A Human Factor Analysis for Software Reliability in Design-Review Process

SHIGERU YAMADA

Department of Social Systems Engineering, Tottori University
Minami 4-101, Koyama, Tottori-shi, 680-855, Japan

(Received on December 16, 2005)

Abstract: Software faults introduced by human development work have great influence on the quality and reliability of a final software product. The design-review work can improve the final quality of a software product by reviewing the design specifications, and by detecting and correcting a lot of design faults.

In this paper, we conduct an experiment to clarify human factors and their interactions affecting software reliability by assuming a model of human factors which consist of inhibitors and inducers. Finally, extracting the significant human factors by using the quality engineering approach based on the orthogonal array \( L_{18}(2^1 \times 3^7) \) and the signal-to-noise ratio, we discuss the relationships among them and the classification of detected faults, i.e., descriptive-design faults and symbolic-design ones, in the design-review process.

Key Words: software reliability, design-review, human factor, descriptive-design faults, symbolic-design faults, orthogonal-array \( L_{18}(2^1 \times 3^7) \), analysis of variance

1. Introduction

Software faults introduced by human errors in development activities of complicated and diversified software systems have caused a lot of system failures of modern computer systems. Since these faults concern with mutual relations among human factors in such software development projects, it is difficult to prevent from software failures beforehand in the software production control. Additionally, most of these faults are detected and corrected after software failure occurrences during the testing phase.

If we can make the mutual relations among human factors [1,2] clear, then the problem for software reliability improvement is expected to be solved. So far, several studies have been carried out to investigate the relationships among software reliability and human factors by performing software development experiments and providing fundamental frameworks for understanding the mutual relations among various human factors (for example, see [3,4]).

In this paper, we focus on a software design-review process which is more effective than the other processes for elimination and prevention of software faults. Then, we adopt a quality engineering approach for analyzing the relationships among the quality of the

* Communicating author’s email: yamada@sse.tottori-u.ac.jp
design-review activities, i.e., software reliability, and human factors to clarify the fault-introduction process in the design-review process.

We conduct a design-review experiment of graduate and undergraduate students as subjects. First, we discuss human factors categorized in inhibitors and inducers in the design-review process, and set up controllable human factors in the design-review experiment. Especially, we lay out the human factors on an orthogonal array based on the method of design of experiment [5]. Second, in order to select human factors which affect the quality of the design-review, we perform a software design-review experiment reflecting an actual design process based on the method of design of experiment. For analyzing the experimental results, we adopt a quality engineering approach, i.e., Taguchi-method. That is, applying the orthogonal array $L_{18}(2^1 \times 3^7)$ with inside and outside factors to the human factor experiment and classifying the faults detected in design-review work into descriptive-design and symbolic-design ones, we carry out the analysis of variance by using the data of signal-to-noise ratio (defined as SNR) [6] which can evaluate the stability of quality characteristics, discuss effective human factors, and obtain the optimal levels for the selected inhibitors and inducers.

2. Design-Review and Human Factors

2.1 Design-review

The design-review process is located in the intermediate process between design and coding phases, and has software requirement specifications as inputs and software design specifications as outputs. In this process, software reliability is improved by detecting software faults effectively [7].

2.2 Human factors

The attributes of software designers and design process environment are mutually related for the design-review process. Then, influential human factors for the design-specification as outputs are classified into two kinds of attributes in the following [8, 9, 10]:

(a). Attributes of the design reviewers (Inhibitors)
Attributes of the design reviewers are those of software engineers who are responsible for design-review work. For example, they are the degree of understanding of requirement specifications and design-methods, the aptitude of programmers, the experience and capability of software design, the volition of achievement of software design, etc. Most of them are psychological human factors which are considered to contribute directly to the quality of software design-specification.

(b). Attributes of environment for the design-review (Inducers)
In terms of design-review work, many kinds of influential factors are considered such as the education of design-methods, the kind of design methodologies, the physical environmental factors in software design work, e.g., temperature, humidity, noise, etc. All of these influential factors may affect indirectly to the quality of software design-specification.
3. Design-Review Experiment

3.1 Human factors in the experiment

In order to find out the relationships among the reliability of software design-specification and its influential human factors, we have performed the design-review experiment by selecting five human factors as shown in Table 1 as control factors which are concerned in the review work.

