User Experience Index Scale —Quantifying Usability by Magnitude Estimation—

● Shinobu Utamura ● Chikako Murase ● Yukiyo Hamatani
● Yukinori Nagano

(Manuscript received November 7, 2008)

Conventional methods of evaluating usability often take a problem-discovery approach that focuses on making improvements, thereby limiting their use in product evaluation during the development process. Moreover, there are currently no effective techniques for obtaining the overall usability of a target product in a quantitative manner, which means that some doubt remains in usability-evaluation results. Magnitude usability, which applies the magnitude estimation method of psychophysics, is a sophisticated technique that can solve these problems. Using magnitude usability, Fujitsu Design Ltd. has developed the user experience (UX) index scale, which can be used in diverse human-centered design processes. The UX index scale can be easily understood by non-specialists and can be used to uncover problems and make policy decisions in various types of product-development scenarios.

1. Introduction

Human-centered design (HCD), or product development based on user-oriented design, has begun to attract attention among product developers. In HCD, developers aim to learn about diverse user characteristics, determine requirements based on those characteristics, and repeat the steps of design, evaluation, and improvement in an upward, spiral-like manner. The HCD process is essential in order to provide products that have a high degree of user satisfaction.

One step of product development based on HCD is design evaluation against requirements. At this time, it is common to perform a usability evaluation to check whether an easy-to-use product is being designed for the target user. Usability evaluations consist mainly of tests involving users and heuristic evaluations conducted by specialists. In either case, the idea is to uncover things that need to be improved upon and, for ones that can be reflected in product development, to make changes and raise the level of quality. This problem-discovery approach has been common in conventional product evaluations, but with this approach, it has been difficult to measure "overall usability" including the degree to which the product itself is easy to use and the user's sense of usability. For example, one common method uses questionnaires to evaluate a variety of elements related to ease-of-use (such as efficiency, consistency, and ease-of-learning) on 5 to 7 levels and then determine scores for ranking purposes. The results of such questionnaires, though, take time to assess and tabulate and may not reveal the overall picture. A target product may also be ranked by comparing product versions before and after improvements and by benchmarking against competitors' products, but differences in the scales of these methods cannot be accurately determined, so the results may be vague.

Magnitude usability (MU) developed by Mick McGee is a technique that provides a quick and easy solution to this problem. It is a groundbreaking approach that applies the magnitude estimation method of psychophysics to usability. Fujitsu Design Ltd. took notice of this technique from early on and has been conducting trials in development departments with the aim of solidifying the use of HCD. To this end, it has developed the user experience (UX) index scale as an expansion of MU and has been using and improving it as needed.

In this paper, we first describe the usefulness of MU and point out a problem with it. We then describe the development of the UX index scale as a new technique that compensates for that problem and explain why this technique is significant. Finally, we examine the effectiveness of the UX index scale through a study within Fujitsu Design Ltd. and suggest methods for its application.

2. Magnitude usability

2.1 Usability magnitude estimation

Usability magnitude estimation (UME), the basis for MU, was developed in 2003 as a new method for measuring usability.[1] The magnitude estimation method is the classical measuring method in psychophysics for making a direct quantitative estimate of sensory magnitude in response to a series of physical stimuli. Here, numerical values provide relative assessments without the use of units. For example, when a subject is presented with a certain length of string (physical quantity), a suitable numerical value will be assigned to the length as sensed by the subject (sensory quantity). Then, when the subject is presented with another length of string, that numerical value will be used as a basis for expressing the sensed length of the new string relative to the previous length. If the new string is felt to be twice as long as the original string, a value twice as large is applied. This magnitude estimation method enables a subject to numerically estimate the intensity of a stimulus in comparison with a standard stimulus. This measurement method is used widely, and UME was developed thinking that it would be highly applicable to measuring diverse sensory perceptions in response to complex physical stimuli as in the study of usability when designing user interfaces.

