International Journal of Service and Knowledge Management
International Institute of Applied Informatics
2025, Vol. 9, No. 1, IJSKM864

Evaluation of a Method to Support the Design and Evaluation of Rules that Take into Account the Mechanism of Functional Performance

Naoki Okamoto [*], Yoshiko Ohno [†],
Fumihito Oura [†], Seiko Shirasaka [†]

Abstract

In recent years, there has been a need to take preventive measures against rule violations and rules becoming a dead letter, which is a major factor in organizational accidents and scandals. As one of the measures, a method has been proposed to support the design and evaluation of rules that take into account the mechanism of functional performance of rules. However, an evaluation including understandability and usability of the method has not yet been conducted. Therefore, the purpose of this study is to evaluate a method to support the design and evaluation of rules that take into account the mechanism of functional performance including the perspectives of understandability and usability. To evaluate the method, we conducted user evaluations using questionnaires on the method and the results of the method, experiments to compare the method with another method, and third-party evaluations by the regulated persons, and identified several problems with understandability and usability. We explain our results and conclusions of this study and our future research topics, including problems of the method, in the last part of the paper.

Keywords: Organizational Rules, Prevention of Rule Becoming a Dead Letter, Risk Assessment, Rule Design, Rule Management

1 Introduction

According to Baldwin (1990) [1] , there are a large number of rules that are regularly ignored or disobeyed in virtually all fields of regulation and administration. And O'Dea & Flin (2001) [2] state that the failure to follow rules are the third most important perceived cause of accidents.

In response to organizational accidents and scandals caused by such rule violations, Taniguchi (2018) [3] states that in order to prevent organizational accidents and scandals, it is important to establish appropriate rules and ensure that the established rules are followed. In addition,

[*] The System Design and Management Research Institute, Graduate School of System Design and Management, Keio University
[†] Graduate School of System Design and Management, Keio University

Taniguchi (2017) [4] also discusses the usefulness of preventing rules from becoming a dead letter in order to prevent organizational accidents and scandals.

Thus, in the field of organizational accident and scandal research, there is a common recognition of the importance of safety management through rules (hereafter, rule management) [5]. Yoshino & Saito (2018) [6] point out the challenges of this rule management, especially with regard to the methodology of rule design, which is a "feed-forward approach to design" in which all problems related to the rule are identified and designed in advance. He states that not only a "feed-forward design approach", in which all problems related to the rule are identified and designed in advance, but also a "practical approach", in which the regulator designs the rule sequentially while evaluating various practices in the field, is necessary. In other words, rule management to prevent organizational accidents and scandals requires a rule design method that takes into account "setting appropriate rules," "ensuring that the established rules are followed," and "evaluating field practices to prevent the rules from becoming a dead letter.

Based on the problem in the previous studies by Hale & Borys (2013) [7], and Takahashi et al. (2021) [8], Okamoto et al. (2023) [9] propose a method to support rule design and evaluation that takes into account the mechanism of rule functional performance in order to prevent organizational accidents and scandals by defining the state in which a rule performs its function as "the rule is appropriate for purpose," "the established rule is followed," and "the rule is not becoming a dead letter.

As a methodology for designing and evaluating rules that takes into account the mechanisms of functional performance of rules, Hale & Borys (2013) [7] propose a rule management framework, which has a cyclical structure, emphasizing the fact that rule management is a dynamic process of adapting rules to the changing realities of the activity and its environment. This framework has the structure of a PDCA cycle in which rules are repeatedly evaluated and designed, with an emphasis on the monitoring and learning part of the loop. While this is indeed a methodology for designing and evaluating rules that takes into account the mechanism of functional performance of rules, it does not describe a specific method for designing and evaluating rules.

A method that supports the design and evaluation of specific rules is the risk assessment method for rules regarding intentional rule violations by Takahashi et al. (2021) [8]. They organized "direct causes of intentional rule violations" and "factors behind intentional rule violations" and developed a checklist for rule violation prevention based on these factors. However, they note that this method is only intended for rules in the chemical industry, and its application to rules in other industries and empirical studies are problems to be addressed. In addition, this method does not have a PDCA cycle structure like the framework of Hale & Borys (2013) [7], and it is difficult to say that it can take into account the mechanism of functional performance of rules.

Based on these problems of previous research, Okamoto et al. (2023) [9] proposed specific methods to support the design and evaluation of rules that take into account the mechanism of functional performance of rules, and we empirically evaluated these methods.

However, there were several problems with Okamoto's (2023) study. First, in Step2, they didn’t take into account the risk of rules being broken due to individual circumstances. According to Becker (1968) [10], some people become criminals not because their basic motivation differs from that of other people, but because their benefits and costs differ. This is based on the economic principle that individuals make rational decisions to maximize their self-interest, suggesting that the decision to violate rules depends on a rational calculation of costs and benefits in a given situation. For example, someone who normally obeys traffic signals may break the rule if the situation is extremely urgent. Therefore, this research adds and evaluates the 'risk dependent

on the situation for adhering to rules,' which Okamoto et al. (2023) [9] did not consider. Also, Okamoto et al. (2023) [9] noted that the methodology needs to be improved to increase usability, as users have pointed out the amount of work and complexity of the method. However, there has been no evaluation of the understandability and usability of the method. Therefore, the purpose of this study is to evaluate the method proposed by Okamoto et al. (2023) [9], including the perspectives of understandability and usability. The novelty of this study is to evaluate the understandability and usability of the method.

