Value Added by the Axiomatic Usability Method for Evaluating Consumer Electronics

Yinni Guo[1], Yu Zhu[2], Gavriel Salvendy[1,3], and Robert W. Proctor[1,4]

[1] School of Industrial Engineering, Purdue University, West Lafayette, IN
[2] Department of Statistics, Purdue University, West Lafayette, IN
[3] Department of Industrial Engineering, Tsinghua University, Beijing, China
[4] Department of Psychological Sciences, Purdue University, Lafayette, IN
Yinni.Guo@gmail.com, {yuzhu,Salvendy}@purdue.edu, proctor@psych.purdue.edu

Abstract. In this paper we demonstrate how to use the axiomatic evaluation method to evaluate usability of consumer electronic products. The axiomatic evaluation method examines three domains of a product: customer, functional, and control domains. This method collects not only usability problems reported by the users, but also usability problems found through the mapping matrix between the three domains. To determine how well this new usability evaluation method works, an experiment was conducted to compare the axiomatic evaluation method with a think-aloud method. 60 participants were randomly assigned to use one method or the other to evaluate three popular consumer electronic devices. Number of usability problems discovered and completion time were collected and analyzed. Results showed that the axiomatic evaluation method performed better than the think-aloud method at finding usability problems for the mobile phone and about user expectation and control.

Keywords: axiomatic evaluation, consumer electronics.

1 Theoretical Background of Axiomatic Evaluation

The axiomatic evaluation method is a usability evaluation method developed based on axiomatic design theory. In a prior paper [1], we introduced the conceptual model of the axiomatic evaluation method. In the current paper we demonstrate the potential value of using the method to evaluate usability of consumer electronic products. Axiomatic design consists of four domains and two axioms [2]. The four domains are customer domain, functional domain, physical domain and process domain (see Fig. 1). The customer domain ([CA]) consists of the needs for which the customers are looking in a product. The functional domain ([FR]) consists of functional requirements, which are the minimum independent requirements that completely characterize the functional needs of the product. The physical domain ([DP]) consists of design parameters, which are the key variables in the physical domain that characterize the design that satisfies the specified functional requirements. Finally, the process domain ([PV]) consists of process variables, those in the process domain that characterize the

S. Yamamoto (Ed.): HIMI/HCII 2013, Part III, LNCS 8018, pp. 457–466, 2013.

process that can generate the specified design parameters. For successive domains in Fig. 1, the domain to the left represents “what we want to achieve”, whereas the domain to the right represents the design solution of “how we propose to satisfy the requirements specified in the prior domain.”

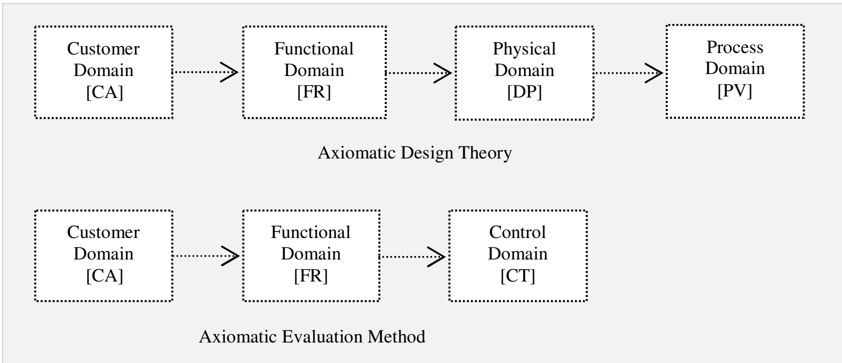

Fig. 1. Axiomatic Design Theory and Axiomatic Evaluation Method

The design flow suggested by axiomatic design theory consists of mapping from one domain to the other. The mapping process between domains can be expressed mathematically in terms of vectors. For example, a set of functional requirements in the functional domain can be written into a functional requirement vector ([FR]). Similarly, a set of design parameters has been chosen to satisfy the functional requirements constitutes the design parameter vector ([DP]). The mathematical expression can be then written as [FR] = [A][DP] or as the following equation, where [A] is the mapping matrix that characterizes the product design:

$$\begin{bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} DP_1 \\ DP_2 \\ DP_3 \end{bmatrix}$$

