

Development of a Usability Questionnaire for Automation Systems

Akihiro Maehigashi¹(✉), Kazuhisa Miwa¹,
Kazuaki Kojima², and Hitoshi Terai³

¹ Graduate School of Information Science, Nagoya University, Nagoya, Japan
maehigashi@cog.human.nagoya-u.ac.jp, miwa@is.nagoya-u.ac.jp

² Learning Technology Laboratory, Teikyo University, Tokyo, Japan
kojima@lt-lab.teikyo-u.ac.jp

³ Faculty of Humanity-Oriented Science and Engineering,
Kindai University, Osaka, Japan
terai@fuk.kindai.ac.jp

Abstract. In this study, we positioned automation systems as the third-generation artifacts and developed a generalized usability questionnaire with 18 questions for automation systems as daily used artifacts. This questionnaire could be used to evaluate various types of automation systems and is useful for the development and improvement of automation systems as artifacts used in our everyday life.

Keywords: Automation system · Usability test · Questionnaire

1 Introduction

1.1 Automation System

Today, technological progress offers many opportunities to use automation systems such as cleaning robots and automated driving systems in everyday life. An automation system is a technology that autonomously conducts a task on behalf of humans [1]. Their use is expected to increase further in the near future.

Automation systems provide a completely new interaction between people and artifacts. Classical first-generation artifacts, such as knives and hammers, are used to support users' physical activities. The physical structures of such artifacts are usually simple and their functions are easy to understand. Therefore, in the interaction between people and classical artifacts, the artifacts normally perform along the users' intentions [2]. Furthermore, cognitive artifacts, such as computers and smartphones, are used to support users' cognitive activities by storing, expressing, and manipulating information [3]. Such artifacts are considered second-generation artifacts. As the inner structures of such artifacts are complex, their functions difficult to comprehend. Therefore, in the interaction between people and cognitive artifacts, discrepancies between the users' intentions and the artifacts' performance often occur [4].

Automation systems, in contrast, autonomously conduct tasks in the users' physical and cognitive activities. Moreover, when using classical or cognitive artifacts, the central player of each task is the user, and the artifacts support the user. However, when automation systems are employed, the central player of each task is the system, and the users monitor and supervise the systems. Thus, automation systems are considered third-generation artifacts. Additionally, because automation systems conduct far more complex information processing than first- and second-generation artifacts, users generally cannot understand the internal functions of the systems. However, if the discrepancies between the users' intentions and the automation systems' performance occur, they could lead to fatal accidents [1].

1.2 Usability Test

Artifacts used in everyday life are often evaluated through usability tests. Usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [5]; this is the world standard definition of usability [6]. The evaluation targets are usually cognitive artifacts, mainly computer software systems, to be developed and improved [6].

On the other hand, automation systems are usually evaluated based on the users' trust in the systems or their assigned workload [1]. The evaluation targets, in this case, are automated control systems in nuclear and manufacturing plants and automated operating systems of airplanes and ships. With the increased use of automation systems in daily life, their usability must be evaluated as well. Objective data on the use of automation systems may be obtained in multiple ways. However, automation systems cannot be evaluated using existing usability test questionnaires, such as SUS (system usability scale) [7], SUMI (software usability measurement inventory) [8], and QUIS (questionnaire for user satisfaction) [9], because the features of the automation systems, their autonomy and the complex inner information processes, are not considered in these questionnaires. Therefore, a new evaluation index is required, which will consider the features of automation systems.

In this study, we developed a generalized usability questionnaire for evaluating automation systems, which can be applied under different environmental settings.

2 Questionnaire Design

The usability elements effectiveness, efficiency, and satisfaction defined in ISO [5] were reflected in the usability questionnaire for automation systems. Effectiveness is "the accuracy and completeness with which users achieve specified goals," efficiency is "the resources expended in relation to the accuracy and completeness with which users achieve goals," and satisfaction is "the freedom from discomfort, and positive attitudes towards the user of the product" [5].

We added three new usability elements to the questionnaire to evaluate automation system usability. The first element is understandability the comprehensibility of the intentions of automation systems. Automation systems autonomously conduct tasks; however, as users cannot usually comprehend the internal functions of the systems, it is desirable for users to understand the intentions of the systems without discrepancies [10]. The second element is discomfort the absence of comfort or ease in using automation systems. Because automation systems receive information from the external environment and make decisions using this information, automation systems sometimes achieve tasks with users' unpredicted behaviors [10]. Furthermore, the task-performing process influences users' subjective evaluations of automation systems [11]; users feel comfort or discomfort because of the task-performing process, regardless of the task performance of the systems. The third element is motivation the users' desire to conduct the tasks by themselves. Because automation systems autonomously conduct tasks instead of the users, users tend to significantly depend on the systems [1, 12]. Additionally, users' overreliance on the systems reduces their motivation and ability to conduct tasks by themselves [13]. Automation systems should not take away the users' motivation and ability to conduct tasks.