This paper is organized as follows. We mention the method of the previous study and the risks added in this study in section 2. We show evaluation method in section 3, and the evaluation results in section 4. We explain the discussion in section 5, and the conclusion and future research topics in section 6.

2 Method of Previous Study and Risks Added in This Study

In this section, we explain the method of the previous study by Okamoto et al. (2023) [9] and the risks added in this study.

The method is intended for "persons in a position to design and operate rules in an organization (hereafter referred to as "regulators")," which are the entities that execute rule management, and is primarily to assist in the design and evaluation of rules when they are being designed or revised. To embody these requirements, They have defined three steps based on Bax et al. (1998) [11] point that "analysis of the regulatory regime is essential to understand how workers assess the legitimacy of rules," and Amalberti et al. (2006) [12] that "risk should be managed rather than eliminated so that rules are not followed".

In Step 1, the user analyzes the regulatory regime of the subject rule. In Step 2, the user assesses the risk of the rule not being appropriate, not being followed, or becoming a dead letter. In Step 3, the user considers measures to address the risks. To implement each step, they have developed the following tools: "Confirmation Sheet for functional performance of rules," "Risk Assessment Sheet," and "Countermeasure Examination Sheet". In this study, we added the risk of rules being broken due to individual circumstances to the risks listed on the "Risk Assessment Sheet" used in Step2.

In Step 1, the user analyzes the mechanism by which the subject rule functions using the "Confirmation Sheet for Mechanism of Functioning of Rules" (Figure 1). This sheet defines the components and the relationships between the components for a rule to function. The components of the sheet are "A: Regulators", "B: Standard setting", "C: Sanctions", "D: Monitoring", "E: Enforcement", "F: Regulated Persons", "G: Environment for action", "H: Person in charge of disseminating", "I: Other rules", and "J: Monitoring Change".

First, component A is the regulator, which is the user of this method.

Components B, C, D, and E, enclosed by "Rule," indicate what is required for the rule to work, and are defined based on the three elements ("Standard Setting," "Monitoring and Enforcement," and "Sanctions") that Scott [13] has organized to make a regulation functional. Sanctions reduce the temptation to violate the regulated person, Monitoring ensures that the regulated person is following its rules, and Enforcement directs behavior to conform to the standards setting for the regulated person.

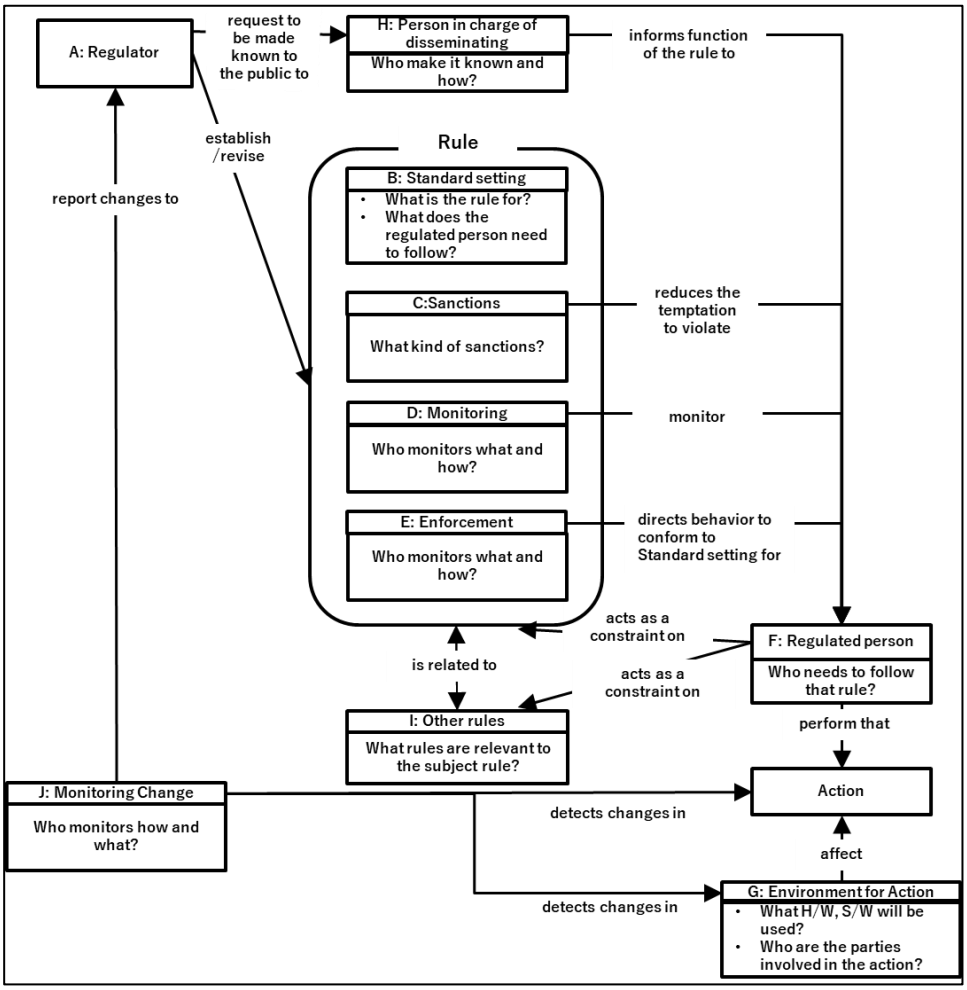

Figure 1: Confirmation Sheet for functional performance of rules

Component F means the subject that the subject rules are intended to regulate. The regulated person acts as a constraint on the subject rules and other relevant rules [14].

