

UsersAward – first Results of a Pilot Study in Germany

Jochen Prümper

FHTW – Berlin

University of Applied Sciences

Treskowallee 8

D-10313 Berlin

j.pruemper@fhtw-berlin.de

Thomas Vöhringer-Kuhnt & Jörn Hurtienne

bao GmbH

Büro für Arbeits- und Organisationspsychologie

Kösterstraße 1b

D-14165 Berlin

t.kuhnt@bao.de, j.hurtienne@bao.de

ABSTRACT

Within the framework of a project on the transfer of established methodologies for measuring and enhancing user satisfaction with workplace software from Sweden to other European countries, the Swedish UsersAward questionnaire has been adapted to German demands. Together with the questionnaire “ISONORM 9241/10”, it was used as part of a pilot study examining customer satisfaction with eight different software products. 90 participants from eight different companies took part in the survey. The data has been analysed to gain detailed insight into the usefulness and quality of the translated instrument with respect to the German context of software application and certification. In addition to the outcome of the software certification process itself, the results of a reliability and validity analysis of the whole instrument and the proposed subscales are reported. Cronbach’s Alpha and correlations with the ISONORM 9241/10 are calculated. Recommendations for revisions of upcoming versions of the UsersAward questionnaire are specified.

Author Keywords

UsersAward, ISO 9241-10, user certification, software quality

INTRODUCTION

Collaboration between users, software developers and scientists has a tremendous impact on the quality of the design of interactive software systems. Especially with regard to satisfaction with a software product, the participation of end users during the development and implementation of interactive systems is essential. A quality assurance process for software systems was established in Sweden by UsersAward in 1998, according

to the certification of displays and other business products by TCO (Tjänstemännens Central-Organisation).

The project was launched by the Swedish Trade Union Confederation (LO) in cooperation with the Swedish Agency for Innovation Systems (VINNOVA) and four universities [4]. Within this network, union and consumer organizations, researchers, user companies and software providers work together to develop strategies for better workplace software and to create a de-facto standard for a *user-driven* software certification process. To accomplish this task, different measures have been taken: The “IT-Kartan, Användare och IT-System i svensk näringsliv” [2] describes the status quo in the Swedish industry concerning the difficulties users have to deal with when using the IT-tools at their workplace. Corresponding “IT-Maps” for other industrial sectors of the Swedish economy are being developed. Additionally, first steps have been taken to establish a quality label for existing user-friendly software products. Since 2000, UsersAward has annually awarded an IT-prize to a software product, which has been nominated by end-users and fulfils certain quality criteria [5]. These quality criteria also build the basis for the quality-label “User-Certified”, for which software providers can put forward their software packages. Recently, some endeavours have been made to transfer this process to other European countries, such as Finland, Austria and Germany. In this paper, first results of a UsersAward pilot study in Germany are presented.

USERSAWARD’S QUALITY CRITERIA

The quality and success of a software product at the work place are determined both by the context of use within the organisation and the characteristics of the software itself.

Based on preliminary research, UsersAward devised six success factors and developed a questionnaire for measuring their fulfilment. The questionnaire quantifies users’ satisfaction with a software product on six dimensions: total benefit (4 items), deployment process (5 items), technical design (10 items), support for work tasks (6 items), support for communication and cooperation (5 items), and quality assurance (2 items). Each item is measured on a 6-point Likert scale from 1 (statement

doesn't apply at all) to 6 (statement totally applies). Additionally, a "don't know" and a "this question is not important" answer category are provided. The software fulfils the conditions to achieve the IT-prize and the quality label, respectively, if a mean value of at least 4.0 is attained for a software product on 80% of all questionnaire statements by all interviewed persons. Furthermore, a mean value of at least 4.0 has to be attained on 67% of all questionnaire statements by both men and women, and a mean value of at least 4.0 has to be attained on 67% of all questionnaire statements for each user category [5].