Based on these considerations, we developed a usability questionnaire for automation systems comprising six questions for each element effectiveness, efficiency, satisfaction, understandability, discomfort, and motivation and a total of 36 questions, rated on a five-point scale: 1. Strongly disagree, 2. Disagree, 3. Neither agree nor disagree, 4. Agree, 5. Strongly agree.

3 Experiment

We conducted an experiment using and evaluating automation systems with the designed usability questionnaire to investigate its reliability and validity.

3.1 Experimental Task

Line-Tracing Task. We used a line-tracing task as an experimental task, a modified version of the line task used by Maehigashi et al. [14]. In this task, participants traced a line that scrolls downward past a circular vehicle. When the vehicle veered off the line, the performance score reduced according to the operational error. The participants were allowed to switch to either the auto-tracing mode (operation performed entirely by the system) or the manual mode (operation performed by participants using left and right arrow keys). The circular vehicle in the task had a diameter of 24 pixels. The window scrolling speed was 24 pixels per second. In the line task, the line was five-pixels wide. When operational errors occurred, a flashing red square frame appeared around the task window as error feedback.

Auto-tracing Systems and Task Situations. We set up two types of auto-tracing systems: the operational pathway-displaying system and the non-displaying system (Fig. 1). The operational pathway-displaying system indicated the pathway (light green) along which the system tried to operate the vehicle. The non-displaying system, in contrast, did not indicate the pathway. The participants used and evaluated these systems in various situations: (1) the system accurately traced the line or failed to trace it, (2) the vehicle stably traced the line or unstably snaked along the line in the auto-tracing mode, and (3) pressing a selector on the keyboard or typing a command, switch, when it was necessary to switch from the manual to the auto-tracing mode.

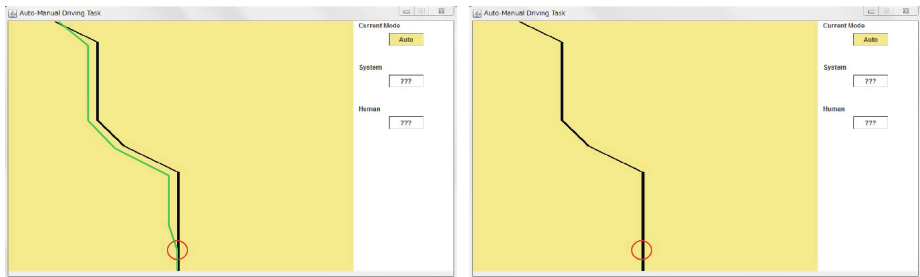


Fig. 1. Two types of auto-tracing systems: operational pathway-displaying system (left) and non-displaying system (right).

The auto-tracing system performance was controlled by the success rate of the operation command. In the experimental task, the system's operation in the auto-tracing mode was generally reflected in the vehicle movements with a 50 Hz sampling frequency. We controlled the percentage of how much the system's operation reflected in actual vehicle movements. In one situation, the system's operation, which was always perfect, was reflected in 100 % of the vehicle's movements. Therefore, the vehicle did not veer off from the line. In another situation, however, the system's operation was reflected in only 40 % of the vehicle's movements. The remaining 60 % of the system's operation was accepted in the experimental task system as no operational command. Therefore, the vehicle did not appropriately trace the line.

Auto-tracing system behavior was controlled by the observed vehicle behavior, independent from the system performance. If the system performed perfectly, the center of the circular vehicle was on the line. In another situation, although the system performed perfectly, the vehicle unstably snaked along the line. However, the circular vehicle was always on the line.

Switching from the manual to the auto-tracing mode was controlled by the command input method. The participants either pressed the spacebar or typed a command to switch from the manual to the auto-tracing mode. Switching back from the auto-tracing to the manual mode was performed by pressing the spacebar.

3.2 Method

Participants. Forty-eight university students participated in the experiment.