Component G is the environment for action, which is the hardware and software used to act to follow the rules and the personnel involved in the action by the regulated person. The reason for the environment of action component is that, as Schulz [15] points out, the rules of an organization become a dead letter over time, which increases with the degree of change in the environment. As for what exactly the environment for action, it is hardware and software, which are one of the factors for rule violations organized by Alper & Karsh [16], and those whom the rules do not directly require to comply, but who are involved in the actions of the regulated party. The environment in which the action is to be taken affects the action because the content and methods of the action may change if the environment of the action changes. (For example, a change in the internal customer information management system will also change the action of how customer information is handled.)

Component H is who and how the rules are disseminated and has the role of informing regulated persons of the function of the rules. The reason this component is necessary is that if the rules are not made known to the regulated, they will not be followed forever. Taniguchi [4] points out that disseminating the function of a rule to the field may be more important than tightening the monitoring regime to prevent the rule becoming a dead letter,

and for a rule to function, it must be made known, including education about why the rule is necessary and the value of following it.

Component I is the rule that represents a rule of a higher or lower hierarchy of the subject rule or a rule that is incidental to the content that the rule prescribes, and these rules are closely related to the subject rule. This component is necessary because the research of Alper & Karsh [16] has revealed the problem of inconsistency among rules and lack of priority as a factor that causes rules to be broken, and for rules to function, the related rules must be related without inconsistency and with clear prioritization.

Component J is who and how the actions and the environment of actions are monitored, and has the role of monitoring the actions and the environment for actions and reporting changes, if any, to the regulator. Alper & Karsh [16] point out that 'rules fail because systems, both internal and external to an organization, are in a constant state of change,' and furthermore, according to Baldwin [17], the lack of a mechanism for detecting change is a major problem to regulate. Therefore, for regulations to work, they need to address changes in both the action (internal) and the action's environment (external).

Next, in Step 2, the risk of the target rule not functioning is assessed using the "Risk Assessment Sheet" with reference to the results of Step 1. Specifically, user assess the risks of the rules not being appropriate, the risks of the rules not being followed, and the risks of the rules becoming a dead letter.

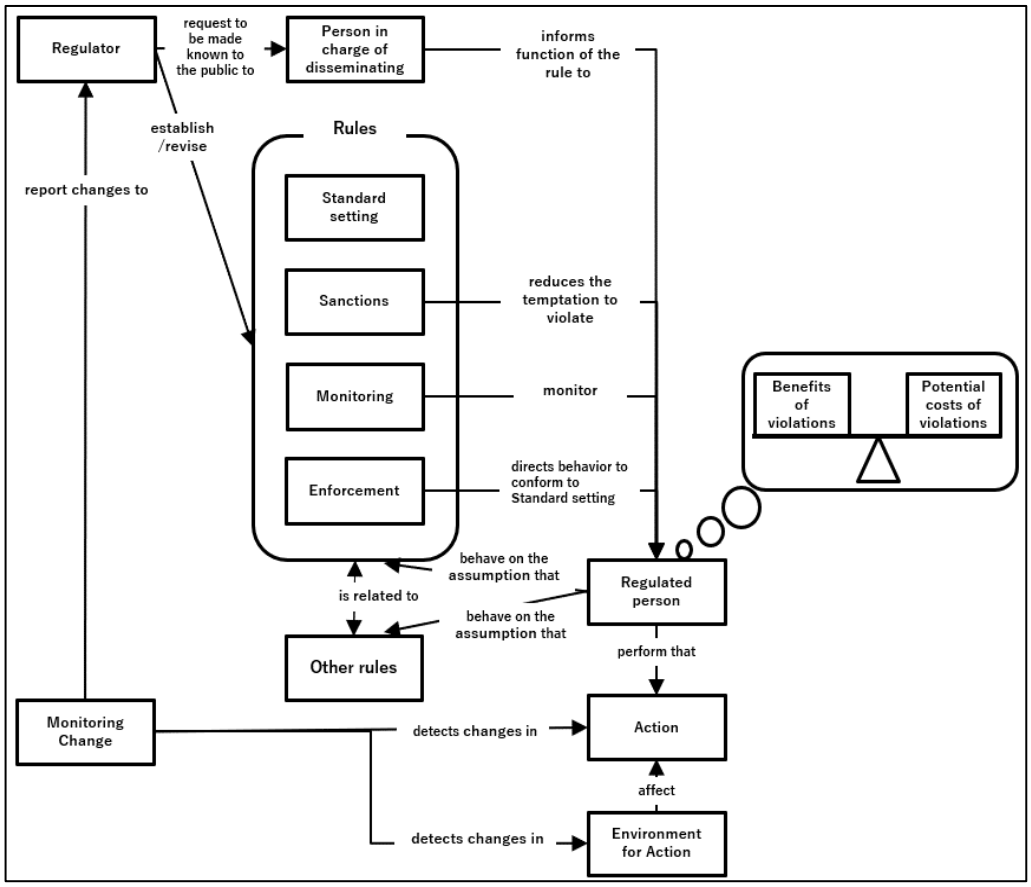

Figure 2: The space in which the “Risk of the rules not functioning” exists

Figure2 shows the space in which the “risk of rules not functioning” exists in this study. The “risk of the rules not functioning” exists in each component and between components for rules to function, as well as in the perception of the regulated person. Specifically, the risks include in previous studies by Schulz (1998) [15], Hale & Borys (2013) [7], Takagi et al. (2011) [14], Alper & Karshet (2009) [16], Takahashi et al. (2021) [8], and Okamoto et al. (2023) [9]. In addition, in this study, from an economic approach, “risks related to the benefits of violating the rule”, and “risk related to the potential costs of violating the rule” were added as ”the risk of rules being broken due to individual circumstances”. The “Risk Assessment Sheet” lists those risks.

