymity was guaranteed to avoid biased results; however, a unique identifier (application code plus sequential number) was associated with each participant in order to enable the re-interview of the outliers.
Moreover, an additional introductory form with specific questions about age, gender, profession and experience was administered along with the SUMI questionnaire to categorise responses.

2.2. Results

Measurement data concerning in-house developed applications revealed a good level of user acceptance: except for Control [15,17], all other usability dimensions [15,17] resulted above the commercial standard with a Global score for software quality ≥ 60 (Fig.1); according to expectations, this was not the case for out-sourced products, with almost all scores below 50 (Fig.2). Furthermore, the confidence intervals plotted in Fig.1

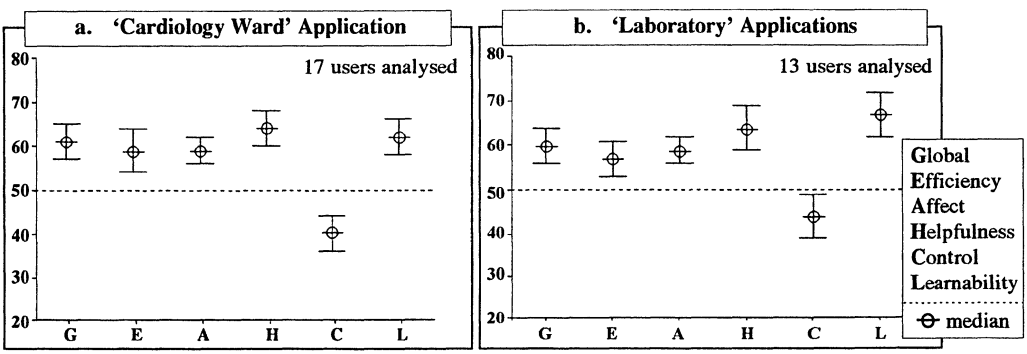

Fig.1. Usability profile for in-house developed clinical procedures.
The medians lie 95% within the confidence intervals (the opening and closing point of each bar correspond to the upper and lower confidence limit); the score-value 50 represents the standard for commercial software and is used as the term of comparison for quality of use: to be above or below 50 means to be ahead or behind the state of the art; Global scale is the general perceived usability benchmark or metric.

show a fair evaluation agreement among user groups (about 5 points around the median), whereas those in Fig.2 indicate more uncertainty (larger intervals).
The results of the applied Mann-Whitney U Test excluded any dependence between the category attributes (age, gender, profession and experience) and usability scores, with the exception of the helpfulness sub-scale related to the experience parameter ($p = 0.03$).
In order to determine whether or not differences between comparable products were significant, the same test was performed:

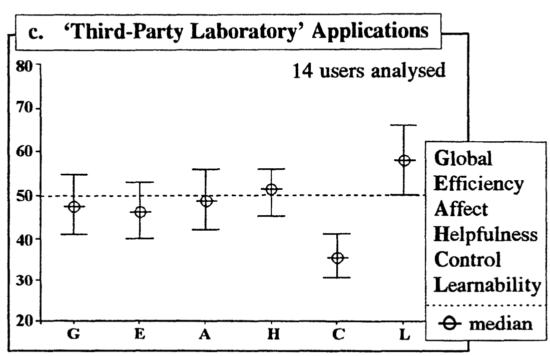

Fig.2. Usability profile
for out-sourced clinical procedures.

- ***a*** *vs* ***b*** (Fig.1):
  no statistical difference ($p > 0.05$).

- ***b*** *vs* ***c*** (Figs.1 and 2):
  the test yielded significant p values for 4 sub-scales, thus confirming the suspected hypothesis (in detail: Efficiency: 0.007; Helpfulness: 0.01; Control: 0.02; Learnability: 0.005; Affect: 0.057; Global: 0.051).

The Kruskal-Wallis ANOVA by Ranks Test confirmed the results.

3. Discussion

Research findings pointed out some interesting suggestions to support future operative choices.

The substantial equivalence between the assessment results for the two in-house applications studied (Fig.1) makes it possible to set up a single work-plan. First of all, notwithstanding the positive estimation which emerged globally, the weak points regarding Control sub-scale evaluation should not be underestimated. Consequently, users will be individually interviewed in order to clear up problems and find solutions; in this way, they will be involved in the design-development process as widely recommended [18,19].

The SUMI Item Consensual Analysis [15,17], which gives the detailed comparison between expected market trends and obtained response patterns, could be useful for locating critical areas; in our setting, significant negative discrepancies were documented for items connected to:

- users' autonomy in task accomplishment
  (the observed majority agreed with: '*The software has at some time stopped unexpectedly*', '*If this software stops it is not easy to restart it*', '*Getting data files in and out of the system is not easy*')

- system efficiency
  (the observed majority agreed with: '*This software responds too slowly to inputs*')

- software capacity to arouse users' interest
  (the observed majority disagreed with: '*Working with this software is mentally stimulating*').