The axiomatic design theory examines the design quality with respect to two axioms, identified by Suh [2] through examining common elements always present in good designs. The first axiom, the *independence axiom*, states that the independence of functional requirements that characterizes the design goals must be maintained. In other words, when there are two or more functional requirements, the design solution must be such that each functional requirement can be satisfied without affecting others. The second axiom, the *information axiom*, states that among those designs that satisfy the independence axiom, the design that has the smallest information content is the best. In a real design scenario, there can be many designs that satisfy the independence axiom. However, one of those designs is likely to be superior. The information axiom provides a quantitative measure of a given design, and it is useful in selecting the best among those designs that are acceptable.

The axiomatic design theory helps to overcome shortcomings of the product development process based on a recursive "design/build/test" cycle, which requires continuing modifications and changes as design flaws are discovered through the testing [3]. The axiomatic design theory has been applied to a variety of products and systems such as mechanical design, software design, organizations management, and materials design. Although the meanings of the four domains are quite different from one product area to another, the axiomatic design theory successfully enhances the performance, robustness, reliability and functionality of products in different fields.

2 Conceptual Model of the Axiomatic Evaluation Method

The core idea of the axiomatic design theory is to start designing from understanding "what do the customers need", and to continue from there to specify "how could we achieve it" through providing the right functions, the appropriate design parameters for the functions, and the proper process variables for the design parameters. Similarly, in usability evaluation we try answer the questions of "what do the users want in the product" and "how well does the product satisfy user requirements". Therefore it is feasible to apply the framework of axiomatic design theory to usability evaluation, and three domains are set up for the usability evaluation process: customer domain ([CA]), functional domain ([FR]) and control domain ([CT]) (Fig. 1). Customer domain ([CA]) consists of customer requirements of the product, and they can be retrieved by an open-ended questionnaire. Functional domain ([FR]) consists of existing functions of the product, which can be easily retrieved by reading the product manual. Control domain ([CT]) consists of the control keys of the product; these also can be retrieved by examining the product or reading the product manual. The use of the two axioms is similar, but the axiomatic evaluation method also includes constraints for human-computer interaction design to examine the mapping matrix, such as maintaining stimulus-response compatibility [4], following Hick's Law [5] and adhering to Fitts's Law [6]. The biggest difference between the original axiomatic design theory and the axiomatic evaluation method is that in the latter the third domain is the control domain ([CT]) instead of the physical domain ([DP]). The main reason for this change is that, in the evaluation process, one is more interested in evaluating how the existing design parameters with which users interact (the control keys) support the functions, rather than in figuring out how the design parameters with which users will not interact perform. Usability problems can be found by examining the mapping between the three domains. The mapping can be expressed as matrices [X] and [Y]:

$$[CA]=[X][FR] \text{ or } [\text{Customer Requirements}] = [X]\,[\text{Functions}]$$

$$[FR]=[Y][CT] \text{ or } [\text{Functions}] = [Y]\,[\text{Control Keys}]$$

The mapping matrix [X] between customer domain ([CA]) and functional domain ([FR]) can provide an index of function sufficiency, which can be beneficial to prototypes at the beginning stage of product development. If the index is high, the current product could satisfy most customers. If the index is low, designers may want to reconsider the design direction. The mapping between functional domain ([FR]) and control domain ([CT]) shows how easy it is to control the device. Usability problems can be found by checking the mapping matrix [Y] between the functional and control domains which is determined by users' operation. According to the independence axiom, an ideal mapping matrix should be a diagonal matrix or a triangular matrix. The former means that users need only one step to control the function, and each function has its own control key. The latter means that users may need more than one step (press more than one key) to complete the task, but the keys used are not conflicting with the ones used in other functions. So if the [Y] matrix is a diagonal matrix or a triangular matrix, it means there are no conflicting controls between different functions. If the [Y] matrix is neither diagonal nor triangular, the designers should reconsider the control design to avoid possible usability problems (e.g., high error rate caused by using the same combination of keys for two functions).

