Questionnaire Based Usability Evaluation of Hospital Information Systems

Kai-Christoph Hamborg[1], Brigitte Vehse[1], Hans-Bernd Bludau[2]
[1]University of Osnabrück, Germany
[2]University of Heidelberg, Germany
khamborg@uni-osnabrueck.de
bvehse@yahoo.de
hans-bernd.bludau@med.uni-heidelberg.de

Abstract: The widespread distribution of HIS requires professional evaluation techniques. In this study we present a usability questionnaire called IsoMetrics which is based on the international standard ISO 9241 Part 10. The questionnaire was applied to assess the usability of a Hospital Information System. The equivalence of the online and a paper-and-pencil format of the questionnaire were investigated. The results show that the different formats do not affect the subject's ratings. IsoMetrics was proven to be a reliable technique for software evaluation in the field of hospital information systems supporting usability screenings in large organisations.

Keywords: Evaluation, usability, ISO 9241 Part 10, Hospital Information Systems (HIS), online questionnaire

1. Introduction

The widespread distribution of Hospital Information Systems (HIS) in healthcare institutions requires professional evaluation to assess the practical usefulness of these applications. So far, evaluations of HIS have been undertaken focussing mainly on financial aspects or considering the patients interests. A major aspect has been neglected: The user! Nurses, physicians and other healthcare employees, working with the software, spend a lot of time each day by filling in forms, reviewing medical inspection results and handling an amount of information for administration needs.

The usability of a product is considered as a precondition of the usefulness of an application (Nielsen, 1993). It is defined with respect to " the extent to which the product can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specific context of use." (ISO 9241 Part 11, 1998). Unfortunately today not many applications fulfill this demand, and thus cause errors, trouble and stress as well as high costs on the part of the users and the organisation (Landauer, 1995).

Usability evaluation aims at identifying strengths and weaknesses of an application and gives hints for improving its usability. There is a multitude of methods for the purpose of software evaluation (Gediga, Hamborg & Düntsch, 2002). Questionnaires are well suited for the summative evaluation of software applications, especially in larger organisations like hospitals, public administrations etc. They are economic evaluation techniques which can be applied to a larger number of users at the same time with comparatively small financial effort.
In this paper the IsoMetrics Inventory (Gediga, Hamborg & Düntsch, 1999) for summative and formative evaluation of software usability will be presented. Its application in an evaluation study concerned with the usability of a HIS is demonstrated. In this study, we established an online version of the questionnaire, aiming at reducing efforts and at speeding up recurrent surveys and consecutive data evaluation. The equivalence of the paper-and-pencil and the online format is examined as well as the reliability of the questionnaire in the application area HIS.

2. Research methodology

The IsoMetrics questionnaire will be presented in the context of an evaluation study which was conducted at the University Hospital of Heidelberg, Department of Internal Medicine.

2.1 Material and methods

2.1.1 The IsoMetrics questionnaire

The IsoMetrics usability inventory (Gediga, Hamborg & Düntsch, 1999) provides a user-oriented, summative as well as formative approach to software evaluation on the basis of ISO 9241 Part 10. While summative evaluation is typically quantitative and located at the end of a development process, using numeric scores to assess the usability of an application, formative evaluation provides (often qualitative) information about weaknesses useful in improving the usability of a software system during the engineering life cycle or before further development. Accordingly there are two versions of

IsoMetrics, both based on the same items: IsoMetrics^S (short) supports summative evaluation, whereas IsoMetrics^L (long) is best suited for formative evaluation purposes. The inventory is available as English and German language version and can be administered by either paper and pencil or an online (inter-/intranet) version. The current version of IsoMetrics (2.04 german/2.01 english) comprises 75 items operationalising the seven design principles of the international standard ISO 9241 Part 10. ISO 9241 formulates „Ergonomic requirements for office work with visual display terminals (VDTs)" and provides guidance for the ergonomic design of interactive software. It comprises 17 different parts, whereas Part 10 covers seven principles for dialog design (s. Table 1).

Table 1: Dialogue Principles according to ISO 9241 Part 10 (translated from the german version by the authors).

