*Journal of the Ergonomics Society of Korea*
Vol. 14, No.2. 1995


# Criteria-based Quantitative Analysis of Product Usability

**Park, Kyung S* and Lim, Chee Hwan***

## ABSTRACT

An approach to quantitative evaluation of product usability during the design and implementation phases of the product life cycle is presented. The evaluation starts by defining a set of usability characteristics, viz. physical, cognitive and affective characteristic. Each of these characteristics is assessed by a set of usability criteria, such as functionality, visual clarity, learmability, etc. The usability criterion is then evaluated by a checklist consisting of a detailed questionnaire. The fuzzy weighted-checklist method with linguistic variables is used for quantitative analysis. Also, the method for quantifying usability improvements in iterative design processes is considered. This analysis provides a quantitative measure, which reflects the degree of excellence of product usability during the design and development phases.

## 1. Introduction

Usability has become more important with the increase in number of computer users and applications, and it has become a primary factor in determining the accept-ability and consequent success of co mputer products[6,7]. Users are becoming less willing to put up with defficult or uncomfortble interfaces of computer produ cts since experience with some current interfaces has shown them that a product can indeed be easy to learn and pleasant to use.

This phenomenon results from poor interface design. As a result, low usability may incur all kinds of costs : costs for the user in extra training and slower, lower quality work ; and costs for the producer in increased demands on software support, and losses of future sales[4]. High usability is thus desirable, but it has remained a fuzzy concept that has been

* Department of Industrial Engineering, KAIST, Taejon, Korea

difficult to evaluate and measure.

The ideal way to evaluate and measure usability would be to specify the attributes required to make a product usable, and evaluate whether they are present in the implemented product. The problem with usability is that it is very difficult to specify what the attributes should be. There have been many attempts to describe them, including design principles and guidelines[9, 17, 18, 19]. However, most design principles and guidelines remain very general, without providing methods.

As the pace of technological innovation quickens, and the design of user interfaces involves more complex interaction techn iques, the need for simple and efficient evaluation methods is increasing. In this paper we sought to provide the developer or designer with a method that could quantitatively analyze informal and subjec tive data frequently acquired during the design and development phases. Moreover, we have not considered only attribute of product usability, but an overall view of the usability attributes during a particular phase of product life cycle.

The objective of this paper is to present the following aspects of usability engineer ing to aid in the development of new procedures and measures for product usability : (1) quantification of product usabi lity during the design and implementation phases ; (2) a quantitative analysis, by the assignment of characteristics and criteria that specify a figure of merit or degree of excellence for product usability.

## 2. Usability characteristics, criteria and Checklists

To develop a procedure for the usability assessment of a product, various attributes of the usability must be defined. However, because the usability attributes are usually unclear and fuzzy, there is a need to obtain a clear understanding of the attributes. The present paper provides a concrete framework of usability attributes. This framework consists of a hierarchy of four levels. The highest level is usability characteristics, the next lower level is usability criteria, and the lowest levels are usability checklists and subchecklists.

The important usability characteristics, the highest level, are classified into three parts of user interface : physical, cognitive and affective characteristic[11]. Physical characteristic is a necessary condition for usability that includes such physiological and formal aspects as 'see' and 'grasp'. Cognitive characteristic includes such cognitive features as 'know' and 'easy to learn'. Also, affective characteristic includes such human feeling as 'comfortable' and 'pleasant'. The three usability characteristics are evaluated by using the usability criteria.

A usability criterion(or the second level) is measure of the extent to which a certain characteristic is exhibited or possessed by a product. That is to say, a criterion is a dimension recognized as one that will lead to a better, more efficient, less error-prone behavior of the user. To establish a set of major criteria for 'usable' products, we reviewed a number of the literature in this area, and selected several representatives of them.

High level principles for user interface design are contained in ISO 9241-10 "Dialogue principles"[2,3]. These principles that should be applied to the design of dialogues between humans and information systems are listed in Table 1. For example, the self-descriptiveness could be thought of as being composed of aspects concerned with self-documenting, transparency, and decision supporting.