Table 1 is an example of how to fill out the "Risk Assessment Sheet" on the subject of rule for appropriate reporting of entertainment expenses. The user should circle the risks listed that are or may be applicable and describe the rationale.

Table 1: Example of how to fill out the Risk Assessment Sheet (excerpt)

No.	Risk Description	Assessment Results	Rationale for assessment
1-1	• The purpose of the rule is not appropriate.		
1-1	• Compliance with the rules does not achieve the purpose.		
1-2	• Sanction is not defined.		
1-3	• Sanctions have not reduced the temptation to violate.	○	A caution is not enough, and the next time it's forgotten again, it may not be followed.
1-4	• Monitoring is not defined.		
1-5	• Monitoring contents and methods are not appropriate.	○	We are not able to monitor all actions to ensure they are followed.
1-6	• The person charged with monitoring does not perform the monitoring as defined.		
1-7	・Enforcement is not defined.		
1-8	・Enforcement methods are not appropriate		
⋮	⋮	⋮	⋮

Table 2: Example of how to fill out the Countermeasure Examination Sheet (excerpt)

Policy on Response Measures	Points for consideration	Risk No.	Countermeasure
Review the purpose of the rule	• What objectives should be set to prevent accidents and scandals in the organization?	1	
Review the content of the rule to ensure that it achieves the purpose.	• What modifications to the content would accomplish the purpose of the rule?	2	
Adapts to changes in the action environment	• What modifications can be made to the rule to adapt the changes in the environment for action that are already occurred?	15,27	
Define or review sanctions	• What rewards or penalties would reduce the temptation to violate?	3,4,34,35	Clarify penalties for violating rules and regulations, and reflect them in personnel evaluations.
Define or review monitoring	• What should we monitor to determine if rules are being followed? • Who should monitor compliance with regulations? • How should we monitor to determine if the rules are being followed?	5,6,7	All forms are monitored, not just those for which approval is requested.
Define or review enforcement	• Who should impose sanctions? • Is it possible to move the regulated person's actions toward compliance through an architecture?	8,9,10	
⋮	⋮	⋮	⋮

Finally, in Step 3, countermeasures for the applicable risks are examined using the Countermeasure Examination Sheet. Specifically, for those circled in the Step 2 assessment, countermeasures are examined on the Countermeasure Examination Sheet. This sheet presents the response policy and items to be considered in advance, and the user examines the countermeasures in accordance with the policy and items. Table 2 shows an example of the "Response Plan Examination Sheet" based on the rule for appropriate reporting of entertainment expenses.

3 Evaluation Method

To evaluate the method, Okamoto et al. (2023) [9]conducted user evaluations using questionnaires on the method and the results of the method, experiments to compare the method with another method, and third-party evaluations of the results by the regulated parties to confirm the effectiveness and appropriateness of the method. In this study, we conducted objective and subjective evaluation of the understandability and usability of the method.

Okamoto et al. (2023) [9] conducted objective and subjective evaluations of the effectiveness of the method and the deliverables of the method, respectively, and for the subjective evaluation, they conducted a user evaluation using a questionnaire. For the objective evaluation, Okamoto et al. (2023) [9] conducted a comparison experiment with the existing method by Takahashi et al. [8] and an evaluation of the artifacts by the regulated parties based on the factors that contribute to making the rules a dead letter in the workplace by Takagi et al. (2011) [14].

Table 3 shows the evaluation methods and perspectives of the evaluation in this study. In this study, we evaluated the understandability and usability of the method using an evaluation method similar to that of Okamoto et al. (2023) [9]. For the comparison experiment, there were two groups, as shown in Figure 3, and the objective differences between the method used by group A and another method used by group B, and the subjective differences in the within-subject comparison between another method used by group B in the first experiment and the method used in the second experiment were targeted.

Table 3: Evaluation methods and perspectives of the evaluation in this study

Subjective evaluation	Objective evaluation
・ Users evaluate the understandability and usability of the method.	・ We conduct experiments to compare the understandability and usability of method with another methods.

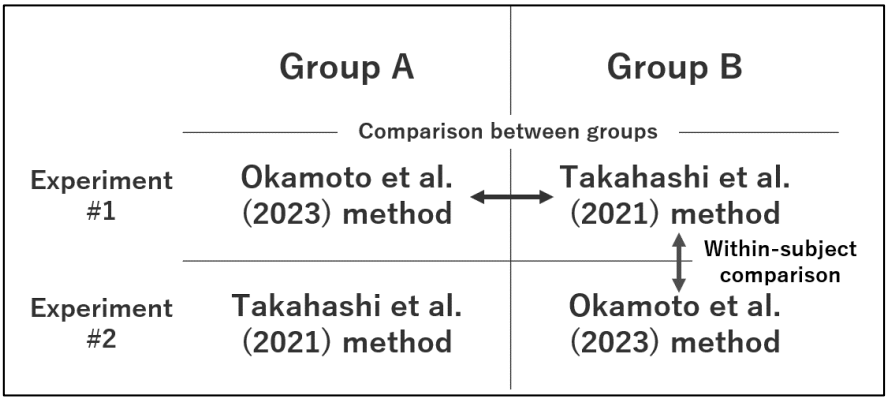

Figure 3: Overview of comparison experiment

4 Evaluation Results

First, the evaluation results shown in Okamoto et al. (2023) [9] are detailed below.