Suitability for the task	A dialogue is suitable, if it supports the user to realise his tasks effectively and efficiently. Only those parts of the software are presented, which are necessary to fulfil the task.
Self-descriptiveness	A dialogue is self-descriptive, if every step is understandable in an intuitive way, or, in case of mistakes supported by immediate feedback. Further, an adequate support should be offered on demand.
Controllability	A dialogue is controllable, if the user is able to start the sequence and influence its direction as well as speed till he reached his aim.
Conformity with user expectations	A dialogue is conform with the user expectations, if it is consistent, complying with the characteristics of the user, e.g. taking into account the knowledge of the user in that special working area, accounting education and experience as well as general acknowledged conventions.
Error tolerance	A dialogue is error tolerant, if the intended deliverable is reached with no or just minimal additional effort despite of obvious faulty steering or wrong input.
Suitability for individualisation	A dialogue is suitable for individualisation, if the system allows customising according to the task as well as regarding the individual capabilities and preferences of a user.
Suitability for learning	A dialogue supports the suitability of learning, if the user is accompanied through different states of his learning process and the effort for learning is as low as possible.

The statement of each item of the IsoMetrics^S Questionnaire is assessed on a five point rating scale starting from 1 ("predominantly disagree") to 5 ("predominantly agree"). A further category ("no opinion") is offered to reduce arbitrary answers.

IsoMetrics consists of the same items as IsoMetrics^S and uses the same rating procedure. Additionally, each user is asked to give a second rating, based upon the request "Please rate the importance of the above item in terms of supporting your general impression of the software." This rating ranges from 1 ("unimportant") to 5 ("important"). A further "no opinion" category may also be selected. In this way, each item is supplied with a weighting index. To evoke information about malfunctions and weak points of the system under study, the question "Can you give a concrete example where you can (not) agree with the above statement?" is posed. This gives users the opportunity to report problems with the software, which they attribute to the actual usability item.

IsoMetrics has proved its practicability in software development projects and field studies. Given ten evaluating users, IsoMetrics^L evokes approximately one hundred remarks addressing weak-points of a given software. Its reliability was examined and confirmed for each of the seven design principles (Gediga, Hamborg & Düntsch, 1999, Gruber, 2000). In order to validate the IsoMetrics inventory, the scale means of five different software systems were analysed and compared. It could be shown that programs with different ergonomic qualities were discriminated by the corresponding scales (Gediga, Hamborg & Düntsch, 1999).

2.1.2 Software

The software examined, "IS-H*MED" release 4.63B by T-Systems, Austria is based on the IS-H solution by SAP, Germany. It is mainly table-oriented software with a broad range of functions:

- *Creation of discharge letter*: A discharge letter is most often dictated on tape by a physician and afterwards typed by a secretary. Proof-reading and corrections are realised online, using a MS Word plug-in.
- *View of laboratory and diagnostic findings*: For each patient, an overview of existing laboratory and diagnostic findings is available. A list of the findings permits the physician a detailed look.
- *Documentation of diagnostic finding*: In-patients can be selected by a physician from a listing of the patients to feed in diagnostic findings. The ICD10-Code of the diagnoses might be entered directly – or with the help of a plug-in called KODIP. This plug-in covers the complete ICD-10 via a thesaurus and offers additional information about the grouping accounting rules etc.
- *Diagnose related grouping*: After the individual diagnostic findings and resulting

medical procedures (e.g. a heart catheter) are entered into the computer, a calculation of the Diagnose Related Group's (DRG) may be accomplished.

- *Order of medical examinations* supports the electronical ordering of medical examinations for a patient.
- *Documentation of physical examinations*: This function allows to document the results of an inspection, e.g. an ultrasound sonic examination, or a radiology report. The reports are mainly written with help of a MS Word plug-in.
- *Nursing category*: A staffing calculation methodology derived from the traditional nursing hour per patient day (HPPD), taking into account a systematic approach estimating effort for a patient with a specified disease.
- *Meal order*: The meal order starts with a listing of the in-patients on a ward. Detailed orders according to the needs of the patients can be entered.

2.2 Preliminary enquiry

Before the evaluation study started, a preliminary enquiry was conducted to collect personal data of the potential participants (nurses, doctors, secretaries and other staff of the department). For that purpose a questionnaire was applied addressing computer-experience in general as well as experience with IS-H*MED, area of work, used IS-H*med functions, age and sex. 182 persons completed the questionnaire and were willing to take part in the subsequent evaluation study. Results of the survey were treated confidentially.

