Published in *Conference Proceedings European Software Quality Week*, Brussels, November 1998

Questionnaire based usability testing

Drs. Erik P.W.M. van Veenendaal CISA

Abstract

Usability is an important aspect of software products. However, in practice not much attention is given to this issue during testing. Testers often do not have the knowledge, instruments and/or time available to handle usability. This paper introduces the Software Usability Measurement Inventory (SUMI) testing technique as a possible solution to these problems. SUMI is a rigorously tested and validated method to measure software quality from a user perspective. Using SUMI the usability of a software product or prototype can be evaluated in a consistent and objective manner. The technique is supported by an extensive reference database and embedded in an effective analysis and reporting tool.

SUMI has been applied in practice in a great number of projects. This paper discusses three practical applications. The results, usability improvements, cost and benefits are described. Conclusions are drawn regarding the applicability and the limitations of SUMI for usability testing.

1. A closer look at usability

Several studies have shown that in addition to functionality and reliability, usability is a very important success factor [10] But although it is sometimes possible to test the software extensively in a usability lab environment, in most situations a usability test has to be carried out with minimum resources.

The usability of a product can be tested from mainly two different perspectives "ease-of-use" and "quality-in-use". Quite often the scope is limited to the first perspective. The ease or comfort during usage is mainly determined by characteristics of the software product itself, such as the user-interface. Within this type of scope usability is part of product quality characteristics. The usability definition of ISO 9126 is an example of this type of perspective:

Usability the capability of the software to be understood, learned, used and liked by the user, when used under specified condition [3]

Two techniques that can be carried out at reasonable costs evaluating the usability product quality, are expert reviews and checklists. However, these techniques have the disadvantage that the real stakeholder, e.g. the user, isn't involved. In a broader scope usability is being determined by using the product in its (operational) environment. The type of users, the tasks to be carried out, physical and social aspects that can be related to the usage of the software products are taken into account. Usability is being defined as "quality-in-use". The usability definition of ISO 9241 is an example of this type of perspective:

Usability the extent to which a product can be used by specified users to achieve goals with effectiveness, efficiency and satisfaction in a specified context of use [6]

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

Clearly these two perspective of usability are not independent. Achieving “quality-in-use” is dependent on meeting criteria for product quality. The interrelationship is shown in figure 1.

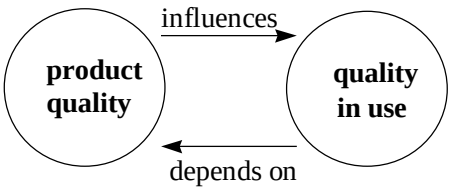

Figure 1 : Relationship between different types of usability

Establishing test scenarios, for instance based on use cases [7], can be applied to test usability in accordance with ISO 9241. However, usability testing with specified test cases / scenarios is a big step for most organization and often not even necessary. From a situation where usability is not tested at all one wants a technique that involves users, is reliable but still requires limited resources.

Within the European ESPRIT project MUSiC [ESPRIT 5429] a method has been developed that serves to determine the quality of a software product from a user’ perspective. Software Usability Measurement Inventory (SUMI) is a questionnaire based method that can been designed for cost effective usage.

2. What is SUMI?

Software Usability Measurement Inventory (SUMI) is a solution to the recurring problem of measuring users' perception of the usability of software. It provides a valid and reliable method for the comparison of (competing) products and differing versions of the same product, as well as providing diagnostic information for future developments. It consists of a 50-item questionnaire devised in accordance with psychometric practice. Each of the questions is answered with "agree", "undecided" or "disagree". The following sample shows the kind of questions that are asked:

- This software responds too slowly to inputs
- I would recommend this software to my colleagues
- The instructions and prompts are helpful
- I sometimes wonder if I am using the right command
- Working with this software is satisfactory
- The way that system information is presented is clear and understandable
- I think this software is consistent.

The SUMI questionnaire is available in English (UK and US), French, German, Dutch, Spanish, Italian, Greek and Swedish.

SUMI is intended to be administered to a sample of users who have had some experience of using the software to be evaluated. In order to use SUMI effectively a minimum of ten users is recommended. Based on the answers given and statistical concepts the usability scores are being calculated. Of course SUMI needs a working version of the software before SUMI can be measured. This working version can also be a prototype or a test release.