Twelve subjects from four organizations applied the method to a total of seven rules. These rules were not limited to safety rules, but also included "rules concerning behavior or conduct that require the regulated person to be obligated or prohibited," and the rule governing the company's approval of business activities and rule concerning secondary employment were included in the evaluation. Subjects were then asked to evaluate the effectiveness of the methods and the artifacts of each method. In addition, several to several dozen regulated persons conducted third-party evaluations of the deliverables based on the rule formulation factors by Takagi et al. [14]. Based on the results of these evaluations, they confirmed the validity of the method for the purpose of this study and verified whether the content implemented with the proposed method was able to realize the requirements of the method.

Table 4: Results of questionnaire to subjects on the effectiveness of the method

No.	Questionnaire items	1	2	3	4	5	Average point
1	Do you feel that this method is effective in analyzing and examining whether the purpose of the rule is appropriate?	0	3	3	2	3	3.45
2	Do you feel that this method is effective in analyzing and examining whether the content of the rule is appropriate for the purpose of the rule?	2	1	2	4	2	3.27
3	Do you feel that this method is effective in making the rules to function?	1	3	3	2	2	3.09
4	Do you feel that this method is effective in modifying the rules as the action changes or as the environment for the action changes?	1	1	4	2	3	3.45
5	Do you feel that this method is effective in examining the regime that the regulations have in place to deal with changes in actions and in the environment in which actions take place?	1	2	5	2	1	3.00

Table 5: Results of paired t-tests in comparative experiments

Questionnaire items	T	df	p value (two-tailed test)
Effectiveness in analyzing whether the purpose of the rule is appropriate	3.993	10	0.003
Effectiveness in analyzing whether the content of the rule is appropriate in light of the purpose of the rule	2.301	10	0.044

Regarding the results of the subjective evaluation, in which subjects rated the effectiveness of the proposed method on a five-point ranking scale ("not very much = 1," "not much = 2," "neither = 3," "much = 4," and "very much = 5"), which are shown in Table 4. The average point for all survey items was at least 3 points. ("*" mark indicates a reverse question)

Regarding the results of the objective evaluation, they confirmed the effectiveness of the method and whether the method was able to realize the requirements of the method through comparison experiment. In the group comparison (t-test without assuming equal variance), the method had higher average points than another method in terms of "establishing appropriate rules," and significant differences were found. For "effectiveness in analyzing whether the purpose of the rule is appropriate," a statistically significant trend was confirmed for the mean, $p = 0.003$, as shown in Table 5. A statistically significant trend was also observed for "effectiveness in analyzing whether the content of the rule is appropriate in light of the purpose of the rule," a statistically significant trend was confirmed for the mean, $p = 0.044$, as shown in Table 5. On the other hand, no significant differences were found for the other items in the between-group and within-subject comparisons of the comparison experiments. In the comparison experiment, they also conducted a subjective comparison using a questionnaire, and a total of 9 responses were obtained. The results showed that the majority of the respondents agreed that the proposed method was effective in terms of " the rule does not become a dead letter," which was one of the goals of the proposed method. On the other hand, the existing method was rated higher in terms of usability.

Regarding the results of the third-party evaluation, Table 6 shows the results reflecting the evaluation of the method and the current rules (t-test without assuming equal variances).

There is a positive statistically significant difference for the proposed method with $p = 0.022$ for the item "I feel that if you break this rule, someone will notice right away" as a "mechanism to prevent rules from being broken". On the other hand, there is a negative statistically significant difference for the proposed method with $p = 0.007$ and $p = 0.030$ for the items "I feel that the rule is changed frequently" and "I feel that the rule is complicated and has many exceptions and examples" as "reliability of the rule".

The reason why the degrees of freedom in Table 6 vary from item to item is that the number of third-party evaluators for each rule is different, and only results with discussion points are listed.

Table 6: Results of paired t-tests in third-party evaluations

Questionnaire items	T	df	p value (two-tailed test)
A mechanism to prevent rules from being broken (I feel that if you break this rule, someone will notice right away)	-3.91	31	0.022
Reliability of the rule (I feel they change the rules too often)	-2.880	36	0.007
Reliability of the rule (I feel that the rule is complex, with many exceptions and case divisions)	-2.260	36	0.030

Next, we show the results of the evaluation in this study.

Twelve subjects from four organizations applied the proposed method to a total of seven rules. In this experiment, the scope was also not limited to safety rules, but also included "rules concerning behavior or conduct that require the regulated person to be obligated or prohibited," and the rule on the company's approval of business activities, and the rule on secondary employment. Subjects were then asked to evaluate the understandability and usability of the method and another method.

The results of the questionnaire based on the subjectivity of the subjects are shown in Tables 7 and 8. A total of 11 responses were obtained, and the comprehensibility was rated on a 5-point ranking scale ("not very much = 1," "not much = 2," "neither = 3," "much = 4," and " very much = 5"). The results, shown in Table 7, indicate that all survey items received an average rating of at least 3 points.