METHOD

The German Pilot Study

The German pilot study took place within the framework of a university class concerning the measurement of customer satisfaction with software. A German version of the Swedish UsersAward questionnaire was applied to eight software products. End users and their managers from eight different companies took part in the survey.

To guarantee the feasibility of the study in the context of a university class, a slightly different approach concerning the sample in the study was preferred. According to the Swedish UsersAward procedure 10 percent of the users of a software package in a company are asked to answer the questionnaire and three of them are interviewed with regard to the motivation for their judgments given in the questionnaire. In the German pilot study 10 users per software package and company took part in the written survey, and one of them was interviewed. Additionally, one manager per company was invited to fill in the questionnaire and to take part in the oral interview.

Participants

The participants of the study included 82 end users (36 female) and eight managers (one female) from different companies. Age groups ranged from 19 to 60 for the end users (mean = 33, std = 9.1) and 29 to 59 for the managers (mean = 39, std = 8.6). Computer experience ranged from 2 to 25 years for the end users (mean = 12, std = 5.5) and from 12 to 26 years for the managers (mean = 19, std = 5.2). The end users' experience with the examined software ranged from 2 to 216 months (mean = 54, std = 46) and managers had worked with the software under inspection between 12 and 219 months (mean = 79, std = 75).

Material

Software packages

The software examined in the pilot study included:

- a. A novel software package for preparing credit agreements used by the majority of employees of a big German financial service provider.
- b. An individual call-centre software package used in a financial customer service centre of a big German

bank.

- c. A standard software package for accounting transaction and tourist traffic management used in a small traffic agency.
- d. An information system used at a compulsory health insurance company. The software contains several modules; two of them were evaluated.
- e. A standard software package for enterprise resource planning (ERP) in small and medium-sized enterprises used by a software service provider. The module for order processing was inspected.
- f. The sales module of an ERP software package used in a small IT-warehouse.
- g. A stock data management system of a life-insurance agency.
- h. An established integrated data handling and information system of a big health insurance company.

Questionnaires

Users and managers were asked to fill in two questionnaires with respect to the software examined in their company. The first questionnaire was the German translation of the UsersAward instrument. In addition to the translation of the statements, the scale was adapted to the German school grading system ranging from 1 (here: "statement fully applies") to 6 (here: "statement doesn't apply at all"). Hence, in comparison to the Swedish questionnaire the scale polarity is reversed. However, to make the results comparable to Swedish findings, data were recoded in accordance with the Swedish scale polarity.

The second questionnaire used in the pilot study was the ISONORM 9241/10 [3], measuring the conformance with the requirements of ISO 9241 ("Ergonomic requirements for office work with visual display terminals"), Part 10 ("Dialog principles"). The latter questionnaire operationalises the seven ergonomic principles "suitability for the task", "self-descriptiveness", "controllability", "conformity with user expectations", "error tolerance", "suitability for individualization" and "suitability for learning". For each principle, five bi-polar statements are to be rated on a 7-point scale ranging from "- - -" (1) to "+++ " (7). The cut off criterion for a software package to fulfil the ISO requirements is a mean value of ≥ 5 (which refers to the "+" on the rating scale).

Procedure

After the companies and the software packages for the case studies had been selected, the questionnaires were sent to a contact person in each company, who organized the distribution of the questionnaires in house. One of ten end users who had filled out the UsersAward questionnaire was interviewed and asked for his/her motivation and reasons for his/her answers in detail. One manager from each

company took part in the study in written and oral form. The statements during the interviews were written down and stored in electronic form, the questionnaire data was transferred into a statistical software package. Data from all samples (end users and managers) are included in the analysis.

RESULTS

In this section, the overall results with respect to the quality demands for the software certification and scale characteristics are reported. Furthermore, reliability and validity measures for the German instrument are provided.

Certification Criteria

One result of the pilot study is that none of the inspected software products achieved the necessary amount of confirmed statements to fulfil the certification quality criteria, relating to the overall results of the UsersAward questionnaire for all participants (Figure 1).