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

One of the most important aspects of SUMI has been the development of the standardization database, which now consists of usability profiles of over 2000 different kinds of applications. Basically any kind of application can be evaluated using SUMI as long as it has user input through keyboard or pointing device, display on screen, and some input and output between secondary memory and peripheral devices. When evaluating a product or series of products using SUMI, one may either do a product-against-product comparison, or compare each product against the standardization database, to see how the product that is being rated compares against an average state-of-the-market profile.

SUMI gives a global usability figure and then readings on five subscales:

- *Efficiency*: degree to which the user can achieve the goals of his interaction with the product in a direct and timely manner
- *Affect*: how much the product captures the user's emotional responses
- *Helpfulness*: extent to which the product seems to assist the user
- *Control*: degree to which the user feels he, and not the product, is setting the pace
- *Learnability*: ease with which a user can get started and learn new features of the product.

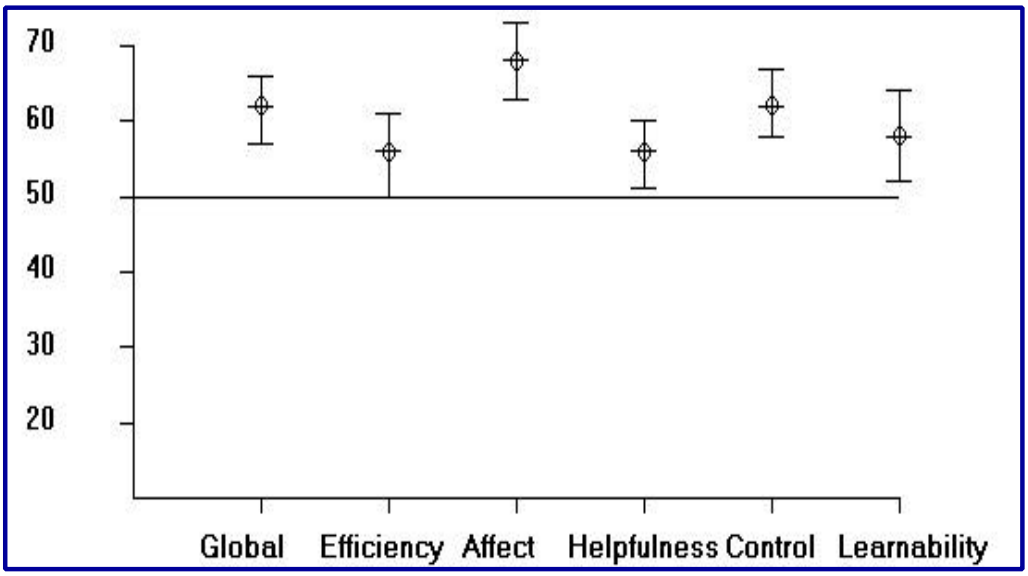

Figure 2: a sample profile showing SUMI scales

Figure 2 shows an example of SUMI output; it shows the scores of a test and the spreading of these scores (measured by the standard deviation) against the average score of the reference database, reflected by the value 50. Consequently the usability scores shown in the sample profile are positive, e.g. more than state-of-the-art, with a reasonable level of spreading.

SUMI is the only available questionnaire for the assessment of usability of software, which has been developed, validated and standardized on a European wide basis. The SUMI subscales are being referenced in international ISO standards on usability [5] and software product quality [4]. Product evaluation with SUMI provides a clear and objective measurement of users' view of the suitability of software for their tasks.

This provides a solid basis for specialized versions of SUMI. Recently MUMMS has been developed for MultiMedia products (Measuring Usability of Multi Media Systems).

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

Any SUMI test must be carried out by asking people that perform realistic, representative tasks. Employing a method such as usability context analysis [11] helps identify and specify in a systematic way the characteristics of the users, the tasks they will carry out, and the circumstances of use. Based on the results the various user groups can be described and used to define how these user groups can be represented in the test.

3. Practical Applications

3.1 Project 1: Project Management Package

Approach
Subject to the usability evaluation by means of SUMI was a software package offering project administration and control functionality. The software package is positioned as a multi-project system for controlling the project time, e.g. in terms of scheduling and tracking, and managing the productivity of projects, e.g. in terms of effort and deliverables. The package has been developed by a Dutch software house that specializes in the development of standard software packages.