Usability was rated on a 5-point scale ("not very much = 1," "not much = 2," "neither = 3," "much = 4," and "very much = 5"). The results are shown in Table 8, and again, all survey items received an average rating of 3 points or more.

As for the negative evaluations that received less than 3 points, based on the free descriptions in the questionnaire, they were mainly related to inadequate language supplementation, lack of guidance in each step, and lack of guidance in the consideration of measures to be taken.

Table 7: Results of a questionnaire administered to the subjects on the understandability of the method

No.	Questionnaire items	1	2	3	4	5	Average point
1	Please rate your understanding of Step 1 on a scale of 1 to 5.	0	2	3	3	3	3.64
2	Please rate your understanding of Step 2 on a scale of 1 to 5.	0	2	4	3	2	3.45
3	Please rate your understanding of Step 3 on a scale of 1 to 5.	0	1	3	3	4	3.91
4	Please rate your understanding of this method as a whole on a 5-point scale.	0	1	1	7	2	3.91

Table 8: Results of a questionnaire administered to the subjects on the usability of the method

No.	Questionnaire items	1	2	3	4	5	Average point
1	Do you feel that you would like to use this method in the future?	1	0	1	7	2	3.82
2	Do you feel you would recommend this method to a colleague?	1	0	4	4	2	3.55

We explain the results of the comparative experiments. Six questionnaires were obtained for the method of group A, five for the method of group B, and six for another method of group B.

Table 9 shows the statistics for the objective differences between the method in group A and another method in group B. Table 10 shows the results of a t-test that does not assume equal variances. As these results show, there was no statistically significant difference between the method in group A and another method in group B.

Table 11 shows the statistics for the subjective differences in the within-subject comparison of the method used by group B in the first experiment and the method used in the second experiment, and Table 12 shows the results of t-tests that do not assume equal variances. As these results show, there was no statistically significant difference between the method and another method in group B.

Table 9: Results of Statistics on objective differences between group A and group B

Questionnaire items	Method	Frequency	Mean	Standard Deviation	Standard Error of the Mean
(Understandability) Please rate your understanding of this method as a whole on a 5-point scale.	Okamoto et al. (2023) [9]	6	4.00	0.000	0.000
	Takahashi et al. (2023) [8]	6	3.83	0.753	0.307
(Usability) Do you feel that you would like to use this method in the future?	Okamoto et al. (2023) [9]	6	4.00	0.632	0.258
	Takahashi et al. (2023) [8]	6	3.50	1.049	0.428
(Usability) Do you feel you would recommend this method to a colleague?	Okamoto et al. (2023) [9]	6	3.67	0.816	0.333
	Takahashi et al. (2023) [8]	6	2.83	1.329	0.543

Table 10: Results of paired t-tests for objective differences between Group A and Group B

Questionnaire items	T	df	p value (two-tailed test)
(Understandability) Please rate your understanding of this method as a whole on a 5-point scale.	0.542	5.000	0.611
(Usability) Do you feel that you would like to use this method in the future?	1.000	8.212	0.346
(Usability) Do you feel you would recommend this method to a colleague?	1.309	8.303	0.226

Table 11: Results of Statistics on subjective differences in within-subject comparisons

Questionnaire items	Method	Frequency	Mean	Standard Deviation	Standard Error of the Mean
(Understandability) Please rate your understanding of this method as a whole on a 5-point scale.	Okamoto et al. (2023) [9]	5	3.80	1.304	0.583
	Takahashi et al. (2023) [8]	6	3.83	0.753	0.307
(Usability) Do you feel that you would like to use this method in the future?	Okamoto et al. (2023) [9]	5	3.60	1.517	0.678
	Takahashi et al. (2023) [8]	6	3.50	1.049	0.428
(Usability) Do you feel you would recommend this method to a colleague?	Okamoto et al. (2023) [9]	5	3.40	1.517	0.678
	Takahashi et al. (2023) [8]	6	2.83	1.329	0.543

Table 12: Results of paired t-tests for subjective differences for subjects in Group B

Questionnaire items	T	df	p value (two-tailed test)
(Understandability) Please rate your understanding of this method as a whole on a 5-point scale.	-0.051	6.151	0.961
(Usability) Do you feel that you would like to use this method in the future?	0.125	6.942	0.904
(Usability) Do you feel you would recommend this method to a colleague?	0.652	8.104	0.532

5 Discussions

In this section, we outline considerations and limitations derived from the previous and the present studies.

5.1 Considerations

First, we discuss the findings from Okamoto et al. (2023) [9].

Based on the evaluation results, they explained the effectiveness of the proposed method and its comparative advantage over existing methods in terms of whether the rules are designed and

evaluated in consideration of the three aspects of the purpose of this study: the rule is appropriate for the purpose, the established rule is followed, and the rule does not become a dead letter.

In terms of "the rule is appropriate for the purpose," the subjective ratings were as shown in Table 4, No. 1 and No. 2, with No. 1 and No. 2 receiving 3.45 and 3.27 points, respectively, indicating a positive opinion. On the other hand, one respondent commented, "I could consider whether the rules were appropriate for the purpose, but I felt that I could not determine whether the purpose was appropriate." This comment suggests that one of the limitations of this method is that it may not be able to assist in judging whether the purpose of the rules is correct or not. The objective evaluation is shown in Table 5, suggesting a comparative advantage over another method. In addition, the results of the third-party evaluation differed depending on the rule. This suggests that the results of this method may vary depending on the type and characteristics of the rule.