Table 1. Comparision of usability criteria of and interactive product

| ISO 9241 Part-10 Dialogue Principles | User-perceived Quality of Interactive systems | SUMI | Ravden & Johnson's driteria |
|---|---|---|---|
| | | | *visual clarity** |
| suitability for the task | | | *appropriate functionality |
| *controllability | user control | control | control |
| suitability for individualization | flexibility in task handing | | *flexibility |
| *error tolerance | fault tolerance | | error prevention and correction |
| suitability for learning | ease of learning | *learnability | |
| conformity with user expectations | *correspond -ence with user expectations | | compatibility |
| *self-descriptiveness | self-descriptiveness | helpfulness | information feedback user guidecne & support |
| | | | *consistency** |
| | problem adequate usability** | efficiency** *effect** | *explicitness** |

* A selected criterion for our study
**A criterion that has no common ground

Dzida et al.(1978) presented an analytecal approach for identifying essential properties of interactive systems as perceived by their users. In addition, they reported that user-perceived quality was defined as a multedimensional property of interactive systems comprising at least seven aspects.

To measure user satisfaction, and heace assess user perceived software quality, Kirakowski et al.(1992) developed the Software Usability Measurement Inventory (SUMI) as part of the Metrics for Usability Standards in Computing(MUSiC) project. SUMI decomposes usability into five dfferent components : efficiency, effect, helpfulness, control and learnability. SUMI can benefit software designers and people involved in the purchasing of software, by indicating how easily the software can be used. Ravden and Johnson(1989) presented a set of software ergonomics criteria, which a well-designed user interface should aim to meet. The nine criteria are listed in Table 1.

As we compared the summarized usability criteria with each other, we found several common grounds of usability criteria. Thus, the criteria with common ground were selected for our usability criteria(or the second level). Among the criteria without common ground, there were important criteria for product usability. For example, the visual clarity could be a more important criterion as the graphic user interface and multimedia technology are introduced. Moreover, the consistency is one of the most important criteria that are considered in the design of user inter face[8]. Thus, these criteria were added to our usability criteria.

The affective characteristic refers to the user's feeling, good, warm, happy or the opposite as a result of interacting with the product. The criteria for the affective characteristic were derived from the adjective items in Image Technology area[13]. The selected criteria for the affective characteristic were satisfaction, reliability, safety, originality and pleasantness. As a result of the review and the comparison, we established a set of usability criteria in Table 2.

A usability checklist(or the third level) is a questionnaire that contains a set of questions. A set of exhaustive checklists is considered for each of the criteria. Each question of a checklist measures one particular aspect to be relfected by a criterion. The usability checklists have been developed with reference to a number of sources. The sources for the checlist include, among others, Smith and Mosier (1986), Ravden and Johnson(1989) and Shneiderman(1992).

Table 2. A set of usability criteria

| | |
|---|---|
| appropriate functionality | The system should meet the needs and requirements of users when carrying out tasks. |
| visual clarity | information displayed on the screen should be clear, well-organized, unambiguous and easy to read. |
| controllabilty | The interface should be sufficiently controllable in structure, in the way information is pressented and in terms of what the user can do, to suit the needs and requirements of all users. |
| flexibility | The interface should be sufficiently flexible in structure, in the way information is presented and in terms of what the user can do, to suit the needs and requirements of all users. |
| error tolerance | The system should be designed to minimize the possibility of user errors, with inbuilt facilities for detecting and handling those which do occur. |
| learnablity | The way the system works should be easy to learn. |
| correspondence with user expectations | The way the system looks and works should be compatible with user conventions and expectations. |
| self-descriptiveness | Information and relevant feedback, guidance and support should be provided, both on the computer and in hard-copy document form, to help the user understand and use the system. |
| consistency | The way the system looks and works should be consistent at all tims. |
| explicitness | The way the system works and is structured should be clear to the user. |
| satisfaction | The way the system looks and works should be satisfying to the user. |
| reliability | The way the system looks and works should be reliable to the user. |
| safety | The way the system looks and works should be safe to the user. |
| originality | The way the system lood and works should be creative to the user. |
| pleasantness | The way the system looks and words should be plesant to the user. |

Whenever a checklist question needs to be considered by assessing a number of points, a subchecklist(or the foruth level) is prepared, which consists of a number of questions. These checklists are mainly meant for anomaly detection.