The SUMI test was part of an acceptance test carried out on behalf of a potential customer. Due to the very high number of users, a number of different user groups, their inexperience with project management software and the great variety of information needs, usability was an important characteristic. It was even looked upon as the critical success factor during implementation. Two main user group were distinguished. One user group was mainly involved in input processing of effort and time spent. For this user group especially operability and efficiency is of great importance. Another user group was characterized as output users. Especially receiving the right management information is important for the output users. Per user group a SUMI test has been carried out.

Regarding the usage of the SUMI technique for the usability evaluation a specific acceptance criteria was applied. SUMI provides quantitative values relating to a number of characteristics that lead to a better understanding of usability. As part of the acceptance test, the SUMI scale was used that provides an overall judgement of usability, the so-called “global scale”. Based on the data in the SUMI database, it can be stated that the global score has an average value of 50 in a normal distribution. This means that by definition for a value exceeding 50 the user satisfaction is higher than average. In the test of the project management package the acceptance criteria applied that for each user group the global scale and the lower limit of the 95% confidence interval must both exceed the value of 50.

Results
The "global scale" regarding both user groups was below the desired 50. For the input user group the score was even a mere 33. The output user group showed a slightly better score. Not only the “global scale” but also most other subscales were scoring below 50.

Because the results did not meet the acceptance criteria that were set a number of usability improvement measures needed to be taken. Examples of measures that were taken based on the results of the SUMI test are:

- extension and adaptation of the user training
- optimization of efficiency for important input functions

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

- implementation of specific report generation tools for the output user with a clear and understandable user-interface.

3.2 Project 2: PDM system

Approach
At the R&D department of a large copier manufacturer a Product Data Management System (PDMS) is implemented. During the trial phase usability appeared to be an issue and could become a major risk factor during implementation. The time and effort needed to be spent on usability formed a point of discussion between development and the user organization. It was decided to apply SUMI to acquire an insight into the current user perception of the PDMS.

A number of randomly selected users that were involved in the PDMS trail phase were requested to fill out the questionnaire. Twenty six users were selected in this way, of whom twenty-one returned the questionnaire. Six users stated that they didn't use the PDMS often enough. The feedback thus resulted in a 77% response.

Results
The table below shows the overall scores for the various SUMI subscales:

	Global	Efficiency	Affect	Helpfulness	Control	Learnability
Median	36	31	43	36	36	35

Table 1: SUMI scores PDMS

The various scores are relatively low all round. There didn't seem to be a too large divergence of opinion, except perhaps for learnability. An analysis of the individual user scores did not show any real outlayer (see next table). Two users (one and five) had an outlayer score for one scale (too high). Since it was only on one scale, they were not deleted from the respondent database.

	G	E	A	H	C	L
User 1	60	52	59	**69**	47	32
User 2	57	48	53	62	41	61
User 3	25	19	46	35	22	33
User 4	17	14	28	11	26	23
User 5	61	63	55	44	**60**	64
User 6	24	23	23	36	22	14
User 7	53	62	44	..	..	..
User ..	..	..	..	..	..	..

Table 2: SUMI scores per user

As stated earlier the various scores were relatively low all round. In general one can say that the user satisfaction regarding the system is too low and corrective action is needed. Some more detailed conclusion were:

- *Efficiency*
  According to the users PDMS doesn't support the user tasks in an efficient way. One has to carry out too many and too difficult steps. As a consequence one cannot work

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

efficiently and has the opinion that the system is insufficiently customized to their needs.

- *Helpfulness*
  An important conclusion is the fact that the messages are often not clear and understandable; as a consequence the system doesn't provide much help when one has to solve a problem. The possibilities that the user has in each situation are not clearly shown.
- *Control*
  The user often have the feeling that they are not in control and find it difficult to let the system behave in the way they want it to. They feel save when they only use commands they know. However, they do find it easy to jump from one task to another.

On the basis of the SUMI evaluation it was decided to define a number of follow-up actions:

- a detailed analysis of the problems as being perceived by the users. A number of users is interviewed and asked to explain, by means of practical examples, the answers given to the SUMI questions;
- a study on outstanding change requests and probably increase their priority;
- an improved information service to the users on changed functionality to provide them with more knowledge on how the system operates;
- a re-evaluation of the training material with user representatives;
- a SUMI test was to be carried out on a regular basis (every two/three months) to track the user satisfaction during implementation of the PDMS.