By means of a cluster analysis six "user-types" according to the used IS-H*med functions were discriminated (see table 2). Moreover three user categories were distinguished due to the general as well as the IS-H*med specific experience: Novices, intermediate and expert users.

Table 2: Specification of the user types (percentage in brackets illustrate how many persons of a user type use the corresponding functions.)

User types	Used IS-H*MED functions
„Prevailing medical secretaries" (user type 1) N = 39	View of laboratory and diagnostic findings (67%) Creation of discharge letter (54%) Documentation of physical examinations (51%) Order of medical examinations (49%) Other (15%)
„Physicians "(user type 2) N = 41	Documentation of diagnostic findings (100%) Diagnose related grouping (100%) View of laboratory and diagnostic findings (93%) Order of medical examinations (93%) Creation of discharge letter (90%) Documentation of physical examinations (59%)
"Nursing staff I"(user type 3) N = 60	Meal order (100%) Nursing category (95%) Diagnose related grouping (88%) Order of medical examinations (80%) Other (25%) Documentation of physical examinations (22%) View of laboratory and diagnostic findings (12%)
"Prevailing physicians (user type 4) N = 22	Documentation of diagnostic findings (100%) View of laboratory and diagnostic findings (86%) Creation of discharge letter (82%) Order of medical examinations (82%) Documentation of physical examinations (73%) Other (14%)
"Prevailing nursing staff"(user type 5) N = 11	Other (91%) Meal order (82%) Diagnose related grouping (55%) View of laboratory and diagnostic findings (27%) Documentation of physical examinations (27%) Order of medical examinations (18%)
"Nursing staff II"(user type 6) N = 9	Nursing category (100%) Meal order (100%) Order of medical examinations (89%) Documentation of physical examinations (89%) Other (78%) Diagnose related grouping 22%)

2.3 Main inquiry

2.3.1 Participants and procedure

The evaluation study was conducted in January and February 2003. Participation was voluntary, no financial incentives were offered. We received 132 responses (online as well as paper-and-pencil Questionnaires) from the 182 participants who took part in the preliminary study and from additional spontaneous participants.

After the exclusion of questionnaires with too much missing data (s. chapter 2.3.2 Data analysis) 106 responses remained. Mean age of these participants was 38 years (SD = 8,81; range: 24-61 years). 55 persons (51,9 %) were female, 36 (34 %) male; 15 participants (14,1 %) did not answer the question about their gender. According to computer-experience, the sample included 22 novice, 27 intermediate and 30 expert users. 27 persons did not give information about their general computer experience or their experience with the IS-H*MED system.

2.3.2 Data analysis

Questionnaires with more than 20% missing data (more than 15 items not answered) were excluded from further analysis (s. Gediga, Hamborg & Willumeit, 1998). In case of less or equal than 15 omissions, missing values were replaced by the mean scale value (‘3’) of the items. The same procedure was applied if the answer was 'no opinion'. This procedure was controlled by comparing reliabilities based on the records without missing data with the reliabilities based on the records with replaced missing data. The procedure showed no differences of the reliabilities. Some items of the questionnaire (A1, A8, T12, E8, F1, F7, F14, L1, and L7) are formulated negative. The values of these items were inverted by the transformation $r_i' = 6 - r_i$ for further analysis.

To analyse the equivalence of the paper-and-pencil and the online version of the IsoMetrics questionnaire two matched groups (à N = 29) from a sub sample of all participants were established with regard to the user-type, computer experience, age and sex. The equivalence of both formats was assessed with respect to the scale mean values and reliabilities.(For a detailed description of this analysis, see Hamborg, Vehse, Ollermann & Bludau, 2004).

The reliability of the scales was computed in a next step. After that, the mean values for both questionnaire formats were calculated to assess the ergonomic quality of the application. Moreover the IS-H*med profile was compared with the profiles of two reference systems.