The characteristics and criteria discussed earlier are aaplicable to any computer prod uct. Thus, the methodology for quantifying usability in this paper can be applied to any product. The checklists considered here are applicable to the graphical user interface, which may include some forms of the windows, icons, menus, and a pointing device interface style (WIMP).

## 3. Quantitative anaysis to measure product usability

Ravden and Johnson(1989) regarded the quantitative analysis of product usability as an unrealistic goal, because their method aimed at identifying the strengths and weakness of the product usability. How ever, quantifiable outcomes in the usability analysis could help the decision making during the design and development phases. For example, the outcomes in the usability analysis may be defined in absolute terms, using pre-specified levels that have been judged to be acceptable for the design, or they may be expressed in relative terms, involving comparisons between alternative design options, or alternative available products.

One of the prevalent analysis techniques for evaluating or comparing complex, un structured situations(or systems) is the Analytic Hierarchy Process(AHP)[12]. The AHP as a methodology for prioritizing alter-natives associated with a system could help the decision making or the policy making. However, the AHP may have limitations in that it is difficult to quantify the quality of an alternative or a system in absolute term. To perform quantitative analysis of usability in this area, therefore, it requires a methodology for quantifying the usability of a product in absolute terms as well as in relative terms.

### 3.1 The method for quantitative anaylsis

The method used for quantitative analysis is the fuzzy-weighed checklist method proposed by Park and Kim(1990). This method is a fuzzy version of the weighted checklist technique used for evaluating or comparing compley system cor subjects). The procedure of this method consists of converting the ratios of weights obtained by pairwise comparison to a best combination of normalized weights and obtaining the fuzzy composite score and its linguistic approximation. The method can be of service to design engineers and system analysts for system evaluation with qualitative criteria.

To apply this method to the present application, a hierarchy of four levels was defined as shown in Fig. 1. Each hierarchi cal level is assigned a weight relative to its importance. The highest level is broken down into lower levels of 100% weight. These levels are again split into lower levels of 100% weight and so on, until the lowest level is reached. The highest level in this application is usability characteristics, the next lower level is usability criteria, and the lowest two levels are usability checklists and subchecklists. (Fig. 1.)

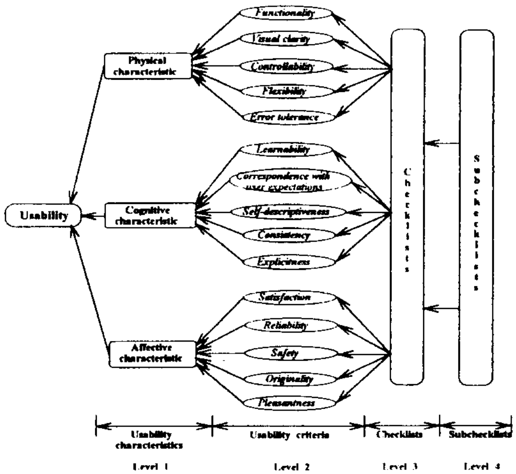


Figure 1. Usability characteristics and criteria tree

To perform reliable usability evaluation of a product(or an interactive system), several people are usually involved. When multiple evaluators are involved, it is necessary that their judgments are synthesized into a single judgment. Aczel and Saaty(1983) proved that synthesizing the individual ratios of weights was effectively performed by the geometric mean. Thus, the geometric means of the individual ratios of weights are converted to a combination of normalized weights.

The system designers or users evaluate the lowest two levels, i.e checklists and sub checklists. Figure. 2 shows the example of the usability checklists and subchecklists. The evaluators grade a product on a rating scale.