**Table 1: Controllable Factors in the Design-Review Experiment**

<table>
<thead>
<tr>
<th>Control Factor</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGM of classical music in the review work environment [Inducer A]</td>
<td></td>
</tr>
<tr>
<td>Time duration of design-review work [Inducer B]</td>
<td></td>
</tr>
<tr>
<td>Degree of understanding of the design-method [Inhibitor C]</td>
<td></td>
</tr>
<tr>
<td>Degree of understanding of requirement-specification [Inhibitor D]</td>
<td></td>
</tr>
<tr>
<td>Check list (Inducer E, Inducer E, Inducer E)</td>
<td></td>
</tr>
</tbody>
</table>

- **BGM of classical music in the review work environment (Inducer A)**
  Design-review work for detecting faults requires concentrated attentiveness. We adopt a BGM of classical music as the factor of work environment in order to maintain review-efficiency.

- **Time duration of design-review work (Inducer B)**
  In this experiment, we set the subjects design-review work to be completed in approximately 20 minutes. We adopt the time duration of software design-review work with 3 levels such as 20 minutes, 30 minutes and 40 minutes as the factor of work time.

- **Check list (Inducer E)**
  We prepare the check list (CL) which indicates the matters to be noticed in review work. This factor has the following 3 levels: Detailed CL, common CL, and without CL.

- **Degree of understanding of the design-method (Inhibitor C)**
  Inhibitor C of two inhibitors is the degree of understanding of the design-method of R-Net (requirements network). Based on the preliminary tests on the ability to understand the R-Net technique, the subjects are divided into the following 3 groups: High, common, and low ability groups.

- **Degree of understanding of requirement-specification (Inhibitor D)**
  Inhibitor D of two inhibitors is the degree of understanding of requirement-specification. In the similar case as Inhibitor C, based on the preliminary tests on the ability of geometry, are divided into the following 3 groups: High, common, and low ability groups.
3.2 Summary of experiment
In this experiment, we conduct an experiment to clarify the relationships among human factors affecting software reliability and the reliability of design-review work by assuming a human factor model [8, 9, 10] consisting of the inhibitors and inducers. The actual experiment has been performed by 18 subjects based on the same design-specification of a triangle program which receives three integers representing the sides of a triangle and classifies the kind of triangle such sides form [11].

We measured the 18 subjects’ capability of both the degrees of understanding of design-method and requirement-specification by the preliminary tests before the design of experiment. Further, we seeded some faults in the design-specification intentionally. Then, we have executed such a design-review experiment in which the 18 subjects detect the seeded faults.

We have performed the experiment by using the five control factors with three levels as shown in Table 1, which are assigned to the orthogonal-array $L_{18}(2^1 \times 3^7)$ of the design of experiment as shown in Table 3.

3.3 Classification of Detected Faults
We distinguish the design parts as follows to be pointed out in the design-review as detected faults into the descriptive-design and symbolic-design parts.

- **Description-design faults**
The descriptive-design parts consist of words or technical terminologies which are described in the design-specification to realize the required functions. In this experiment, the descriptive-design faults are algorithmic ones, and we can improve the quality of design-specification by detecting and correcting them.

- **Symbolical-design faults**
The symbolical-design parts consist of marks or symbols which are described in the design-specification. In this experiment, the symbolical-design faults are notation mistakes, and the quality of the design-specification can not be improved by detecting and correcting them.

3.4 Data analysis with classification of detected faults
For the orthogonal-array $L_{18}(2^1 \times 3^7)$, setting the classification of detected faults as outside factor $R$ and the control factors A, B, C, D, and E as inside factors, we perform the design-review experiment. Here, the outside factor $R$ has two levels such as descriptive-design parts($R_1$) and symbolical-design parts($R_2$).

4. Analysis of Experimental Results
4.1 Definition of SNR
We define the efficiency of design-review, i.e., the reliability, as the degree that the design reviewers can accurately detect correct and incorrect design parts for the design-specification containing seeded faults. There exists the following relationship among the total number of design parts, $n$, the number of correct design parts, $n_0$, and the number of incorrect design parts containing seeded faults, $n_1$:

$$n = n_0 + n_1$$  \hspace{1cm} (1)
Therefore, the design parts are classified as shown in Table 2 by using the following notations:

\[ n_{00} = \text{the number of correct design parts detected accurately as correct design parts}, \]
\[ n_{01} = \text{the number of correct design parts detected by mistake as incorrect design parts}, \]
\[ n_{02} = \text{the number of incorrect design parts detected by mistake as correct design parts}, \]
\[ n_{03} = \text{the number of incorrect design parts detected accurately as incorrect design parts}, \]

where two kinds of error rate are defined by

\[ p = \frac{n_{01}}{n_0}, \quad (2) \]
\[ q = \frac{n_{02}}{n_0}. \quad (3) \]