For the ergonomic quality of software systems should be assessed with respect to the context of use (user, task, equipment and environment, see ISO 9241 Part 11, 1998) an analysis of variance with user-type and computer experience (user group, experience, etc.) as independent and the 7 IsoMetrics-Scales as dependent variable was calculated. To identify special differences between the identified user-types, post-hoc tests have been calculated. To get more detailed information about ergonomic shortcomings of the software, ratings of the single IsoMetrics items were analysed at least.

2.4 Results

2.4.1 Equivalence of the online and paper-based questionnaire

Analysis of the scale means revealed no marked differences between the two matched samples using the online respectively the paper-and-pencil version of IsoMetrics^S (table 3).

Table 3: Means of the online- and paper-pencil version

IsoMetrics Scale	Format of the Questionnaire	N	Mean	SD	T	df	sig. (2-sides)
Suitability for the task (15)	Online	29	2.54	.793	-.771	56	.444
	Paper	29	2.70	.841			
Self-descriptiveness (12)	Online	29	2.33	.764	-1.559	56	.125
	Paper	29	2.66	.835			
Controllability (11)	Online	29	2.72	.795	-1.423	56	.160
	Paper	29	3.03	.864			
Conformity with user expectations (8)	Online	29	2.87	.634	-1,534	56	.131
	Paper	29	3.13	.692			
Error tolerance (15)	Online	29	2.61	.586	-.683	56	.497
	Paper	29	2.72	.618			
Suitability for individualisation (6)	online	29	1.94	.763	-.245	48.39	.807
	Paper	29	2.00	1.161			
Suitability for learning (8)	online	29	2.52	.708	-.784	56	.437
	Paper	29	2.70	1.058			

Reliabilities (Cronbach´s Alpha) of the IsoMetrics subscales were checked and proved to be at least satisfactory (table 4). As well as the scale means, the reliabilities of the IsoMetrics version are not different except for the subscale “suitability for individualisation” (table 4). Within the scope of power analysis we checked whether the sensitivity of the tests was good enough to detect substantial mean differences. A half point mean difference between the online and the paper-and-pencil format was taken as the lower bound of our interest. Data analysis revealed that all tests would have been able to detect this difference. Because the empirical data didn´t show any difference larger than 0.5 and no significant results, the profiles of both formats were considered as equal.

Table 4: Analysis of reliabilities (Cronbach's alpha) of the paper-and-pencil and the online version of IsoMetrics

IsoMetrics scale	Paper-pencil N = 29	online N = 29	Z (paper-pencil vs. online)	IS-H*med overall means	overall Rel. N = 106
Suitability for the task	0.910	0.921	-0.25	2.77	.90
Self-descriptiveness	0.917	0.901	0.33	2.68	.90
Controllability	0.873	0.849	0.34	2.97	.86
Conformity with user expectations	0.704	0.708	-0.03	3.06	.71
Error tolerance	0.791	0.780	0.10	2.85	.84
Suitability for individualisation	0.962	0.849	2.60*	2.12	.90
Suitability for learning	0.918	0.817	1.55	2.84	.87

The results concerning the mean values and reliabilities corroborate the assumption that the two formats can be treated as equivalent. Therefore data of the online and the paper-and-pencil version were merged for further analysis.

2.4.2 Rating of the systems ergonomic quality

The following results give an overview of the overall rating of the system. The scale mean values of the ratings range between 2 and 3 (except the scale „ conformity with user expectations“, which is slightly above, see table 4).

Accordingly the ergonomic quality of the system as assessed by its users can be considered quite low.