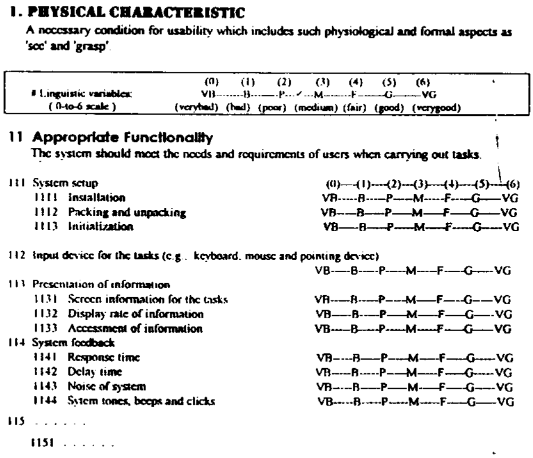


Figure 2. An example of usability checklist

## 3.2 Rating scale

The scale used to rate the checklists is a 7-point scale(0 to 6) using 14 fuzzy sets of linguistic values(verygood, good, fair, medium, poor, bad, verybad, unknown). For fuzzier rating, hyphenated ratings such as fair-good can be used. There are the templates of fuzzy sets used for linguistic rating in Table 3. Rating of the checklists is the most important part of the analysis. An individual has to evaluate the question and assign a linguistic value to it. (Table 3.)

Table 3 Templates of fuzzy sets used for lingustic rating

| linguistic value | membership function | shape | core |
|---|---|---|---|
| verygood (VG) | [5,6,6,6] | triangle | 6 |
| good (G) | [4,5,5,6] | 〃 | 5 |
| fair (F) | [3,4,4,5] | 〃 | 4 |
| medium (M) | [2,3,3,4] | 〃 | 3 |
| poor (P) | [1,2,2,3] | 〃 | 2 |
| bad (B) | [0,1,1,2] | 〃 | 1 |
| verybad (VB) | [0,0,0,1] | 〃 | 0 |
| good-verygood (G-VG) | [4,5,6,6] | trapezoid | 5-6 |
| fair-good (F-G) | [3,4,5,6] | 〃 | 4-5 |
| medium-fair (M-F) | [2,3,4,5] | 〃 | 3-4 |
| poor-medium (P-M) | [1,2,3,4] | 〃 | 2-3 |
| bad-poor (B-P) | [0,1,2,3] | 〃 | 1-2 |
| verybad-bad (VB-B) | [0,0,1,2] | 〃 | 0-1 |
| "unknown" (U) | [0,0,6,6] | rectangle | 0-6 |

### 3.3 Evaluation Procedure

The evaluation starts from the lowest level, i.e., checklists or subchecklists. That level is evaluated and the linguistic value for rating is assigned. The assigned values are multiplied by the normalized weights, and then the composite score is obtained by summing these results. The fuzzy composite score at that level is carried forward to the next higher level. Again, computing the weighted average is performed to obtain the composite score for that level, and this process continues until the highest level is reached.

Because the composite score from the weighted average is a fuzzy set, the median of composite score is used for transforming such a fuzzy set into a representative numerical score[15]. The final composite score(and its linguistic approximation) gives the figure of merit or degree of excellence of product usability during the design and development phases.

### 3.4 An example

For demonstration, a personal database integration program, Korea™(shareware version), was evaluated. This system, a program running on an IBM-PC, processes personal data, including personal informa tion, personnel administration, and adjust ment of accounts. The intended user population of this system was the general public.

Seven graduate students(ages 22 to 33) participated in this study and were unpaid volunteers. All participants had at least five year' experience with a computer system and had programming training or experi ence. They read a description of the purpose of the study and completed a background questionnaire. After a short tutorial on this system, they began working on the given tasks, such as generation of personal information, browsing and search of stored data, a change of operating options, etc. After doing this they completed the checklists and pairwise comparison on the components of each hierarchical level.

The cirteria-based quantitative analysis, then, was performed by the above-metioned evaluation procedure. The information gath ered in this evaluation was analyzed by the quantitative analysis program written in Visual Basic for Windows 3.0™. The main screen of this analysis program is presented in Figure 3. and the analysis output of this system usability is presented in Figure 4. The analysis output shows the normalized weights and the composite scores(or the linguistic values) of each hierarchical level.

The criteria-based quantitative analysis, then, was performed by the above-mentioned evaluation procedure. The information gathered in this evaluation was analyzed by the quantitative analysis program written in Visual Basic for Windows 3.0™. The main screen of this analysis program is presented in Figure 3, and the analysis output of this system usability is presented in Figure 4. The analysis output shows the normalized weights and the composite scores(or the linguistic values) of each hierarchical level.