In terms of "the established rule is followed," the subjective ratings were as shown in Table 4, No. 3, receiving 3.09 points, indicating a slight but positive result. Based on the user's comments, it was suggested that the proposed method has various perspectives to make the rule be followed and that it is now possible to examine it from multiple perspectives. On the other hand, it was suggested that the amount of work and complexity of the work is a problem. In addition, considering the result of the decrease in "reliability of regulations" in the objective evaluation, the method asked respondents to consider and evaluate countermeasures for all applicable risks one by one, but since not all countermeasures will be adopted when regulations are actually enacted or revised, the method allows the selection of several countermeasures. Therefore, it is suggested that the method does not take into account the step of determining a balanced response by selecting several response measures and considering the impact of the regulation.

In terms of "the rule does not become a dead letter," the subjective ratings were as shown in Table 4, No. 4 and No. 5, with No. 4 and No. 5 receiving 3.45 and 3.00 points, respectively, indicating a positive opinion. Users commented that the monitoring of changes was easy to understand because it was set up in a positive way, and they felt that the background behind the rule becoming a dead letter was not taken into account. This suggests that the method is effective in modifying rules in accordance with changes in operations and the business environment, but does not cover modification of rules based on past background. The objective evaluation showed no significant differences from another method. However, when comparing the subjective evaluations of the users, most of them said that the method was more effective, suggesting its comparative advantage over another method. In addition, they did not confirm whether the proposed measures would really prevent the rules from becoming obsolete. Therefore, one of the limitations of this study is that it does not support whether or not the proposed measures can deal with the subsequent deformity and whether or not the occurrence of deformity can be noticed.

We discuss the findings of this study.

First, in the subjects' questionnaire, assuming that 4 points or more on a 5-point ranking scale is a positive evaluation, the comprehensibility and usability of each step and the method as a whole averaged less than 4 points for all items, which indicates that there are problems with the comprehensibility and usability of the method. In terms of the specifics of the problem, the comments of those who rated the method less than 3 points indicated several points regarding the method, and we will discuss them in the following sections.

Regarding Step 1, for example, "I didn't know whether to do it in order or not. I didn't know what the procedure manual was referring to, and I wasn't sure if what I was trying to do was the right thing." and "I was not sure how valid the output would be and how convincing they would be because I could not see the shades of gray between what should be done in depth as information and what should not be done in depth.". The comments also indicated that "the information was not clear where information should be deepened and where should not. These comments suggest that there is a problem of unclear procedures and the setting the level of abstraction of the final deliverables in the work.

Regarding Step 2, comments included, " I couldn't work with the assumption that the rules would stop working." and "I couldn't think deeply because of new words, stiff expressions, and abstract expressions.". This suggested that the difficulty in understanding the design of the connections between each step and the words used was a problem.

Regarding Step 3, there was a comment that "I couldn't work with the idea of keeping the rules functioning.". It was suggested that this comment could be further improved if there was some guide to facilitate understanding of the steps, such as having the participants review the purpose of Step 3 before considering measures to be taken.

Next, in the comparison experiment, no significant difference at the 1% or 5% level for either the objective differences between Group A and Group B or the subjective differences among Group B subjects. On the other hand, when we asked subjects to comment on the comparison between the method and another method, they replied, "Another method was simpler and required less knowledge to understand," and "Another method was easier in terms of work process and time. This suggests that there are problems in both understandability and usability.

5.2 Limitations

The limitations of the study by Okamoto et al. (2023) [9] are as follows.

First, it is necessary to add an item to assist in determining whether the purpose of the rule is correct. This is because if the regulated person is not convinced of the purpose of the rule, the rule may not be followed.

Second, since the results may differ depending on the type and characteristics of the rule, it is necessary to have the method used by regulators in various organizations and industries to confirm its effectiveness as a widely used method.

On the other hand, since the evaluation results changed depending on the rules, it is necessary to consider whether there is merit in developing design and evaluation methods for each type of industry, such as agriculture, industry, information technology industry, and service industry, based on the characteristics of the rules.

It is also necessary to examine what kind of regulatory method would be appropriate, since the results are likely to change depending on the regulatory method, such as mandatory versus prohibited rules.

Next, since it was pointed out that the amount of work and complexity involved in the method needs to be improved to increase the usability of the method depending on user's purpose by reducing the amount of work involved in the method and making the operation of the method more efficient.

In addition, in order not to reduce the reliability of the rule, it is considered necessary to add a step to assist in determining a balanced response strategy, taking into account the impact of the

regulation after the response strategy has been considered.

Also, since rules may be broken by human will in the first place, proposals for regulatory methods combined with other forms of regulation, such as regulation by architecture that regulates without notice, may be considered.

Additionally, since this study has not confirmed whether the proposed measures can detect changes, it is necessary to confirm whether the measures considered in the method can really detect changes in the behavioral environment and changes in behavior.

Finally, they did not confirm whether the method can maintain the state in which the rule does not become a dead letter. Therefore, it is necessary to confirm whether or not the method can maintain the state where rule becoming a dead letter does not occur by means of empirical experiments.

We show the limitations of this study.

First, it is necessary to make the method more understandable to the user. This is because users are expected to use the method on their own and must implement the method without support.

Second, it is necessary to improve the clarity of the connection between each step and to make the words used easier to understand so that users do not forget the deliverables and objectives of the method.

Third, it is necessary to improve the method to reduce the working time in order to increase its usability. This is because users manage many rules, and the longer the working time, the less practical they become to use.