2.2 Master usability scaling

After the usefulness of UME was recognized, a new concept called master usability scaling (MUS)[2] was born. Based on UME, MUS introduces standard reference tasks with the aim of representing all data in different usability evaluations on a single scale. Although the master-scaling procedure of Berglund[3] establishes a basic concept for work in this field, MUS simplifies the application of standard reference tasks. In either case, the use of a master scale precludes the selection of a target product, which means that a direct quantitative comparison of usability can be made between otherwise disparate products such as a cell phone and personal computer. This is a function that could not be performed under traditional usability evaluation methods. McGee called this comparison axis the “universal usability continuum”.

2.3 Expected UME

An evaluation method called expected UME was next developed by Rich et al.[4] This method applies UME to measure the degree of user satisfaction. Specifically, it collects user expected values before task execution during the usual evaluation and compares them with actual values after task execution. Studies can then be performed on setting priorities for actions to be taken with regard to those tasks and on solutions to be implemented. Albert and Dixon[5] developed the basic concept for using expectation measures in usability testing: if expected and actual values are collected by UME, the difference in those values can be used to measure the effects of the current design on satisfaction and overall usability and to assign priority to improvements in a more accurate manner.

In the above way, MU has become the generic name for the above-mentioned measurement techniques based on UME.

2.4 Strengths and weaknesses of MU

The most outstanding feature of MU is its success in quantifying "overall usability", which is difficult for conventional methods. Furthermore, since MU enables diverse products to be compared and evaluated directly, it is said that the scope of HCD activities that apply MU has broadened. The MU technique can also be applied to experimental procedures by simply adding the practice of relative assessments to conventional user tests. Moreover, since the analysis of such assessments uses standard statistical-analysis techniques, the planning, execution, and analysis of such evaluation experiments is extremely simple for usability engineers.

At the same time, actual trials have revealed that a problem still exists with MU. While the use of MUS enables the usability of various products to be expressed on a single scale, the numerical values themselves have no meaning. This makes it difficult to assess what is actually good or poor in a product or to assess its degree of perfection during product evaluation and development. In other words, the McGee's universal usability continuum does not include any absolute values on its axis.

3. User experience index scale

3.1 Significance of index scale development

In order to solve the "intangible scale" problem described above, Fujitsu Design Ltd. is examining the use of tangible graduations that can act as a "ruler". Theoretically speaking, we can say that MUS takes the results of all user tests and presents them on a single axis. That is, data is collected and plotted on an endless number line. Now, from among that data, if we select comparative items (referred to as "benchmarks" below) that correspond to good or poor usability relative to a product targeted for evaluation (referred to as "target product" below), we get suitable graduations for that product's scale. In short, numerical values are given meaning through the use of concrete "things" (**Figure 1**).

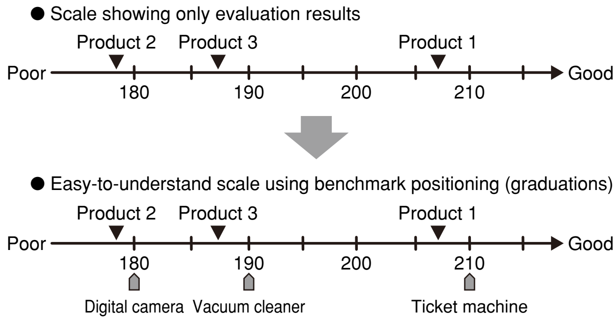


Figure 1
Example of scale improvement.

For example, when examining the usability of a certain product, it would be helpful if the extent to which that product is easy to use could be understood in comparison with a common product such as a home electrical appliance. Even better than home appliances as references would be devices that have high public exposure, such as train-station ticket machines and convenience-store copiers, which are used by many and diverse users, making them an ideal product group for making usability comparisons. If a certain product has equal or better usability with respect to such a familiar public product, we can say that the product in question is easy to use.

In recent years, conducting exchanges over the Internet has become quite common as reflected by the submission of electronic applications on the Websites of local governments and the execution of online transactions on the Websites of financial institutions. The problem of how to connect the virtual world of Websites with the real world has attracted much attention. However, if results such as "I could do three times as much shopping as expected by virtual Web-based shopping compared with actual shopping" were to be obtained, evaluation results would probably be more persuasive.