Procedure. Half the participants used the operational pathway-displaying system, and the other half used the non-displaying system. Each participant conducted the task in the eight different situations created by combining the task situations mentioned above. Performing the task in each situation took five minutes. After the participants completed the task in each situation, they evaluated the system with the usability questionnaire. After the evaluation, they took a two-minute break. Before they engaged in the task in each situation, they were instructed to achieve as high a score as possible and to use the auto-tracing system to evaluate the system.

Results. First, factor analysis based on the maximum likelihood method with varimax rotation was performed on the 384 data points collected from the 48 participants. As a result, six factors were extracted that explained 73.77% of the variability in the dataset (Table 1).

The first factor loadings exceeded .50 in six questions for effectiveness and in two questions for efficiency, and were below .50 in one question for motivation. Therefore, the first factor is considered to indicate effectiveness. The second factor loadings exceeded .50 in four questions for efficiency. Consequently, the second factor is considered to indicate efficiency. The third factor loadings exceeded .50 in six questions for satisfaction. Therefore, the third factor is considered to indicate satisfaction. Regarding the fourth factor, the factor loadings exceeded .50 in six questions for understandability and in one question for discomfort. Therefore, the fourth factor is considered to indicate understandability. Regarding the fifth factor, because the factor loadings exceeded .50 in five questions for discomfort, it is considered to indicate discomfort. Finally, the sixth factor loadings exceeded .50 in five questions for motivation; thus, the sixth factor is considered to indicate motivation.

Next, we selected three questions whose factor loadings were among the three highest for each factor, and we prepared a questionnaire comprising these 18 selected questions. We performed factor analysis using the maximum likelihood method with varimax rotation of the collected data for these 18 questions. As a result, six factors were extracted, and the validity of the questionnaire was confirmed (Table 2). The accumulated proportion of the six factors was 72.30%. In each factor, the factor loadings exceeded .50 for three questions in each element.

Furthermore, we calculated Cronbach's alpha coefficient among the three questions for each factor. The coefficient was .959 for the first factor, effectiveness; .833 for the second, efficiency; .872 for the third, satisfaction; .836 for the fourth, understandability; .911 for the fifth, discomfort; and .933 for the sixth, motivation. Since the alpha coefficient for each factor exceeded .80, the reliability of the questionnaire was confirmed.

Table 1. Thirty-six questions and factor loadings. Factor loadings greater than .50 are in gray. (R) indicates reverse scored question.

	Factors					
	1	2	3	4	5	6
Effectiveness						
This system did not conduct the task accurately. (R)	.79	.10	.25	.24	.23	-.10
I think that I would not make any mistakes by using this system.	.81	.06	.29	.21	.22	-.15
I was able to perform the task accurately by using this system.	.82	.12	.27	.22	.21	-.17
Using this system incurs errors during the task. (R)	.77	.09	.23	.19	.13	-.15
This system prevented me from committing errors during the task.	.81	.13	.31	.26	.20	-.10
This system decreases the task accuracy. (R)	.76	.13	.32	.29	.19	-.13
Efficiency						
Using this system incurs fatigue. (R)	.40	.58	.26	.26	.30	-.08
This system made the task easier.	.81	.17	.27	.26	.21	-.17
This system let me complete the task with less effort.	.84	.17	.28	.27	.13	-.12
I think that someone's support is necessary to use this system. (R)	.01	.69	.23	.22	.16	-.17
I was able to get used to this system in a short amount of time.	.11	.73	.20	.28	.15	-.11
By using this system, performing the tasks became more difficult. (R)	.44	.61	.28	.30	.19	-.13
Satisfaction						
I would never use this system again. (R)	.30	.20	.81	.16	.06	-.05
I did not feel any anxiety to use this system.	.19	.07	.70	.16	.25	-.24
I am satisfied with this system.	.34	.19	.82	.14	.10	.02
I was worried about using this system. (R)	.31	.14	.82	.14	.12	-.13
I am not satisfied with this system. (R)	.30	.17	.83	.08	.11	-.08
I would like to use this system again.	.40	.18	.78	.16	.03	-.07
Understandability						
I was able to predict the action of the system during the task.	.22	.13	.08	.74	.08	-.02
This system behaved against my expectation. (R)	.27	.13	.15	.72	.38	-.07
I could predict how this system behaved.	.27	.16	.19	.68	.13	.01
This system never behaved against my expectation.	.27	.10	.13	.67	.43	-.01
I think nobody can predict this system. (R)	.12	.20	.08	.76	.03	-.21
I had no idea what this system would do. (R)	.23	.21	.16	.73	.26	-.11
Discomfort						
This system behaved smoothly.	.39	.27	.17	.26	.64	-.02
I did not get used to this system's behavior.	.37	.36	.19	.44	.12	-.11
This system sometimes behaved unnaturally. (R)	.25	.14	.09	.62	.58	-.09
This system behaved naturally.	.38	.19	.17	.38	.65	-.13
I did not feel any discomfort with this system.	.34	.12	.24	.46	.62	-.09
This system behaved awkwardly. (R)	.28	.30	.13	.30	.70	-.14
Motivation						
Based on the experience of using this system, I feel like performing better by myself.	-.02	-.10	.05	-.07	.05	.62
It is more fun to perform the task by myself than relegating the task to the system.	-.13	-.04	-.08	-.03	-.12	.75
Based on the experience of using this system, I feel like performing the task by myself.	-.27	-.08	-.18	-.13	-.06	.84
Based on the experience of using this system, I think I should conduct the task by myself, only when it is necessary.	-.48	-.19	-.22	-.23	-.11	.22
I can perform the task by myself better without using the system.	-.33	-.10	-.16	-.04	-.12	.73
I feel like delegating the task to the system for as long as possible. (R)	-.57	.01	-.22	-.10	-.13	.33