Currently the follow-up is in progress and no new SUMI test has yet taken place. As a consequence nothing can be said regarding the improvement of the usability. However, by means of the SUMI test usability has become a topic within the PDMS project that gets the attention (time and effort) it apparently needs.

3.3 Project 3: Intranet site

Approach
By means of MUMMS, the specialized multimedia version of SUMI, the usability of an intranet site prototype of a large bank was evaluated. The intranet site was set up by the test services department to get well-known and to present themselves to potential customers. The fact that during the test only a prototype version of the intranet site was available meant that some pages were not yet accessible. For MUMMS a special subscale has been introduced, with the objective to measure the users' multimedia "feeling":

- *Excitement*: extent to which end-users feel that they are "drawn into" the world of the multimedia application.

In total ten users (testers) were involved in the MUMMS evaluation. The set of users can be characterized by:

- not having been involved during the development of the intranet site
- potential customers
- four users with internet experience
- six users without internet experience
- varying by age and background (job title).

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

Results
The table below shows the overall scores for the various MUMMS subscales:

	Affect	Control	Efficiency	Helpfulness	Learnability	Excitement
average score	69	74	62	67	67	68
median	71	77	67	69	67	72
standard deviation	9	12	11	8	6	12

Table 3: Overall MUMMS score table

The various scores were moderately high all round. However, there seems to be a divergence of opinion on the control and excitement scales. Some low scores are pulling down the control and efficiency scales (see next table). Two users from the sample were giving exceptionally low average scores. They were analyzed in detail but no explanation was found.

	A	C	E	H	L	E	Average
User 1	71	81	67	71	74	77	73
User 2	74	74	74	71	67	71	72
User 3	81	84	67	67	74	74	74
User 4	54	51	54	57	64	44	**54**
User 5	71	74	43	58	55	76	63
User 6	64	84	67	81	67	69	72
User 7	51	81	74	54	74	64	66
User 8	71	81	64	74	71	81	73
User 9	77	81	76	84	77	74	78
User 10	64	47	51	57	57	44	**53**

Table 4: MUMMS scores per user

As stated the usability of the Intranet site was rated moderately high from the users' perspective, although there seemed to be a lot of divergence in the various user opinions. Some more detailed conclusion were:

- *Attractiveness*
  The attractiveness score is high (almost 70%). However some users (4, 7 and 10) have a relatively low score. Especially the questions "this MM system is entertaining and fun to use" and "using this MM system is exiting" are answered in different ways. It seems some additional MM features should be added to further improve the attractiveness for all users.
- *Control*
  A very high score for control in general. Again two users can be identified as outlayers (4 and 10) scoring only around 50%, the other scores are around 80%. Problems, if any, in this area could be traced back to the structure of the site.
- *Efficiency*
  The average score on efficiency is the lowest, although still above average. Users need a more time than expected to carry out their task, e.g. find the right information.

On the basis of the MUMMS evaluation it was decided to improve the structure of the internet site and to add a number of features before releasing the site to the users. Currently the update

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

of the intranet site is being carried out. A MUMMS re-evaluation has been planned to quantify the impact of the improvement regarding usability.

4. Applicability of SUMI

On the basis of the test carried out in practice, a number of conclusions have been drawn regarding the applicability of SUMI and MUMMS:

- it is easy to use; not many costs are involved. This applies both to the evaluator and the customer. On average a SUMI test can be carried in approximately 3 days; this includes the time necessary for a limited context analysis and reporting;
- during testing the emphasis is on finding defects, this often results in a negative quality indications. SUMI however, provides an objective opinion;
- the usability score is split into various aspects, making a thorough more detailed evaluation possible (using the various output data);
- MUMMS provides, after detailed analysis and discussion, directions for improvement and directions for further investigation. SUMI can also be used to determine whether a more detailed usability test, e.g. laboratory test, is necessary.