However, only meeting the requirement of a diagonal matrix or a triangular matrix is not enough. According to the information axiom, the design that has the smallest information content is the best alternative. Therefore, designers should also make sure that users do not need to take too many steps in order to complete a certain task. Other constraints like stimulus-response compatibility, Hick's Law, and Fitts's law for movement times, mentioned previously, can be used to determine usability problems as well. For instance, Hick's law describes the time it takes for a person to make a decision as a result of the number of possible choices that he or she has. In axiomatic evaluation, if the mapping matrix [Y] shows that there are too many options under one menu, it may take users a long time to select the option they want. Another example is Fitts's law, which is used to model the time to move the hand or a cursor from a current position to a target position. In axiomatic evaluation, if the mapping matrix [Y] shows that the keys used for one function are located far away from each other, it is possible that Fitts's law was not followed well.

The proposed axiomatic evaluation method could be used in the formative stage of product development before a final design is accepted for release. Compared to traditional usability evaluation methods, the axiomatic evaluation method is likely to discover more usability problems related to user requirement and control. Examining the mapping matrix between customer domain ([CA]) and functional domain ([FR]) will reveal what the customers need and what is barely used. Examining the mapping matrix between functional domain ([FR]) and control domain ([CT]) will reveal the problems about control.

3 Experiment

3.1 Experimental Design

Because the axiomatic evaluation method is a new tool for usability evaluation, we conducted an experiment to assess how well this evaluation method worked and in which aspect it would perform better or worse than other usability evaluation methods. Since the axiomatic evaluation method is task-specific and intended to be used in the formative stage before the design is finalized, we chose another task-specific evaluation method used at the same stage for comparison, the think-aloud method. This method was selected also because it did not require participants to have much knowledge about usability. The experiment used a between-subjects design: 60 participants were recruited and randomly assigned to the group using the think-aloud method or the group using the axiomatic evaluation method. Three popular consumer electronic devices: music player, digital camera and mobile phone representing different levels of complexity were evaluated by each participant in a randomized order.

Using the Axiomatic Evaluation Method. Participants in the axiomatic evaluation group were first asked to complete a background questionnaire. Then, the experimenter introduced the purpose of the study and encouraged the participants to raise any problems or questions regarding to the test products. Although the participants were not required to tell the experimenter everything in their mind, as in the think-aloud method, they were encouraged to voice any problem or question relating to the test device. This way, the experimenter could collect usability problems found by the participants without interfering with their performing the tasks. As the experiment started, participants were asked to fill out a questionnaire of what functions they really used and their expectations of the product. After that, participants were asked to evaluate the product's functions that they said they would really use. The order of the tasks was randomized for each participant. The evaluating procedure was videotaped, and the problems raised by the participants were written down by the experimenter. After evaluating each test product, participants were asked to fill out a satisfaction questionnaire. Upon completion of the experiment, the experimenter went through the videos to fill out the mapping matrix [Y] between function and control keys according to participants' action—which keys did they click in order to perform a certain task. By examining the errors participants made, and by examining the mapping matrix, usability problems not reported by the participants could be identified, as well as the reasons for the usability problems.

Using the Think-Aloud Method. The think-aloud method is widely used in the same stage of product development in laboratories and industries. This method, developed by Lewis [7] and refined by Ericsson and Simon [8], has proved to be successful in collecting qualitative data from a small number of users [9]. The think-aloud procedure involves participants thinking out loud as they perform a set of specified tasks. In a think-aloud evaluation, users describe whatever they are looking at, thinking, doing, and feeling, as they go through their task. This allows the experimenter to understand how the task is completed. The experimenter records everything the participant says, without attempting to interpret the actions and words. Test sessions are usually videotaped or audiotaped so that experimenter can go back and refer to what participants did and how they reacted. However, there is a limitation of the think-aloud method: it seems unnatural to test users and may influence users' problem-solving behavior.