Table 2. Eighteen questions and factor loadings. Factor loadings greater than .50 are in gray. (R) indicates reverse scored question.

	Factors					
	1	2	3	4	5	6
Effectiveness						
I think that I would not make any mistakes by using this system.	.79	.29	.18	.12	.19	-.25
I was able to perform the task accurately by using this system.	.78	.28	.18	.14	.22	-.20
This system prevented me from committing errors during the task.	.76	.29	.27	.08	.19	-.25
Efficiency						
I think that someone’s support is necessary to use this system. (R)	.29	.76	.29	.19	.15	-.14
I was able to get used to this system in a short amount of time.	.18	.77	.33	.22	.21	-.10
By using this system, performing the tasks became more difficult. (R)	.32	.68	.33	.23	.26	-.17
Satisfaction						
I am satisfied with this system.	.17	.23	.81	.14	.11	-.19
I was worried about using this system. (R)	.18	.22	.86	.12	.12	-.06
I am not satisfied with this system. (R)	.19	.21	.86	.02	.09	-.15
Understandability						
I was able to predict the action of the system during the task.	.09	.22	.07	.83	.11	-.19
I think nobody can predict this system. (R)	.18	.14	.08	.65	.22	-.02
I had no idea what this system would do. (R)	.20	.23	-.14	.66	.36	-.10
Discomfort						
This system behaved smoothly.	-.11	.19	-.14	.23	.78	-.08
This system behaved naturally.	.10	.26	.10	.22	.80	-.19
This system behaved awkwardly. (R)	.33	.21	-.13	.32	.66	-.18
Motivation						
It is more fun to perform the task by myself than relegating the task to the system.	-.07	-.09	-.07	-.04	-.11	.75
Based on the experience of using this system, I feel like performing the task by myself.	-.18	-.12	-.13	-.15	-.06	.87
I can perform the task by myself better without using the system.	-.23	-.12	-.15	-.09	-.14	.79

4 General Discussion

Based on the conducted experiment, the usability questionnaire with 18 questions was developed, and its reliability and validity were confirmed (Appendix A). In the following, we indicate the relationships between the six usability elements effectiveness, efficiency, satisfaction, understandability, discomfort, and motivation.

We developed a structural equation model to investigate the relationships. The model is based on two assumptions; (1) the users’ evaluations of effectiveness, efficiency, discomfort, and understandability directly reflect the automation system behaviors and (2) the users’ evaluations of satisfaction and motivation are influenced by their evaluations of the other four elements. Therefore, we separated the usability elements into high-level elements satisfaction and motivation and low-level elements effectiveness, efficiency, discomfort, and understandability to develop the model.

Based on the answer data for the 18 questions in the questionnaire, we performed covariance structure analysis using the proposed model. First, the model fit test results showed that the proposed model is sufficient to explain the data: CFI (goodness of fit) = .93, TLI (Tucker-Lewis index) = .92, RMSEA (root mean square error of approximation) = .09, and SRMR (standardized root mean square residual) = .05. Second, the analysis results showed interactive influences among the low-level elements: effectiveness and efficiency ($r = .78, p < .001$), effectiveness and discomfort ($r = .66, p < .001$), effectiveness and understandability ($r = .58, p < .001$), efficiency and discomfort ($r = .71, p < .001$), efficiency and understandability ($r = .68, p < .001$), and discomfort and understandability ($r = .70, p < .001$).