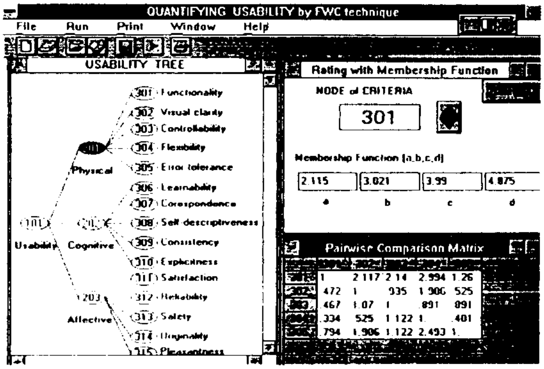


Figure 3. The main screen of the quantitative analysis program

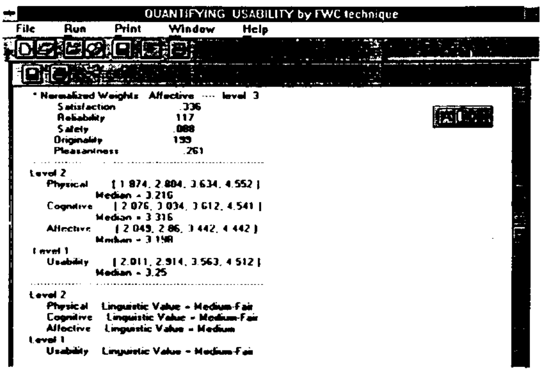


Figure 4. The output of the quantitative analysis program

## 4. Quantifying usability improvements in iterative design

A uniform scoring of usability improvements could be used to quantify relative changes of usability in iterative design processes. The approach normalizes all usability criteria's values(or evaluation scores) concerning the values measured for the initial design. Thus, the initial design has a normalized usability value of 100, and the approach normalizes the subsequent iterations by dividing their measured values by those measured for the initial design.

However, converting evaluation scores of product usability into improvement scores is more challenging, because the approach used in this paper is measured with questionnaires and rating scales. The underlying problem is that rating scales are not ratio scales. It is theoretically impossible to arrive at a perfect fomula to calculate how much better one rating is relative to another.

To solve this problem, thus, we applied the method proposed by Nielsen(1993) to transforming subjective ratings to a ratio scale. The transformation of the raw rating scale scores in each design process is as follows :

1. Rescale the various rating scales linearly to map onto the interal from-1 to +1, with 0 as the midpoint and +1 as the best rating.

2. Apply the arcsine function to the rescaled scores.

3. Apply the exponential function $e^X$ to the stretched scores of step 2 to achieve numbers for which ratios are meaningful.

The purpose of the transformation is to compare relative ratings of the various interface iterations, and the same transformations were applied to each original rating We always look at the ratio between two equally transformed ratings, and all the transformations are monotonic.

The method for quantifying usability improvements in iterative design is per formed by calculating overall usability in relative terms as the geometric mean of the normalized values of the relative improve ments in the individual usability criteria. The geometric mean increases more when all the usability criteria improve a little than when a single criterion improves a lot and the others are stagnant.

Table 4. An example of transforming rating score and calculating improvement score

| cognitive characteristic | composite score in version 1 (vl) | | composite score in version 2 (v2) | | improvement score |
|---|---|---|---|---|---|
| | median | transformed value | median | transformed value | V2/VI (%) |
| *learnability* | 3.6 | 1.2231 | 4.2 | 1.5091 | 123 |
| *correspondence with user expectations* | 4.7 | 1.8266 | 5.1 | 2.1715 | 119 |
| *self-descriptiveness* | 4.5 | 1.6881 | 4.6 | 1.7551 | 104 |
| *vomsistency* | 4.2 | 1.5091 | 5.3 | 2.3956 | 159 |
| *explicitness* | 3.8 | 1.3099 | 4.5 | 1.6881 | 129 |

*geometric mean : 125.6(%)*

For example, if the median of the fuzzy composite score of learnability criterion is 3.6 while completing a task with version 1 and 4.2 with version 2 of an interface, the normalized usability of version 2 with respect to learnability is 123 percent, indicating a 23 percent improvement in usability. The relative change of cognitive characteristic in an example of Table 4 is 125.6 percent. This means 25.6 percent improvement in cognitive characteristic of usability.