In this study, we employed the specific think-aloud procedure that has been used in more than 30 studies (e.g. [10]). After the experimenter introduced the purpose of the study, participants were asked to fill out a background questionnaire. Then, printed instructions of how to perform "thinking aloud" were given to the participants. The instructions were based on the methodology developed by Lewis [7]: "The basic idea of thinking aloud is that you ask your users to perform a test task, but you also ask them to talk to you while they are working on it. Ask them to tell you what they are thinking: what they are trying to do, questions that arise as they work, things they read." After this, participants were asked to watch a video of how to perform the think-aloud method and to do a warm-up practice. When the experiment formally started, participants were given a list of tasks to perform—to use a set of functions of a product. The experimenter videotaped the whole procedure and took notes of the thinking reported by the participant. The participants' thinking included not just the problems they encountered, but also whether they thought a certain function was well designed, or what kind of design they liked, or their suggestion for the product. After the evaluation, participants were asked to fill out a satisfaction questionnaire.

3.2 Participants and Test Products

The number of participants was determined by statistical power analysis and experiment design requirements. In this experiment, each participant would use one of the two methods to evaluate all three products in one of the six testing orders. Assuming the standard deviation of the number of usability problems found would be approximately 5, and the maximum difference between the means of usability problems found would be 20, we calculated that having 4 participants in each testing order can give us a power value of 0.9 (given $\alpha = 0.05$). Thus, we decided to have 5 participants in each of the 6 testing orders for both evaluation methods, 60 participants in total.

Participants who had experience with at least 2 of the 3 consumer electronic products (music player, mobile phone and digital camera) were recruited through e-mail from an electronic product company located in Xiamen, China. More than 90% of the participants had used more than 1 music player for at least 2 years, and over 70% of them had used at least 3 mobile phones for more than 5 years. The participants had less experience of using digital cameras, but still more than half of them had used at least 2 models for more than 2 years. Participants in the think-aloud group and the axiomatic evaluation group had similar distribution in gender, age, education, job category, and experience of using the 3 consumer electronic products.

Three widely used consumer electronics were chosen as the test products: music player, mobile phone and digital camera. The music player had eight major functions (playing music, recording, playing recorded soundtrack, radio, games, e-book, picture display, address book) and six control buttons. The digital camera offered multiple photo shooting modes (automatic, manual, portrait, landscape, moving mode, night mode, video recording) and the button layout was similar to most point and shoot digital cameras. The mobile phone had all common smart phone features. It had a touch screen and a key pad of 21 keys including the 12 number keys.

3.3 Results

Number of Usability Problems Found. The number of usability problems found was the total number of different usability problems discovered by all participants using the same evaluation method. To obtain this number, the experimenter first collected all usability problems found by all participants using the same evaluation method and then filtered out all the different ones. Each usability problem may have different weight, but in this study we weighted them equally. The usability problems found by the two evaluation methods were then sorted into eight main categories: content, menu, panel, display, control, functions, technology and appearance. A summary of usability problems is shown in Table 1. For the music player, the axiomatic evaluation method found 245 problems, which is slightly more than the think-aloud method (212 problems). For the digital camera, the axiomatic evaluation method found 193 problems, a value slightly smaller than the think-aloud method (215 problems). For the mobile phone, the axiomatic evaluation method found over twice as many problems as the think-aloud method (404 vs. 161). A closer look reveals that the "extra" usability problems found with the axiomatic evaluation method were mainly from four categories: content (lack of information, incorrect information, and ambiguous information), panel (location of the keys, shape of the keys), display, control (don't know which key to press, control bug) and functions.