Some remedial actions have already been taken towards performance enhancement (hardware upgrade, functional optimisation) and the integration of more attractive features (user-interface re-design, additional utilities such as network access to Medline and to Harrison's knowledge base as well as new routines for data import/export). Preliminary informal interviews confirmed that the users' perceived lack of control over the system was strongly influenced by unexpected programs faults limited to the initial period after deployment.

The dependence between the level of expertise and perceived helpfulness seems to prove that "where there's knowledge there's a way", stressing the importance of appropriate training. Since it has been demonstrated that sufficient time to learn and work with the technology combined with correct un-biased information positively affect both satisfaction and performance [20], future deployments will be preceded by trial-and-error learning sessions on simulated data until an adequate level of confidence has been reached.

Low global usability evaluation scores obtained for third-party procedures can be explained not only by certain functional inadequacies but also by an evident lack of

manager/end-user commitment; this lead to managers' indifference and users' resistance to products which are not psychologically considered as their own and suggests the need for careful adoption of commercial applications and, when possible, in-house development instead of outsourcing [19].
As soon as users have attained the required level of expertise regarding the upgraded applications, the study will be re-proposed to verify the failure or success of the re-engineering process and the validity of the technological and methodological choices undertaken.

References

[1] Giordano A, Terazzi A, Minuco G: A Specialized Hospital Information System for Rehabilitation Medical Centers: The SIR Project. In: *Proceedings of HEALTH TELEMATICS '95* (Eds: M.Bracale and F.Denoth), Area di Ricerca CNR -Pisa, 1995, pp. 499-504.
[2] Terazzi A, Giordano A, Minuco G: Evolution of an Integrated Hospital Information System. In: *Proceedings of Medical Informatics Europe '96* (Eds: J.Brender, J.P.Christensen, J.R.Scherrer, P.McNair), IOS Press, 1996, Part B, pp. 705-709.
[3] Van Gennip E.M.S.J, Talmon J.L.: Introduction. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 1-5.
[4] Saranummi N: Supporting System Development with Technology Assessment. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 35-44.
[5] Jorgensen T: Measuring Effects. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 99-109.
[6] Engelbrecht R, Rector A, Moser W: Verification and Validation. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 51-66.
[7] Richart R.H: Evaluation of a Medical Data System. In: *Computers and Biomedical Research*, 3, 1970, pp. 415-425.
[8] O'Moore R: The Conception of a Medical Computer System. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 45-49.
[9] Jorgensen T: Methods for Data Acquisition. In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 111-116.
[10] Barrie J.L, Marsh D.R: Quality of data in the Manchester orthopaedic database. In: *BMJ*, 304, 1992, pp. 159-162.
[11] Gremy F, Bonnin M: Evaluation of Automatic Health Information Systems. What and How? In: *Assessment and Evaluation of Information Technologies in Medicine* (Eds: E.M.S.J. van Gennip, J.L. Talmon), IOS Press, 1995, pp. 9-20.
[12] Davis F.D: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: *MIS Quarterly*, September 1989, pp. 319-340.
[13] Binucci L: Software Quality of Use: Evaluation by MUSiC. In: *Proceedings of Objective Quality '95* (Ed: P.Nesi), Springer, 1995, pp. 165-178.
[14] Classe A: Usability. In: *Software Quality Management*, 21, 1994, pp. 36-39.
[15] Kirakowski J: The Software Usability Measurement Inventory: Background and Usage. In: Jordan P, Thomas B, Weerdmeester B, *Usability Evaluation in Industry*, Taylor & Francis.
[16] Kirakowski J, Corbett M: SUMI: The Software Usability Measurement Inventory. In: *Brit J Educ Technol*, 24.3, 1993, pp. 210-212.
[17] Murray P, Kirakowski J, Corbett M: SUMI User Handbook. © 1993 Human Factors Research Group, University College Cork, Ireland.
[18] Philp I, Goddard A, Connel N.A.D, Metcalf A, Tse V, Bray J: Development and Evaluation of an Information System for Quality Assurance. In: *Age and Aging*, 23, 1994, pp. 150-153.
[19] Newman M, Sabherwal R: Determinants of Commitment to Information Systems Development: A Longitudinal Investigation. In: *MIS Quarterly*, 20.1, 1996, pp. 23-54.
[20] Griffith T.L, Northcraft G.B: Cognitive Elements in the Implementation of New Technology: Can Less Information Provide More Benefits? In: *MIS Quarterly*, 20.1, 1996, pp. 99-110.