To give another example, there are few opportunities to benchmark corporate-oriented

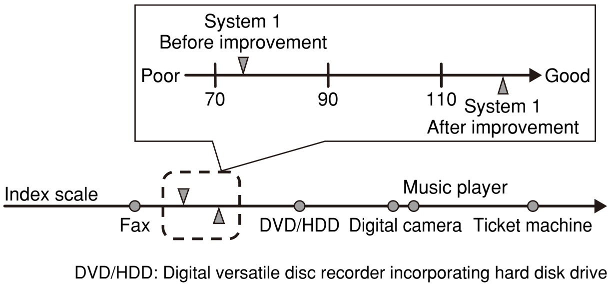


Figure 2
Consumer product evaluation with index scale.

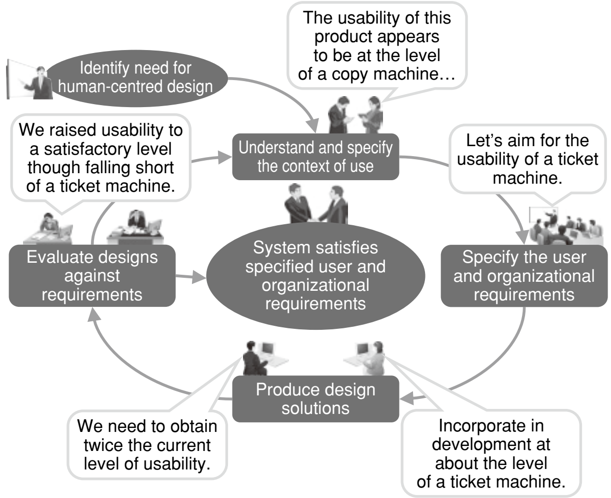


Figure 3
Using index scale for developmental policy making (in HCD process as defined in ISO13407).

systems such as middleware products against competitors' products. Furthermore, the users of such systems are limited to system developers and operators. It is consequently difficult for the developers themselves to determine the extent to which a system is complete. They may perform an evaluation and make some improvements relative to the previous version, but by lining up that result with progressive consumer products as benchmarks to obtain a bird's eye view in terms of operability, fun-to-use, or other factors, they will be able to see that the improvement that they have achieved may not necessarily be sufficient (**Figure 2**).

As described above, an evaluation within the same product group may not necessarily be sufficient for one to understand that product's usability. An index scale enables developers to obtain an objective and quantitative understanding of the usability of a product under development and helps to clarify later development policies and issues.

3.2 Index scale as communication tool

As described above, the creation of an index scale by selecting benchmarks corresponding to evaluation objectives is important because it indicates a clear direction to take in subsequent development steps in various HCD scenarios (**Figure 3**). The same can be said for smooth communication among development personnel and for the increasingly popular Persona method. One feature of the Persona method is that it creates a clear image of the user that can be shared among developers. In a similar way, an index scale can position the current degree of usability of a target product by selecting appropriate benchmarks, enabling that information to be shared among developers. A benchmark can also represent a target value. Saying, for example, that "usability should be of the same level as the task of buying a ticket from an automatic ticket machine" corresponds to a numerical value that is easy to understand and that can be used by developers to set clear objectives in the development process.

3.3 Creation of UX index scale

Our first objective in developing the UX index scale was to assemble a variety of well-known products that could be arranged evenly on a scale from poor to good usability. To this end, we recruited 12 men and women in their 30s and 40s as subjects and collected expected and actual MU values in a user-test format targeting the 10 products and 17 tasks (plus a standard reference task) listed in **Table 1**. (The tests were administered in cooperation with NIFTY Corporation.) The tabulated UME results and

Table 1
Target products and tasks in user tests.