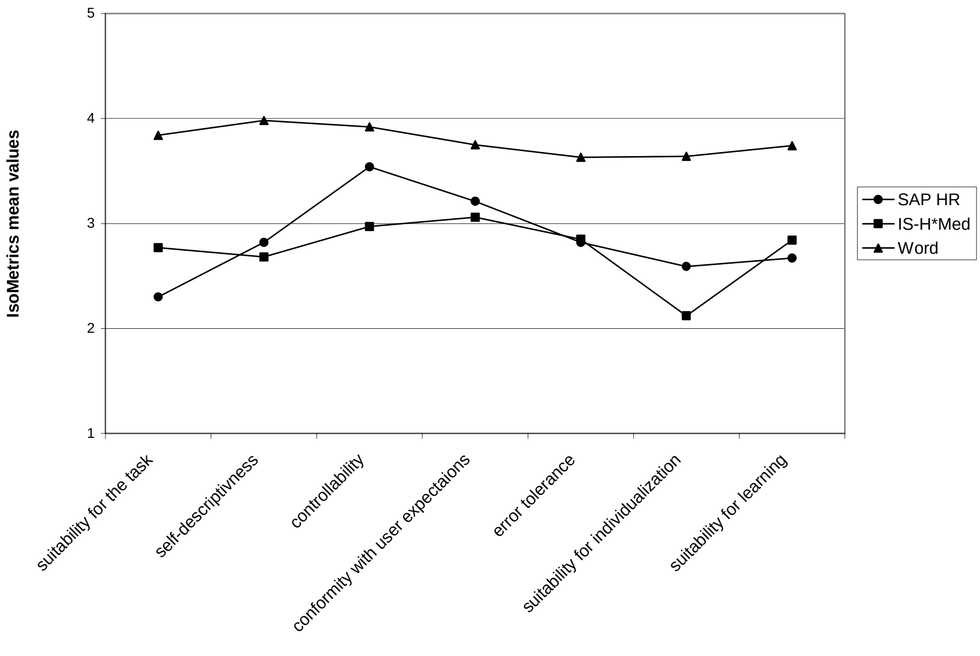

Figure 1: IsoMetrics scale means of IS-H*med and reference systems

2.4.3 Comparison with reference systems

The IS-H*med IsoMetrics mean value profiles were compared with the profiles of two other applications: a) SAP-HR resulting of a study conducted by Gruber (2000) with IsoMetrics, version 2.03 (N = 28) and b) Microsoft Word for Windows (Version 2) which was evaluated in a study by Gediga, Hamborg and Düntsch (1999) with an previous IsoMetrics version (N = 55).

SAP HR is an application supporting several tasks in the field of human resources management like personnel administration, personnel time management, training and event management and payroll accounting. WinWord is the word processing software from Microsoft. The reliabilities of the IsoMetrics scales also were proven to be at least good in this study (table 5).

Table 5: IsoMetrics mean values, standard deviations and reliabilities for SAP R/3 HR (Gruber 2000) and Microsoft WinWord (Gediga, Hamborg & Düntsch, 1999)

	SAP HR			**Microsoft Word**		
IsoMetrics scale	**Reliability**	**Mean**	**SD**	**Reliability**	**Mean**	**SD**
Suitability for the task	.92	2.30	.72	.53	3.84	0.38
Self-descriptiveness	.84	2.82	.63	.82	3.98	0.36
Controllability	.77	3.54	.50	.79	3.92	0.76
Conformity with user Expectations	.77	3.21	.64	.70	3.75	0.70
Error tolerance	.78	2.82	.52	.63	3.63	0.40
Suitability for individualisation	.73	2.59	.71	.78	3.64	1.02
Suitability for learning	.80	2.67	.67	.54	3.74	0.65

The mean values of IS-H*med and the reference systems were compared using t-values and t-tests respectively. In contrast to SAP-HR, IS-H*med was significantly rated higher on the scale “suitability for the task” and lower on the scales “controllability” and “suitability for individualisation”. It has to be mentioned, that the effect sizes (d, Cohen, 1977) of this tests are quite small. Therefore the observed mean value differences can’t be considered as substantial.

Table 6: Comparison IS-H*med vs SAP-HR

IsoMetrics scale	**Mean**		**t**	**df**	**p**	**d**
	IS-H*med	**SAP-HR**				
Suitability for the task	2.77	2.30	2.96	45.97	0.005*	0.19
Self-descriptiveness	2.68	2.82	-1.04	5.56	0.305	
Controllability	2.97	3.54	-4.62	68.20	< 0.0001*	0.23
Conformity with user expectations	3.06	3.21	-1.02	43.97	0.312	
Error tolerance	2.85	2.82	0.30	54.43	0.762	
Suitability for individualisation	2.12	2.59	-2.85	54.82	0.006*	0.19
Suitability for learning	2.84	2.67	1.13	55.22	0.262	

The comparison of IS-H*med and MS Word reveals significant differences on all scales (empirical values exceed the critical t-value, df = 157, p = 0.007, see table 7). The effect sizes (d) and the mean value differences respectively are large (d > .8, Cohen, 1977, p. 26).