The comparison of the two evaluation methods for the mobile phone showed an advantage of the axiomatic method in finding usability problems of user requirement and usability problems about control. This advantage was not evident for the music player and digital camera. One possible reason is that in recent years the mobile phone has become more complex than the music player or digital camera and is likely to have more usability problems. Also, nowadays people use mobile phones so often that they may become more stringent in their evaluation of the phones and thus point out more usability problems. Both reasons lead to the conclusion that participants would discover more usability problems in the mobile phone than the music player or digital camera (as the data in Table 1 show). According to Ericsson and Simon's research [11], think-aloud participants retrieve information from short-term memory. When evaluating a mobile phone (a complex device with more usability problems), a participant may not have been able to report all of the usability problems noticed while performing the task, possibly because doing so would break the flow of performance or because he/she forgot some usability problems after explaining the first few (limit of short term memory). Although a participant was evaluating a complex device when using the axiomatic evaluation method, similar to the think-aloud method, not all of the usability problems s/he noticed could be spoken aloud, but some of the hidden usability problems were caught by examining the mapping matrix (how the participant interacted with the device) recorded in the video. That is why the axiomatic evaluation method could find more usability problems than the think-aloud method on a complex device.

Table 1. Summary of Total Number of Problems Found

Category	Music Player		Digital Camera		Mobile Phone	
	TA*	AE*	TA	AE	TA	AE
About content						
Lack of information	14	5	12	9	2	18
Incorrect information	0	1	2	2	0	5
Ambiguous information	3	5	5	9	3	12
Redundant information	0	1	0	0	8	16
Lack of description in Chinese	0	3	5	7	0	0
Sub-total	17	15	24	27	13	51
About menu						
The sorting	8	7	1	1	21	29
The order or priority	0	0	0	0	0	7
Sub-total	8	7	1	1	21	36
About panel						
Position/location of the keys	5	17	6	13	0	7
Shape	3	3	3	3	0	4
Material/touch feeling	16	9	15	10	6	9
Label on the keys	19	12	9	8	3	4
Lack of keys	0	1	4	2	3	5
Sub-total	43	42	37	36	12	29
About display						
Font size, style, color	8	9	2	2	3	10
Screen or display	3	3	1	1	0	3
Icon	5	1	1	1	3	3
Format	11	10	0	0	2	7
Sub-total	27	23	4	4	8	23
About control						
Don't know which key to press	41	38	3	7	14	23
Control bug	16	17	1	3	14	37
Not convenient design	22	16	6	6	33	47
Sub-total	79	71	10	16	61	107
About functions	11	42	16	20	20	121
About technology	27	35	62	59	20	36
About appearance	0	10	61	30	6	1
Total	212	245	215	193	161	404

* TA stands think aloud, AE stands for axiomatic evaluation.

Completion Time. The completion time was measured by the timer on the video recorder. The experimenter started the recording when the participant was ready to perform a task, and stopped it when the participant said that s/he was finished. The elapsed time was the completion time. Statistical analysis of completion time is listed in Table 2. Normality tests conducted on each of the respective time measures showed no violations of normality, $p > 0.10$. For the think-aloud method, the total completion time, consisting of the time of training, evaluating the products, and filling out the satisfaction questions, was 51.13 minutes. For the axiomatic evaluation method, the total completion time, for which there is no training time but time for the experimenter to review the videos and identify usability problems, was 52.85 minutes. An analysis of variance (ANOVA) showed this small difference in total completion time to be nonsignificant (bottom row of Table 2).

Table 2. Statistical Summary of Completion Time (in min)

Test Segment	Think-Aloud Mean (std)	Axiomatic Evaluation Mean (std)	*F*	*p*
Training	6.47 (0.68)	-	-	-
Background questionnaire	0.80 (0.20)	0.83(0.02)	0.27	0.626
Evaluate music player	15.67 (2.80)	9.78 (2.83)	59.80	<0.0001
Evaluate mobile phone	20.75 (3.03)	15.93 (2.78)	35.86	<0.0001
Evaluate digital camera	6.33 (0.68)	5.07 (1.27)	16.17	0.0002
Satisfaction questionnaire	2.05 (0.33)	2.23 (0.22)	2.40	0.1266
Video review	-	20.2 (5.38)	-	-
Total	51.13 (4.75)	52.85 (7.10)	1.19	0.2791