Target product		Task	
1	DVD/HDD recorder	1	Play back recorded program
		2	Schedule a recording from program schedule
2	Digital camera	1	Take picture in full-auto mode
		2	Take picture manually after entering various settings
3	Music player	1	Play back music
		2	Use image viewer
4	Fax	1	Send a fax using telephone directory
		2	Make an enlarged copy and multiple copies
5	Vacuum cleaner	1	Turn on vacuum cleaner
		2	Empty rubbish container and set tissue filter
6	Recipe book	–	Search for recipe
7	Cup of noodles	–	Prepare a cup of noodles
8	Minivan	–	Fold up the third row of seats
9	Convenience-store copier	1	Make a color copy
		2	Print from a digital camera
10	Train-station ticket machine	1	Purchase an ordinary ticket
		2	Purchase a ticket with a transfer
–	Website login (reference task)	–	Log in to a self-made Website as a reference task

their graphs after we excluded statistically outlying values (one sample) are shown in **Figures 4** and **5**.

The results revealed that tasks using products having high public exposure like subway ticket machines and convenience store copiers ranked high. These results can be said to have matched the expectations of the subjects (Figure 5). One product for which subjects had a good impression beyond expectations was the music player. In this case, most of the subjects had never used a music player before and their expected values were not high at all. However, upon actually using the device, their impressions of it rose dramatically as they were surprised and pleased with various operations and discoveries that exceeded the expected usability range. This can be explained as a jump in the UME value due to the extensive amount of experience possessed by the subjects (users). Also, while the task of playing back a recording on a digital versatile disc recorder incorporating a hard disk drive (DVD/HDD recorder) may appear to be simple, a poor result was obtained contrary to user expectations. This was because with today's remote control devices, it is

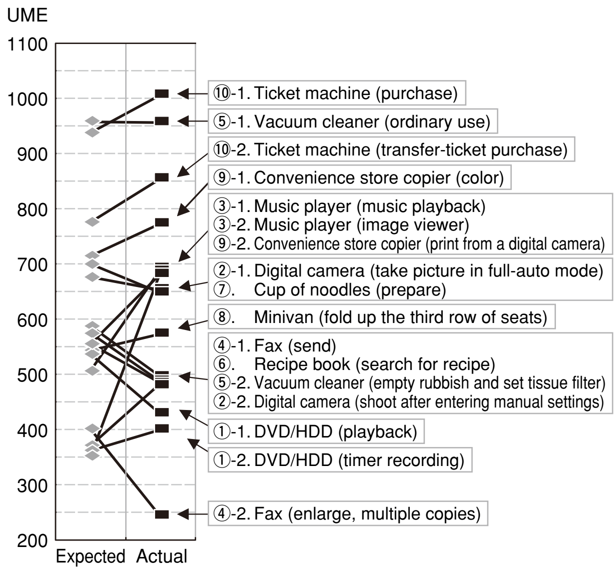
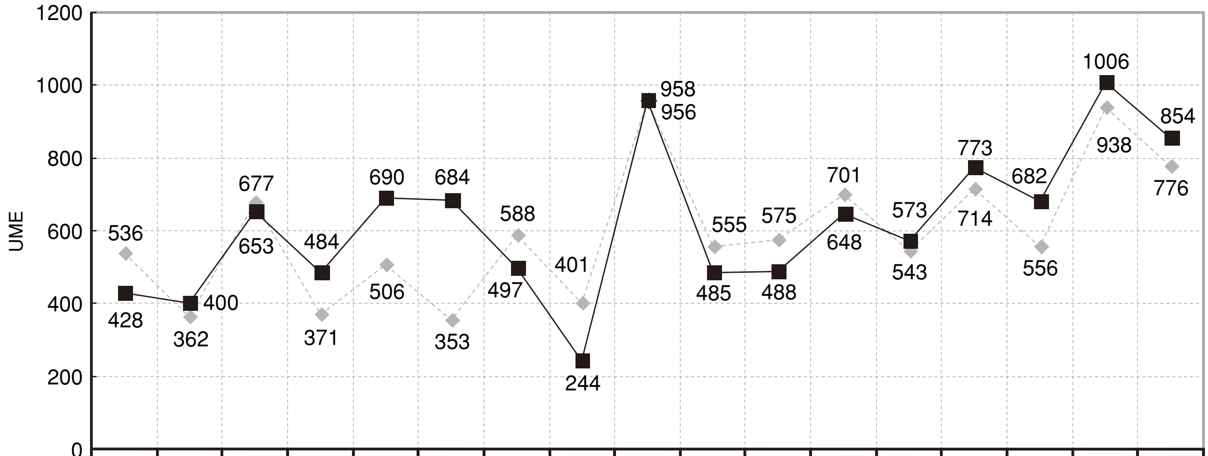