Fourth, the effectiveness of the method needs to be confirmed again because "the risk of rules being broken due to individual circumstances" was added in the sheet used in Step 2 of this study.

Finally, because of the small number of subjects in this study, it is necessary to apply the method to a larger number of users and evaluate its understandability and usability.

6 Conclusion and Future Research Topics

The purpose of this study was to evaluate a method to support the design and evaluation of rules that take into account the mechanism of functional performance including the perspectives of understandability and usability. The results of the evaluation showed that there were no statistically significant differences from another method, but suggested that there are problems with the understandability and usability of the method. Finally, we explain future research topics related to the method.

Okamoto et al. (2023) [9] identified the following research topics.

- Add an item to assist in determining whether the purpose of the rule is correct.
- Have the proposed method used by regulators in various organizations and industries to confirm its effectiveness as a widely used method.
- Examine that methods appropriate to the type and characteristics of the rule are effective.
- Improve to increase the usability of this method by reducing the amount of work involved in this method and making the operation of this method more efficient.
- Add a step to assist in determining a balanced response strategy, taking into account the impact of the regulation after the response strategy.
- Propose regulatory methods combined with other forms of regulation, such as regulation by architecture that regulates without notice.
- Confirm through empirical experiments whether the measures considered in the

proposed method can really detect changes in the action environment and changes in actions and help prevent rules becoming a dead letter.

In addition, future research topics identified in this study include,

- Improve the method and its description so that users can understand the method more clearly.
- Improve the method by clarifying the connection between each step and using words that are easy to understand.
- Reduce working time by efficiency improvement of methods, training for users, and other efforts.
- Confirm the effectiveness of the method with the addition of "the risk of rules being broken due to individual circumstances".
- Apply the method to a larger number of users and evaluate its understandability, usability.

References

[1] R. Baldwin, "Why rules don't work," *Mod. L. Rev.*, vol. 53, p. 321, 1990.

[2] A. O'Dea and R. Flin, "Site managers and safety leadership in the offshore oil and gas industry," *Safety Science*, vol. 37, no. 1, pp. 39–57, Feb. 2001, doi: 10.1016/S0925-7535(00)00049-7.

[3] Taniguchi, Y., "Exploring the process of rule violating behavior," Economic Studies, vol. 68, no. 1, pp. 25-33, 2018.

[4] Taniguchi, Y., "The process of rule becoming a dead letter: an examination based on the triangle theory of injustice," Economic Studies, vol. 67, no. 1, pp. 5-13, 2017.

[5] J. Weichbrodt, "Safety rules as instruments for organizational control, coordination and knowledge: Implications for rules management," *Safety Science*, vol. 80, pp. 221–232, Dec. 2015, doi: 10.1016/j.ssci.2015.07.031.

[6] Naoto Yoshino and Yasushi Saito, "Rule Management in High-Risk Organizations: Reconsidering Research Agendas in Safety Management Studies," Organization Science, vol. 51, no. 3, pp. 19-30, 2018.

[7] A. Hale and D. Borys, "Working to rule or working safely? Part 2: The management of safety rules and procedures," *Safety science*, vol. 55, pp. 222–231, 2013.

[8] A. Takahashi, G. Shimada, and Y. Sato, "Proposal on the concept and evaluation procedure of human errors in chemicals risk assessment for preventing fire and ex-plosion," National Industrial of Occupational Safety and Health, vol. 14, no. 2, pp. 169 -176, 2021.

[9] N. Okamoto, Y. Ohno, F. Oura, and S. Shirasaka, "Proposal of a Method to Support the Design and Evaluation of Rules that Take into Account the Mechanism of Functional Performance," *IIAI Letters on Business and Decision Science*, vol. 3, 2023.

[10] G. S. Becker, “Crime and Punishment: An Economic Approach,” *Journal of Political Economy*, vol. 76, no. 2, pp. 169–217, Mar. 1968, doi: 10.1086/259394.

[11] E. H. Bax, B. J. Steijn, and M. C. De Witte, “Risk management at the shopfloor: The perception of formal rules in high-risk work situations,” *Journal of Contingencies and Crisis Management*, vol. 6, no. 4, pp. 177–188, 1998, doi: 10.1111/1468-5973.00087.

[12] R. Amalberti, C. Vincent, Y. Auroy, and G. de Saint Maurice, “Violations and migrations in health care: a framework for understanding and management: Figure 1,” *Qual Saf Health Care*, vol. 15, no. suppl 1, pp. i66–i71, Dec. 2006, doi: 10.1136/qshc.2005.015982.

[13] C. Scott, “Analysing regulatory space: fragmented resources and institutional design,” *Public law*, pp. 283–305, 2001.

[14] Miho Takagi, Hirokazu Fukui, and Yuko Matsui, "Factors that contribute to making the rules a dead letter in the workplace," Journal of the Institute of Nuclear Safety System, vol. 18, pp. 2–13, 2011.

[15] M. Schulz, “A model of obsolescence of organizational rules,” *Computational & Mathematical Organization Theory*, vol. 4, no. 3, pp. 241–266, 1998.

[16] S. J. Alper and B.-T. Karsh, “A systematic review of safety violations in industry,” *Accident Analysis & Prevention*, vol. 41, no. 4, pp. 739–754, Jul. 2009, doi: 10.1016/j.aap.2009.03.013.

[17] R. Baldwin and J. Black, “Really responsive regulation,” *The modern law review*, vol. 71, no. 1, pp. 59–94, 2008.