Considering the two kinds of error rate, \( p \) and \( q \), we can derive the standard error rate, \( p_0 \), [6] as

\[ p_0 = \frac{1}{1 + \left( \frac{1}{p} - 1 \right) \left( \frac{1}{q} - 1 \right)^2}. \quad (4) \]

Then, the signal-to-noise ratio based on (4) is defined by

\[ \eta_0 = -10 \log_{10} \left( \frac{1}{(1 - 2p_0)^2} - 1 \right). \quad (5) \]

The standard error rate, \( p \), can be obtained from transforming (5) by using the signal-to-noise ratio of each control factors as

\[ p_0 = \frac{1}{2} \left( 1 - \frac{1}{\sqrt{10} \left( \frac{20}{10} \right) + 1} \right). \quad (6) \]

**Table 2: Two Kind of Inputs and Outputs in the Design-Review Experiment.**

<table>
<thead>
<tr>
<th>(i) Observed values</th>
<th>(ii) Error rates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>output</strong></td>
<td><strong>0(true)</strong></td>
</tr>
<tr>
<td><strong>input</strong></td>
<td><strong>0(true)</strong></td>
</tr>
<tr>
<td>0(true)</td>
<td>( m_0 )</td>
</tr>
<tr>
<td>1(false)</td>
<td>( m_0 )</td>
</tr>
<tr>
<td>total</td>
<td>( n_0 )</td>
</tr>
</tbody>
</table>
The method of experimental design based on an orthogonal-array is special one that requires only a small number of experimental trials to help us discover main factor effects. On traditional researches [4, 8], the design of experiment has been conducted by using orthogonal-array $L_{12}(2^{11})$. However, since the orthogonal-array $L_{12}(2^{11})$ has two levels for grasp of factorial effect to the human factors experiment, the middle effect between two levels can not be measured. Thus, in order to measure it, we adopt the orthogonal-array $L_{18}(2^1 \times 3^7)$ can lay out one factor with 2 levels (1, 2) and 7 factors with 3 levels (1, 2, 3) as shown in Table 3, and dispense with $2^1 \times 3^7$ trials by executing independent 18 experimental trials each other.

Considering such circumstances, we can obtain the optimal levels for the selected inhibitors and inducers efficiently by using the orthogonal-array $L_{18}(2^1 \times 3^7)$.

5. Data Analysis with Classification of Detected Faults

The experimental results of observed values classified by two types of design parts discussed in 4.1 are shown in Table 3. The data of the SNR calculated by Eq. (5) are also shown in Table 3.

A result of the analysis of variance for the descriptive-design parts is shown in Table 4, and that for the symbolical-design parts shown in Table 5, which are based on the data of SNR as shown in Table 3. In Tables 4 and 5, $f$, $S$, $V$, $F_0$, and $\rho$ represent the degree of freedom, the sum of squares, the unbiased variance, the unbiased variance ratio, and the contribution ratio, respectively, for performing the analysis of variance. In order to obtain the precise analysis results, the factor with clearly small $S$ is pooled in the factor of error (Factor e). Then, we performed the analysis of variance based on the factor of new pooled error (Factor $e'$). Fig. 1 shows the factor effect for each level in the control factor which affects the design-review result based on the SNR calculated from the observation values.

![Fig. 1: The Estimation of Significant Factors with Classification of Detected Faults](image)
5.1 Descriptive-Design Faults

In design-review work, effective review conditions for correcting and removing algorithmic faults are BGM of classical music, “yes ($A_1$)” and design-review time, “40minutes ($B_2$)”. As reviewer’s capability, the degree of understanding of the design-method (R-Net Technique), “high ($C_1$)”, and that of the requirement-specification, “high ($D_1$)”, are derived as optimal conditions.
5.2 Symbolic-Design Faults