Table 7: Comparison IS-H*med vs MS Word for Windows (Version 2.0)

IsoMetrics scale	**Mean**		**t-Value**	**P**	**D**
	Word	IS-H*med			
Suitability for the task	3.84	2.77	9.39	< 0.0000	1.55
Self-descriptiveness	3.98	2.68	11.48	< 0.0000	1.90
Controllability	3.92	2.97	7.23	< 0.0000	1.19
Conformity with user expectations	3.75	3.06	6.11	< 0.0000	1.01
Error tolerance	3.63	2.85	7.77	< 0.0000	1.28
Suitability for individualisation	3.64	2.12	9.42	< 0.0000	1.56
Suitability for learning	3.74	2.84	6.63	< 0.0000	1.10

It can be concluded, that the usability of IS-H*med is rated low from the user’s point of view and does not meet the ergonomic quality of a standard windows application in this connection but doesn’t differ from other ERP applications substantially.

2.4.4 Evaluation results due to user types

Up to now we have just looked at the pooled ratings of all users taking part in the study. In consideration of the statement that the usability of a software depends on its context of use, we analysed the data with respect to the identified user types.

The IsoMetrics profiles of the different user types are quite similar but located on different levels (figure 2). IS-H*med was rated low especially by doctors while medical secretaries and nursing staff gave the best ratings, but being far from good.

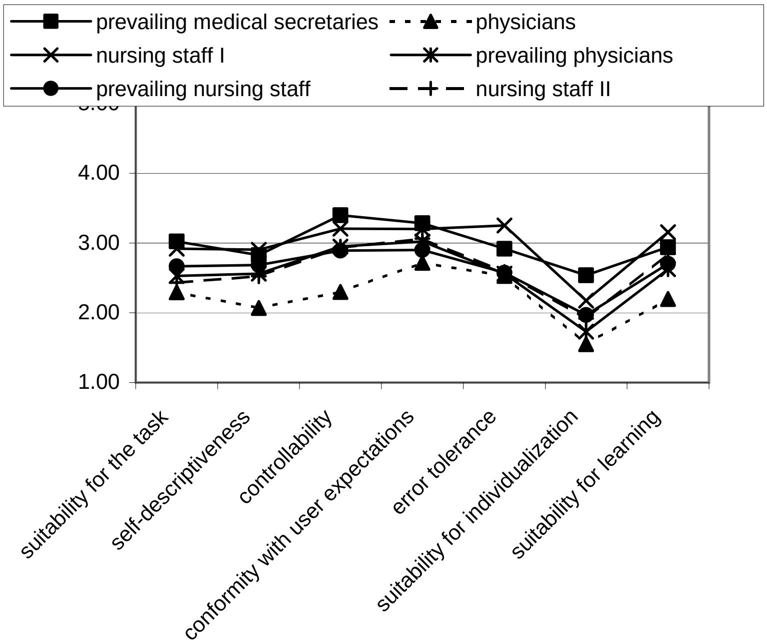

Figure 2: Rating of IS-H*MED according to user types

Statistical analysis shows a significant overall difference between user types on all scales except the scale „Conformity with expectation". Significant post hoc contrasts were found between „Physicians" (user group 2) „Medical secretaries" (user group 1), and "Nursing staff I" (user group 3). Physicians rated the systems "Suitability for the task", "Self-descriptiveness", "Controllability" and "Suitability for learning" worse than "Nursing Staff I" and "Medical Secretaries". "Error tolerance" was rated significantly better by "Nursing staff I" (user group 3) compared to "Physicians".

There are different explanations for this finding. First, by definition of the "user types", "Physicians" are using other functions than "Medical Secretaries" do. It can be assumed that the functions used by secretaries are of higher ergonomic quality. Especially the MS Word plug-in might have influenced the comparatively good rating of the medical secretaries

Secondly, different ratings might depend on the specific IS-H*med experience of the identified "user-types". Results of demographic data show that there are more novices than expert IS-H*med users among doctors. In contrast, nursing staff comprises of more experts than novices. A large proportion of nursing staff (43,8 %) has been working with the software for up to 62 months whereas a large part of the doctors (46,8 %) have been using the software for only up to 12 months. The user type of "Medical Secretaries" included no novice users, four intermediate and four expert users. The low experience of the physicians may be due to the fact that the university hospital is a teaching hospital with a high turnover of physicians dependent on a training scheme.