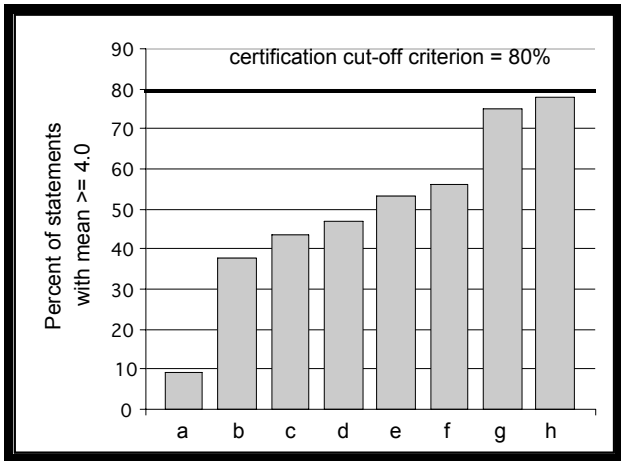


Figure 1: Certification result for all software products a ... h

Regarding the overall mean value, a t-test shows no significant differences between mean ratings of men (mean = 3.9) and women (mean = 3.7), $t = 1.49, p \geq .14, N = 87$, and between users (mean = 3.8) and managers (mean = 4.2), $t = -1.38, p \geq .17, N = 90$, respectively.

Descriptive Statistics

A descriptive analysis has been performed on the data set of the pilot study. We show means and standard deviations for the six dimensions collapsing over all 8 software packages (Table 1).

UsersAward Subscale	min	max	mean	std
Total benefit	2.28	5.62	4.43	0.99
Deployment process	2.45	4.40	3.18	0.75
Technical design	3.35	4.85	4.01	0.48

Support for work tasks	2.90	4.46	3.89	0.59
Support for communication and cooperation	3.50	4.60	4.10	0.44
Quality assurance	2.29	4.25	2.96	0.66

Table 1: Results of the UsersAward evaluation of eight software products

Overall, the software products received the worst evaluation with respect to “quality assurance” (mean = 2.96), while the best results were achieved in “total benefit” (mean = 4.43), followed by “support for communication and cooperation” (mean = 4.10) and “technical design” (mean = 4.01).

Reliability

To estimate the reliability of the German UsersAward questionnaire, Cronbach’s Alpha as an index for internal consistency was calculated for each of the proposed subscales (Table 2).

UsersAward Subscale	Cronbach’s Alpha
Total benefit	0.95
Deployment process	0.77
Technical design	0.79
Support for work tasks	0.79
Support for communication and cooperation	0.56
Quality assurance	0.31

Table 2: Cronbach’s Alpha of subscales

According to DeVellis [1], Cronbach’s Alpha – values over 0.80 are very satisfactory, over 0.70 acceptable and below 0.65 undesirable. Thus, four of the six subscales show satisfactory internal consistency.

Despite the illustrated constraints, the overall Cronbach’s Alpha of the questionnaire is remarkably high (= 0.94 for all 32 items), which suggests that the questionnaire could be shortened by removing certain items. On the contrary, internal consistency could improve by removing items with undesirable item characteristics.

Validity

For the UsersAward questionnaire, mean scale values for the six dimensions were determined in two different ways: across participants (part) and across software products (sw). Then, the correlations with the matching subscales in the ISONORM 9241/10 were calculated.

UsersAward Subscale	r_{sw}^1	r_{part}^2
Total benefit	.07	.18
Deployment process	.38	.43**
Technical design	.67*	.65***
Support for work tasks	.42	.46***
Support for communication and cooperation	.53	.33*
Quality assurance	.14	.39**

Note: * $p < .05$, ** $p < .01$, *** $p < .001$ (Spearman-Rho; one-tailed); ¹ N = 8 software products; ² N = 37 - 60 participants

Table 3: Correlations with ISONORM 9241/10

The subscale “technical design” shows the highest correlation, as could have been expected considering the scope and internal structure of the ISONORM 9241/10 questionnaire.