However, also some disadvantages can be distinguished:

- a running version of the system needs to be available; this implies SUMI can only be carried at a relatively late stage of the project;
- the high (minimum of ten) number of users with the same background, that need to fill out the questionnaire. Quite often the implementation or test doesn't involve ten or more users belonging to the same user group;
- the accuracy and level of detail of the findings is limited (this can partly be solved by adding a small number of open question to the SUMI questionnaire).

5. Conclusions

It has been said that a system's end users are *the* experts in using the system to achieve goals and that their voices should be listened to when that system is being evaluated. SUMI does precisely that: it allows quantification of the end users' experience with the software and it encourages the tester to focus in on issues that the end users have difficulty with. Evaluation by experts is also important, but it inevitably considers the system as a collection of software entities.

A questionnaire such as SUMI represents the end result of a lot of effort. The tester get the result of this effort instantly when SUMI is used: the high validity and reliability rates reported for SUMI are due to a large measure to the rigorous and systematic approach adopted in constructing the questionnaire and to the emphasis on industry-based testing during development. However, as with all tools, it is possible to use SUMI both well and badly. Care taken over establishing the context of use, characterizing the end user population, and understanding the tasks for which the system will be used supports sensitive testing and yields valid and useful results in the end.

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl

References

[1] Bevan, N. (1997), "Quality and usability: a new framework", in: E. van Veenendaal and J. McMullan (eds.), "Achieving Software Product Quality", Tutein Nolthenius, 's Hertogenbosch, The Netherlands

[2] Bos, R. and E.P.W.M. van Veenendaal (1998), "For quality of Multimedia systems: The MultiSpace approach" (in Dutch), in: Information Management, May 1998

[3] ISO/IEC FCD 9126-1 (1998), "Information technology - Software product quality - Part 1 : Quality model", International Organization of Standardization

[4] ISO/IEC PDTR 9126-2 (1997), "Information technology - Software quality characteristics and metrics - Part 2 : External metrics", International Organization of Standardization

[5] ISO 9421-10 (1994), "Ergonomic Requirements for office work with visual display terminals (VDT's) - Part 10 : Dialogue principles", International Organization of Standardization

[6] ISO 9241-11 (1995), "Ergonomic Requirements for office work with visual display terminals (VDT's) - Part 11 : Guidance on usability", International Organization of Standardization

[7] Jacobson, I. (1992), "Object Oriented Software Engineering; A Use Case Driven Approach", Addison Wesley, ISBN 0-201-54435-0

[8] Kirakowski, J., "The Software Usability Measurement Inventory: Background and Usage", in: Usability Evaluation in Industry, Taylor and Francis

[9] Kirakowski, J. and M. Corbett (1993), "SUMI: the Software Usability Measurement Inventory", in: British Journal of Educational Technology, Vol. 24 No. 3 1993

[10] Moolenaar, K.S. and E.P.W.M. van Veenendaal (1997), "Report on demand oriented survey", MultiSpace project [ESPRIT 23066]

[11] National Physical Labotory (NPL) (1995), "Usability Context Analysis: A Practical Guide", version 4.0, NPL Usability Services, UK

[12] Nielsen J. , (1993) "Usability Engineering", Academic Press, ISBN 0-12-518406-9

[13] Preece, J. et al, "Human-Computer Interaction", Addison-Wesley Publishing company

[14] Tienekens, J.J.M. and E.P.W.M. van Veenendaal (1997), "Software Quality from a Business Perspective", Kluwer Bedrijfsinformatie, Deventer, The Netherlands

The Author

Drs. Erik P.W.M. van Veenendaal CISA has been working as a practitioner and manager within the area of software quality for a great number of years. Within this area he specializes in testing and is the author of several books, e.g. "Testing according to TMap" (in Dutch) and "Software Quality from a Business Perspective". He is a regular speaker both at national and international testing conferences and a leading international trainer in the field of software testing. Erik van Veenendaal is the founder and managing director of Improve Quality Services. Improve Quality Services provides services in the area of quality management, usability and testing.

At the Eindhoven University of Technology, Faculty of Technology Management, he is part-time involved in lecturing and research on information management, software quality and test management. He is on the Dutch standards institute committee for software quality.

Improve Quality Services, Waalreseweg 17, 5554 HA Valkenswaard, The Netherlands
Tel 040 – 2089283 • Fax 040 – 2021450 • E-mail info@improveqs.nl