Especially the low ratings of the scales "Suitability for the task", "Self-descriptiveness", "Controllability" and "Suitability for learning" received from the physicians should be taken seriously. Its causes should be scrutinised in detail with respect to remedial action.

2.4.5 Item score analysis

So far we have discussed results on the basis of mean value profiles. This kind of analysis allows the assessment of a given software according to standard design principles, benchmarkings with other applications and the identification of critical usability aspects for specific user groups or task profiles.

In the following we demonstrate an analysis approach that provides some more concrete information about the usability of an evaluated system. It is based on the analysis of the mean values of the single items belonging to the different IsoMetrics scales. This analysis gives a more refined picture of the ratings in comparison to the examination of the scale means presented so far.

To exemplify this approach, we will take a look at the scale "Suitability for the task". There are some obviously high and low rated items. Analysing the wording of these items in association with the item mean values we get some more detailed information about negative and positive usability aspects of the software.

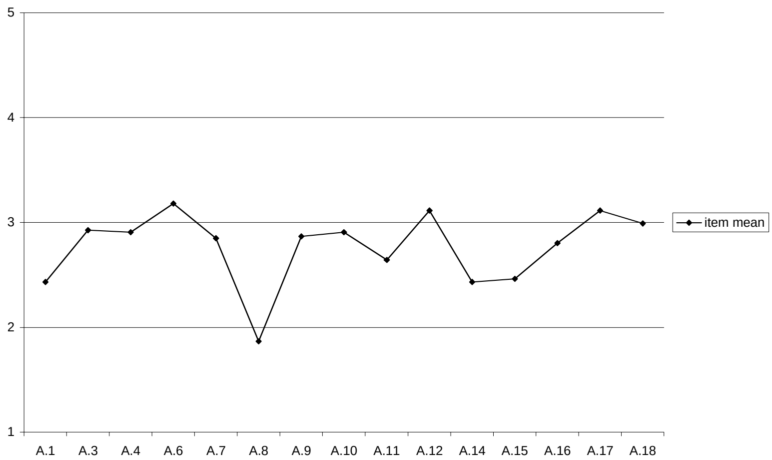

Figure 3: Mean values of the subscale „suitability for the task"

The item means and the standard deviations are shown in brackets after the item description.

Comparatively positive rated items are:

- A.6: The way, in which data is entered is suited to the task I want to perform with the software (M = 3.18; SD = 1.14).
- A.12: The terminology used in the software reflects that of my work environment (M = 3.11; SD = 1.3).
- A.17: I am able to adjust the presentation of results (on the screen, to printer, plotter etc.) to my various work requirements (M = 3.11; SD = 1.26).

Salient negative rated items are:

- A.1: The software forces me to perform tasks, that are not related to my actual work (M = 3.57; SD = 1.28).
- A.8: Too many different steps need to be performed to deal with a given task (M = 4.13; SD = 1.23). (Because the items A.1 and A.8 are formulated negatively the coding has been reversed for statistical data analysis)
- A.14: The software provides me with repeat function for work steps that must be performed several times in succession (M = 2.43; SD = 1.27).
- A.15: I can easily adopt the software for performing new tasks (M = 2.46; SD = 1.12).

Regarding this items, the software appears to conform at least in part to task requirements, especially with respect to data input, adaptability of data output and the wording. Although these items have been rated comparatively positive the system rating was not better than moderate (3). Hence, it cannot be concluded that the software meets the design requirements of this principle (suitability) sufficiently. The low rated items emphasise that the performance of tasks can be awkward and needs extra effort with the evaluated application. Accordingly need for improving the usability of the evaluated software is specified.

3. Discussion

In this paper the IsoMetrics questionnaire was presented. IsoMetrics is a technique to evaluate the usability of software applications with respect to the international standard ISO 9241 Part 10. It could be shown that IsoMetrics is a well suited and reliable technique in the application area of HIS. With help of the online version, that behaves equivalent to the paper-and-pencil format, evaluation studies will become more efficient with respect to the distribution of questionnaires, data management and data analysis.