Similar to the results for the UsersAward questionnaire, the certification criterion of the ISONORM 9241/10 questionnaire (i.e. mean score above 5) is not achieved by any of the eight software products (min = 4.04, max = 4.68, mean = 4.33, std = 0.27). The highly significant correlation on the subscale “technical design” in the product perspective ($r = .67$, $p < .05$, N = 8) verifies the validity of this subscale.

CONCLUSION

The fact that none of the software products reached the quality criteria recommended by UsersAward could be confirmed with the results of the ISONORM 9241/10. The descriptive statistics of the subscales of the UsersAward instrument point to some inconsistencies, probably as a result of ambiguous item verbalisations (e.g. two questions in one statement) or statements, which might not be eligible in the German context of use and customisation of work place software, both because the attitude towards a satisfying software product and towards the (software supported) cooperation in every day work life itself may be different between Sweden and Germany. This might also explain the differences in internal consistency between the more technical orientated subsections on the one hand, and the more organisational orientated subsections on the other hand. Due to the undesirable low values for Cronbach’s Alpha, the items of the sections “support for communication and cooperation” and “quality assurance”

need further investigation by conducting a detailed item analysis. Items with unfavourable item characteristics should be removed or reworded. The unequal number of statements for each of the subscales has to be reconsidered. If a different weighting of the subscales is not intended, each subscale should be made up of the same number of statements, or a scale mean value across all statements of each scale has to be calculated to balance their relevance. Additionally, the polarity of the scales should be harmonized between the Swedish and the German instrument to simplify comparison of the results. As a matter of course, it has to be examined whether the intended dimensions of the questionnaire can indeed be found in the data. Hence, the dimensionality of the instrument has to be verified with a factor analysis and the existence of subscales has to be proven empirically.

All things considered, the German adaptation of the UsersAward questionnaire shows acceptable quality criteria. Some items and subscales might need reviewing, but the reliability of the overall instrument is satisfying. The validation with an external questionnaire points in the right direction, especially under the constraints described. Additional validation measures could be calculated to lend more support to the validity of dimensions other than the technical part. Whether the overall instrument (and not only the subscale on “technical design”) measures what it is intended to measure (i.e. user satisfaction with a software product) requires further research.

REFERENCES

1. DeVellis, R. F. *Scale Development: Theory and Applications*. Sage, Newbury Park, 1993.
2. Lind, T. *IT-Kartan, Användare och IT-System i svensk näringsliv*. LO, Stockholm, 2002.
3. Prümper, J.: Test IT: ISONORM 9241/10. In H.J. Bullinger and J. Ziegler (Eds.), *Human-Computer Interaction – Communication, Cooperation, and Application Design*. Lawrence Erlbaum Associates, Mahwah, New Jersey (1999), 1028-1032.
4. Sundblad, Y., Lind, T. and Rudling, J. IT Product Requirements and Certification from the Users’ Perspective. In: H. Luczak, A.E. Çakir and G. Çakir, (Eds.), *Work With Display Units WWDU 2002 - World Wide Work*. ERGONOMIC, Berlin. (2002), 280-282.
5. Walldius, Å., Sundblad, Y., Bengtsson, L., Lind, T. and Sandblad, B. User-Driven Quality Certification of Workplace Software, the UsersAward Experience. In: H. M. Khalid, M. G. Helander and A. W. Yeo (Eds.): *Work with Computing Systems*. Damai Sciences, Kuala Lumpur, (2004), 150-155.

Quelle: □

Prümper, J., Vöhringer-Kuhnt, T. & Hurtienne, J. (2005). UsersAward – first Results of a Pilot Study in Germany. In Y. Sundblad & A. Walldius (Eds.), *User-driven IT Design and Quality Assurance* (S. 15-18). Stockholm.