Figure 4
Differences between expected and actual UMEs within the same task.



	①-1. DVD/HDD (playback)	①-2. DVD/HDD (schedule recording)	②-1. Digital camera (take picture in full-auto mode)	②-2. Digital camera (shooting after making manual settings)	③-1. Music player (music playback)	③-2. Music player (image viewer)	④-1. FAX (send)	④-2. FAX (enlarge, multiple copies)	⑤-1. Vacuum cleaner (ordinary use)	⑤-2. Vacuum cleaner (throw out rubbish and set tissue filter)	⑥. Recipe book (search for recipe)	⑦. Cup of noodles (prepare)	⑧. Minivan (fold up the third row of seats)	⑨-1. Convenience store copier (color)	⑨-2. Convenience store copier (print from a digital camera)	⑩-1. Ticket machine (purchase)	⑩-2. Ticket machine (transfer-ticket purchase)
Expected	536	362	677	371	506	353	588	401	958	555	575	701	543	714	556	938	776
Actual	428	400	653	484	690	684	497	244	956	485	488	648	573	773	682	1006	854

Items 1–10 correspond to Table 1.

Figure 5
Expected and actual UMEs for each task.

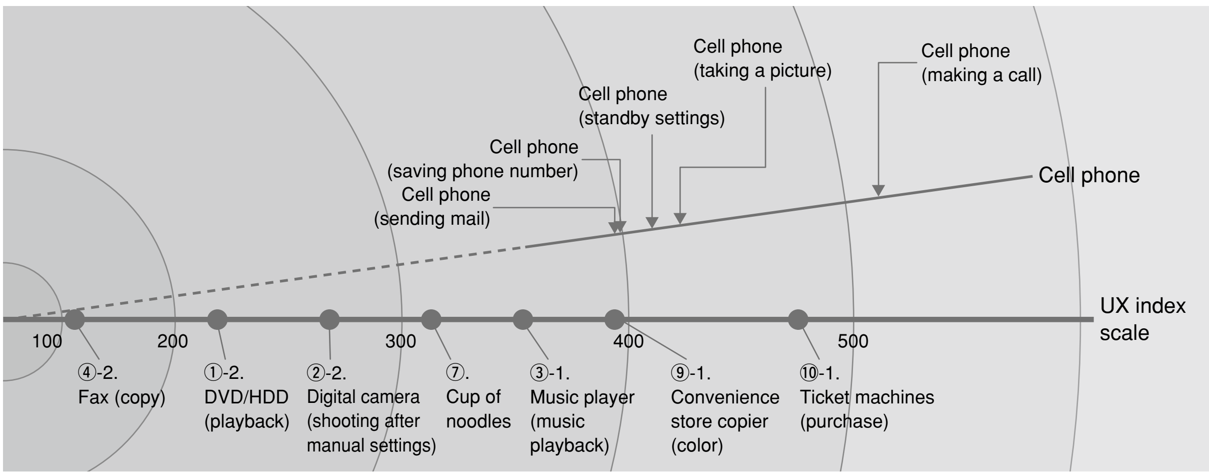


Figure 6
Evaluation of mobile-phone tasks against various benchmarks on UX index scale.

no longer just a matter of pushing the playback button. For the most part, a wide range of index values were obtained here from poor to good usability, giving a representative arrangement of benchmarks on the UX index scale.