IsoMetrics was applied to evaluate a HIS in a field study. Results revealed low ergonomic quality of the evaluated system. The usability of a given software should be treated with respect to given context of use. We demonstrated how to use the IsoMetrics questionnaire to identify critical usability aspects with respect to so called user types, defined by special users dealing with special tasks and functions of the software. Apart from the general low rating of the evaluated software, need for remedial action was identified especially for the group of physicians. The analysis of single items of the questionnaire gives some more concrete hints to weak-points of the evaluated software.

However, principal limits of the application of questionnaires for summative evaluation have to be considered: they will just provide general hints to problem areas of given software but they are not able to detect *concrete* weaknesses nor will they reveal the *causes* why users attribute lack of usability to software. For this purpose, a deeper analysis of problem areas identified by means of a summative questionnaire should be conducted with the help of evaluation techniques like user tests (Dumas & Redish, 1999), walkthroughs (Bias, 1991) or IsoMetrics[L].

Especially in large organisations an incremental strategy of usability evaluation seems to be adequate: carry out a screening with a summative questionnaire and identify general problematic usability aspects. After that conduct a more thorough formative evaluation for the identified and most critical usability aspects to reveal the causes for the lack of usability and derive a remedial plan for action.

This kind of systematic evaluation of Hospital Information Systems will support the clinical work of health care employees by adapting the software to user requirements, improving its functionality continuously and avoid errors and stress reactions as well as the costs associated herewith.

4. Acknowledgements

We would like to express special thanks to Alexander Schubek for proofreading this paper.

References

Cohen, J *Statistical power analysis for the behavioural science* (revised edition). Academic Press, New York (1977).

Bias, R `Walkthroughs: Efficient Collaborative Testing´ *IEEE Software,* Vol 8 Issue 5 (1991) pp 94-95.

Dumas, J S and Redish, J C *A Practical Guide to Usability Testing* (Revised Edition), Intellect Books, Exter (1999).

Gediga, G, Hamborg, K-C and Düntsch, I `The IsoMetrics Usability Inventory: An operationalisation of ISO 9241-10´ *Behaviour and Information Technology*, Vol 18 No 3 (1999) pp 151-164.

Gediga, G., Hamborg, K.-C. and Düntsch, I `Evaluation of Software Systems´ In: A. Kent and J. G. Williams, eds. *Encyclopedia of Computer Science and Technology*, Vol 44, Marcel Dekker Incorporated (2002) pp 162-196.

Gediga, G, Hamborg, K-C and Willumeit, H *The IsoMetrics Manual* (Osnabrücker Schriftenreihe Software-Ergonomie Nr. 7). Osnabrück: Universität Osnabrück, Fachbereich Humanwissenschaften (1998).

Gediga, G and Hamborg, K-C `IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241-10´ In: H. Holling and G. Gediga, eds. *Evaluationsforschung*. Göttingen: Hogrefe (1999) pp 195-234.

Gruber, C (2000). Arbeitseffizienz im Büro. *Psychische Einflüsse auf SAP R/3-unterstützte Arbeitsprozesse*. University of Würzburg, Master thesis [Online]. Available from: http://www.clemens-gruber.de/psychologie/sapr3/ [Accessed: 06.03.2004].

Hamborg, K-C, Vehse, B, Ollermann, F and Bludau, H-B *Softwareevaluation: Zur Äquivalenz der Online- und Papier-Bleistiftversion des IsoMetrics[S] Verfahrens* (Osnabrücker Schriftenreihe Software-Ergonomie Nr. 8). Osnabrück: Universität Osnabrück, Fachbereich Humanwissenschaften (2004).

ISO 9241-10 Ergonomic requirements for office work with visual display terminals (VDTs): Dialogue principles (1996).

ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs): Guidance on usability specifications andd measures (1998).

Landauer, T K *The Trouble with Computers. Usefulness, Usability and Productivity*. MIT Press, Cambridge MA (1995).

Nielsen, J *Usability Engineering*. AP Professional, Boston (1993).