In design-review work, effective review conditions for correcting and removing notation mistakes is an optimal condition that the degree of understanding of requirement-specification is "high(C1)" as reviewer's capability.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$f$</th>
<th>$S$</th>
<th>$V$</th>
<th>$F_o$</th>
<th>$\rho$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1</td>
<td>0.037</td>
<td>0.037</td>
<td></td>
<td>8.180</td>
</tr>
<tr>
<td>$B$</td>
<td>2</td>
<td>29.041</td>
<td>14.521</td>
<td>2.975</td>
<td>32.618</td>
</tr>
<tr>
<td>$C$</td>
<td>2</td>
<td>86.640</td>
<td>43.320</td>
<td>8.8/5**</td>
<td>12.108</td>
</tr>
<tr>
<td>$D$</td>
<td>2</td>
<td>38.300</td>
<td>19.15</td>
<td>3.923</td>
<td>11.889</td>
</tr>
<tr>
<td>$E$</td>
<td>2</td>
<td>37.783</td>
<td>18.892</td>
<td>3.870</td>
<td>10.000</td>
</tr>
<tr>
<td>$A \times B$</td>
<td>2</td>
<td>4.833</td>
<td>2.416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_1$</td>
<td>6</td>
<td>38.759</td>
<td>6.460</td>
<td></td>
<td>352.000</td>
</tr>
<tr>
<td>$e_2$</td>
<td>9</td>
<td>43.929</td>
<td>4.881</td>
<td></td>
<td>100.000</td>
</tr>
</tbody>
</table>

Table 5: The Results of Analysis of Variance (Symbolic-Design Faults)

6. Data Analysis with Correlation among Inside and Outside Factors

6.1 Analysis of Experimental Results

We analyze simultaneous effects of outside factor R and inside control factors A, B, C, D, and E. As a result of the analysis of variance by taking account of correlation among inside and outside factors discussed in 3 and 4, we can obtain Table 6. There are two kinds of errors in the analysis of variance: $e_1$ is the error among experiments of the inside factors, and $e_2$ is the mutual correlation error between $e_1$ and the outside factor. In this analysis, since there was no significant effect by performing F-testing for $e_1$ with $e_2$, F-testing for all factors was performed by $e_2$. As the result, the significant control factors such as the degree of understanding of the design-method (Factor C), the degree of understanding of requirement-specification (Factor D), and the classification of detected faults (Factor R) were recognized. Fig.2 shows the factor effect for each level in the significant factors which affect design-review work.

6.2 Discussion

As a result of analysis, in the inside factors, only Factors C and D are significant and the inside and outside factors are not mutually interacted. That is, it turns out that the reviewers with the high degree of understanding of the design-method and the high degree of understanding of requirement-specification exactly can review the design-specification efficiently regardless of the classification of detected faults. Moreover, the result that outside factor R is highly significant, and the descriptive-design faults are detected less than the symbolic-design faults, can be obtained. That is, although it is a natural result, it is difficult to detect and correct the algorithmic faults which lead to improvement in quality rather than the notation mistakes. However, it is important to detect and correct the algorithmic faults as an essential problem of the quality improvement for design-review work. Therefore, in order to increase the rate of detection and correction of the algorithmic faults which lead to the improvement of quality, it is required before design-
review work to make reviewers fully understand the design technique used for describing design-specification and the contents of requirement specifications.

7. Conclusions

In this paper, in order to improve the reliability of software design-review, we have proposed a quality engineering approach, i.e., Taguchi-method, which can find out the relationships among human factors and the reliability of design-review. Applying the orthogonal array $L_{18}(2^{7} \times 3^{7})$ and SNR, we have performed the experiment of software design-review, and verified the relationships among the selected human factors categorized in inhibitors and inducers and the reliability of design-review.

Further studies on human factors in the software design process by using this approach are needed to support the findings in this paper.
Acknowledgment

This work was supported in part by the Grant-in-Aid for Scientific Research(C) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan under Grant No. 18510124

References


Shigeru Yamada has been working as a Professor at the Faculty of Engineering of Tottori University, Tottori-shi, Japan, since 1993. He has published numerous technical papers in the area of software reliability engineering, project management, and statistical quality control. He has authored several books such as "Introduction to Software Management Model" (Kyoritsu Shuppan, Tokyo), "Software Reliability Models: Fundamentals and Applications" (JUSE, Tokyo), "Statistical Quality Control for TQM" (Corona Publishing, Tokyo), and "Software Reliability: Modeling, Tool, Management" (The Society for Project Management, Chiba).