The results of statistically processing a separately executed UME of a cell phone and arranging it on this index scale for ease of understanding are shown in **Figure 6**. As shown, the task of making a call on this cell phone had a higher value than purchasing a ticket from an automatic ticket machine while other typical cell-phone tasks ranked high as well. These results indicate that the ease-of-use of this cell phone ranks about the same as that of a convenience store copier having high public exposure.

4. Conclusion

This paper described the development of the user experience (UX) index scale for practical use in product development based on human-centered design. This scale uses the magnitude usability technique, which applies the magnitude estimation method of psychophysics. On the basis of empirical psychology experienced by users themselves, the UX index scale can be used to quantitatively and objectively express usability in a broad sense that includes surprises and fun in product operation in addition to ease of use (total user experience). It is an effective technique for uncovering problems and making policy decisions in diverse product-development scenarios.

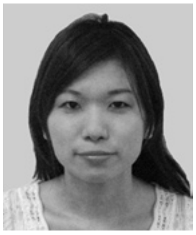
In future studies, we will investigate whether the UX index scale can be applied beyond usability and user experience to measure stimuli that appeal to *kansei* (sensitivity) in the field of product kansei quality. We will also study its usefulness as an index.

References

1) M. McGee: Usability magnitude estimation. Proc. HFES, 47th Annual Meeting, Denver, 2003, pp. 691–695.
2) M. McGee: Master Usability Scaling: Magnitude Estimation and Master Scaling Applied to Usability Measurement. Proc. SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 2004, pp. 335–342.
3) M. B. Berglund: Quality assurance in environmental psychophysics. In S. J. Bolanowski and G. A. Gescheider, (Eds.), Ratio scaling of psychological magnitude: In honor of the memory of S. S. Stevens. Lawrence Erlbaum Associates, Publishers, 1991, pp. 140–162.
4) A. Rich et al.: Expected Usability Magnitude Estimation. Proc. HFES, 48th Annual Meeting, 2004, pp. 912–916.
5) W. S. Albert et al.: Is this what you expected? The use of expectation measures in usability testing. Proc. Usability Professionals Association, 12th Annual Conference, Arizona, USA, 2003.

Shinobu Utamura
Fujitsu Design Ltd.

Ms. Utamura received the B.S. degree in Architecture from Meiji University, Kawasaki, Japan in 1999. She also received the M.S. degree in Design and Environmental Analysis from Cornell University, NY, USA in 2001. She joined Fujitsu Ltd. in 2002 and has been engaged in usability evaluations of products and systems. She has been with Fujitsu Design Ltd. since its spin-off in October 2007.



Yukiyo Hamatani
Fujitsu Design Ltd.

Ms. Hamatani received the B.S. degree in Design and Information and the M.S. degree in Systems Engineering from Wakayama University, Wakayama, Japan in 2005 and 2007, respectively. She joined Fujitsu Ltd. in 2007 and has been engaged in the study of usability evaluation methods and usability evaluations of products and graphical user interfaces. She has been with Fujitsu Design Ltd. since its spin-off in October 2007.

Chikako Murase
Fujitsu Design Ltd.

Ms. Murase received the B.A. degree in Human Relations and the M.S. degree in Psychology from Keio University, Tokyo, Japan in 2003 and 2005, respectively. She joined Fujitsu Ltd. in 2005 and has been engaged in usability evaluations of products and systems. She has been with Fujitsu Design Ltd. since its spin-off in October 2007. She is a member of the Japanese Psychological Association and the Japanese Society for Cognitive Psychology.

Yukinori Nagano
Fujitsu Design Ltd.

Mr. Nagano received the B.E. degree from Kyoto Institute of Technology, Kyoto, Japan in 1993. He joined Fujitsu Ltd. in 1993 and initially worked on the design of icons and graphical user interfaces. Since 1998, he has been engaged in the field of usability and accessibility. He has been with Fujitsu Design Ltd. since its spin-off in October 2007. He is a member of the Human Interface Society of Japan, the Japan Ergonomics Society, and the Color Science Association of Japan.