A generalized linear model was used to test the interaction between type of evaluation (think-aloud and axiomatic) and product (music player, mobile phone, digital camera), and the interaction was significant. The benefit in evaluation time for the axiomatic evaluation method over the think-aloud method was larger for the music player (5.89 min) and mobile phone (4.82 min) than for the camera (1.26 min). However, individual ANOVAs for each product showed evaluation time to be shorter with the axiomatic evaluation method than with the think-aloud method in all cases (Table 2). Because participants spent much less time evaluating the camera than the other two products, the percentage improvement with the axiomatic evaluation method was still 80%, compared to 77% for the mobile phone and 62% for the music player. From the participants' point of view, the axiomatic evaluation method is more efficient than the think-aloud method because they are finishing their part in less time.

4 Discussion and Future Research

This experiment showed that the axiomatic evaluation method performed better at finding usability problems for a mobile phone than did the think-aloud method. But the axiomatic evaluation method did not show advantages in finding usability problems for the music player or digital camera. In particular, when used to evaluate the mobile phone, the axiomatic evaluation method was able to identify more usability problems about user requirement and control. Therefore, we suggest using the axiomatic evaluation method for evaluating products similar to mobile phones, e.g., tablets and laptops that have a lot of features and a high level of complexity, and thus are likely to have more usability problems. On the other hand, axiomatic evaluation may not benefit evaluation of products with a lower level of complexity and just a few features, like web cameras. This experiment also compared the completion time using the two evaluation methods. Although there was no significant difference between the total time for the methods, the axiomatic evaluation method required significantly less time than the think-aloud method for participants to evaluate each individual product.

The ultimate goal of usability evaluation is to improve user experience of products. What is unique about the present study is that it develops and validates a systematic evaluation method that can discover usability problems from "what do the users want" to "how well does the product satisfy user requirements". This type of evaluation procedure could reduce the time and energy of the "design-test-redesign" cycle. The new evaluation method also complements finding usability problems about control. One limitation of the study is that only two usability evaluation methods and three testing products were compared. In the future, the axiomatic evaluation method needs to be applied to more products with a larger range of complexity, and compared with other evaluation methods on more devices.

References

1. Guo, Y., Proctor, R.W., Salvendy, G.: A conceptual model of the axiomatic usability evaluation method. In: Smith, M.J., Salvendy, G. (eds.) HCII 2011, Part I. LNCS, vol. 6771, pp. 93–102. Springer, Heidelberg (2011)
2. Suh, N.P.: The Principles of Design. Oxford University Press, New York (1990)
3. Suh, N.P.: Axiomatic Design: Advances and Applications. Oxford University Press, New York (2001)
4. Proctor, R.W., Vu, K.-P.L.: Selection and Control of Action. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 3rd edn., pp. 89–110. John Wiley, Hoboken (2006)
5. Schneider, D.W., Anderson, J.R.: A Memory-based Model of Hick's Law. Cognitive Psychol. 62, 193–222 (2011)
6. Guiard, Y., Beaudouin-Lafon, M.: Fitts' Law 50 Years Later: Applications and Contributions from Human–Computer Interaction. Int. J. Hum.-Comp. St. 61, 747–750 (2004)
7. Lewis, C.: Using the" Thinking-Aloud" Method in Cognitive Interface Design. IBM Research Report RC 9265. IBM TJ Watson Research Center, Yorktown Heights, NY (1982)
8. Ericsson, K., Simon, H.: Verbal Reports on Thinking. In: Faerch, C., Kasper, G. (eds.) Introspection in Second Language Research, pp. 24–54. Multilingual Matters, Clevedon (1987)
9. Nielsen, J.: Usability testing. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 2nd edn., pp. 1543–1568. John Wiley & Sons, New York (1997)
10. Magliano, J., Millis, K.: Assessing Reading Skill with a Think-Aloud Procedure and Latent Semantic Analysis. Cognition Instruct. 21, 251–283 (2003)
11. Ericsson, K.A., Simon, H.A.: Protocol Analysis: Verbal Reports as Data, rev edn. The MIT Press, Cambridge (1993)