## 5. Conclusions

The present paper was to attempt a numerical summary of subjective data, through the usability characteristics, criteria and checklists. The methodology considered for usability characteristics and criteria is viewed as a step toward a more rational approach to product usability assurance. A set of criteria presented in this study provides a means for quantitatively specifying the level of usability possessed by an interactive product.

The application of a fuzzy set theory approach for criteria-based analysis of usability is of benefit to quantifying and manipulating qualitative statements, subjec tivity of opinion, vagueness of usability concept. The fuzzy weighted checklist method provides useful empirical results when the set of subjects from which to sample is small and when traditional statistical procedures are perhaps inappro priate or offer insufficient power.

The quantitative analysis methodology described here lends itself to sensitivity analysis. The figure of merit can be used to compare components within the product and to evaluated the overall product usability. A component or group of components that shows low scores should be looked at in more detail during the evaluation. A component that has a low score can be further judged by looking at individual scores of characteristics and criteria. This can help to localize the low-usability areas and improve upon them before release. In summary, the quantifiable outcomes that allow some comparison between interfaces of a product or signify the relative improvements in iterative design processes would help the decision making during the design and development phases.

## REFERENCES

[1] Aczel, J., and Saaty, T.L., "Procedures for synthesizing Ratio Judgments", Jour nal of Mathematical Psychology, 27, p. 93-112, 1983.

[2] Bevan, N., and Macleod, M., "Usability specification and measurement", Nation al Physical Laboratory, Teddington, UK, 1993.

[3] Billingsley, P.A., "Ergonomic Standards Go Beyond Hardware", IEEE Tr. on Software, p. 82-84, March, 1994.

[4] Cohill, A.M., Gilfoil, D.M., Pilitsis, J.V., and Carey, T., "Measuring the Utility of Application Software", Advances in human-computer interaction, Volume II, Ablex, Norwood, New Jersey, 1988.

[5] Dzida W., Herda, S., and Itzfeldt, W. D., "User-Perceived Quality of Interac tive Systems", IEEE Tr. on Software Engineering, Vol. SE-4, No. 4, pp. 270-276, 1978.

[6] Goodwin, N.C., "Functionality and Usa bility", Communications of the ACM, 30(3), pp. 229-233, 1987.

[7] Gould, J.D., and Lewis, C., "Designign for Usability : Key Principles and What Designers Think", Communica tions of the ACM, 28(3), pp. 300-311, 1985.

[8] Grudin, J., "The Case Against User Interface Consistency", Communication of the ACM, 32(10), pp. 1164-1173, 1989.

[9] Helander, M., Handbook of Human-Computer Interaction, Elsevier, pp. 755-927, 1988.

[10] Kiralowski, J., Porteous, M., and Cor bett, M., "How to use the software usability measurement inventory (SUMI)", Proceedings of European Conference on Software Quality, Madrid, November, 1992.

[11] Kadeyoo, N., "Ergonomic design ap proach", In Illustrated Ergonomics, Japan Institute of Standards, 1990.

[12] Mitta, D.A., "An application of the AHP : A rank-ordering of computer interfaces", Human Factors, 35(1), pp. 141-157, 1993.

[13] Nagamachi, M., Kansei Engineering, Tokyo, Japan : Kaibundo Press, 1989.

[14] Nielsen, J., "Iterative User-Interface Design", IEEE Computer, November, pp. 32-41, 1993.

[15] Park, K.S., and Kim, J.S., "Fuzzy weighted-checklist with Linguistic Variables", IEEE Tr. on Reliability, 39(3), pp. 389-393, 1990.

[16] Ravden, S.J., and Johnson, G.I., "Eva luating Usability of Human-Computer Interfaces", Ellis Horwood, Chichester, 1989.

[17] Shackel, B., and Richardson, S., Human Factors for Informatics Usability, Cambridge Univ. Press, 1991.

[18] Shneiderman, B., Designing the User Interface, 2nd Ed., Addison-Wesley, pp. 471-500, 1992.

[19] Simit, S.L., and Mosier J.N., "Guidelines for Designing User Interface Software", Report No. MTR-10090m Esd-TR-86-278, The Mitree Co., Beford, MA., 1986.