The analysis results also showed that the four low-level elements influence one of the high-level elements, satisfaction. The path coefficients indicating the influence of effectiveness, efficiency, discomfort, and understandability on satisfaction were .43 ($p < .001$), .53 ($p < .001$), .13 ($p < .001$), and .09 ($p < .001$), respectively. The effects of discomfort and understandability were very weak, but negative. Therefore, negative influence on users' satisfaction was the result of not only their discomfort but also their low understandability. In this experiment, the participants used the operational pathway-displaying system or the non-displaying system to conduct the line-tracing task in various situations. Thus, when the operational pathway-displaying system was used, in particular, in a situation where the auto-tracing system did not trace the line accurately, users' satisfaction prominently reduced because the displayed pathway and the actual traced pathway were obviously different. Consequently, understandability affects satisfaction negatively.

Moreover, the analysis results showed that the low-level elements effectiveness, efficiency, and understandability affect the other high-level element, motivation. The path coefficients indicating the influence of effectiveness, efficiency, and understandability on motivation were .35 ($p < .001$), .22 ($p < .001$), and .06 ($p < .005$), respectively. Discomfort did not affect motivation. Higher effectiveness, efficiency, and understandability of the system reduced users' motivation to conduct tasks by themselves. Additionally, there was no correlation between satisfaction and motivation ($r = -.01, p = .20$).

In summary, we considered automation systems as third-generation artifacts and developed a generalized usability questionnaire for such systems. This questionnaire can be used to evaluate various types of automation systems and is useful for developing and improving automation systems as daily used artifacts.

A Appendix: Usability Questionnaire for Automation Systems

The order of the questions can be randomized, and (R) indicates reverse scored questions.

1. It is more fun to perform the task by myself than relegating the task to the system.

2. I think nobody can predict this system. (R)
3. I was worried about using this system. (R)
4. This system behaved smoothly.
5. I think that someone's support is necessary to use this system. (R)
6. Based on the experience of using this system, I feel like performing the task by myself.
7. I was able to perform the task accurately by using this system.
8. This system behaved naturally.
9. I am satisfied with this system.
10. This system prevented me from committing errors during the task.
11. I was able to get used to this system in a short amount of time.
12. This system behaved awkwardly. (R)
13. I can perform the task by myself better without using the system.
14. I was able to predict the action of the system during the task.
15. I am not satisfied with this system. (R)
16. I think that I would not make any mistakes by using this system.
17. By using this system, performing the tasks became more difficult. (R)
18. I had no idea what this system would do. (R)

References

1. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse. Abuse. *Hum. Factors* **39**, 230–253 (1997)
2. Rasumussen, J.: *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Elsevier Science Publishing, New York (1996)
3. Norman, D.A.: Cognitive artifacts. In: Carroll, M. (ed.) *Designing Interaction*, pp. 17–38. Cambridge University Press, Cambridge (1991)
4. Norman, D.A.: *The Psychology of Everyday Things*. Basic Books, New York (1988)
5. ISO: Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: Guidance on usability. ISO 9241-11:1998 (1998)
6. Hornbæk, K.: Current practice in measuring usability: challenges to usability studies and research. *Int. J. Hum. Comput. Stud.* **64**, 79–102 (2006)
7. Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) *Usability Evaluation in Industry*. Taylor and Francois, London (1996)
8. Kirakowski, J., Corbett, M.: SUMI: the software usability measurement inventory. *Br. J. Educ. Technol.* **24**, 210–212 (1993)
9. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument for measuring user satisfaction of the human-computer interface. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. pp. 213–218. ACM Press, New York (1988)
10. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **46**, 50–80 (2004)
11. Sarter, N.B., Mumaw, R.D., Wickens, C.D.: Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. *Hum. Factors* **49**, 347–357 (2007)
12. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* **52**, 381–410 (2010)

13. Miwa, K., Terai, H.: Theoretical investigation on disuse atrophy resulting from computer support for cognitive tasks. In: Harris, D. (ed.) EPCE 2014. LNCS, vol. 8532, pp. 244–254. Springer, Heidelberg (2014)
14. Maehigashi, A., Miwa, K., Terai, H., Kojima, K., Morita, J.: Experimental investigation of calibration and resolution in human-automation system interaction. *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* **E96–A**, 1625–1636 (2013)