

# 軽微な指摘を抑制したソフトウェアレビューにおける指摘欠陥の分析

三浦 一輝<sup>1</sup> 森崎 修司<sup>1</sup>

**概要:** ソフトウェアレビューにおいて誤字や体裁の誤りをはじめとする軽微な欠陥の指摘はレビューのコスト対効果を下げる原因となる可能性がある。本研究では、システム開発に携わる者を対象として、普段通りのレビューと、軽微な欠陥の指摘を控えるよう依頼した上で実施したレビューを試行し、比較した。欠陥指摘内容を漏れ、提案、曖昧さ、誤り、誤字の5種に分類し、指摘された欠陥件数について結果を比較した。その結果、軽微な指摘を抑制することで仕様書の漏れに関する指摘が増加することがわかった。また、レビューアの開発経験年数を4年以下、5年から9年、10年以上の3種に分類し、開発経験年数が10年以上のレビューアに漏れの件数が増加する結果が得られた。

**キーワード:** 欠陥分類, 設計レビュー, レビュー経験

## 1. はじめに

ソフトウェアレビューはソフトウェアの品質を向上させるための手法である。ソフトウェアレビューを実施することで、ドキュメントに混入された欠陥を早期に検出し除去することによるドキュメントの品質向上、手戻り作業の低減、および開発全体の生産性向上が期待できる。

ソフトウェアレビューの目的の1つは重大な欠陥を検出することである。重大な欠陥を検出した際の生産性向上効果について、要件定義の段階で検出した欠陥の修正コストと運用段階で検出した欠陥の修正コストを比較すると相対的なコストは40~1000倍となる [11]。

しかし、一般的なソフトウェア開発プロジェクトではドキュメントのレビューに費やすことのできる時間は限られている。誤字脱字の指摘といった本来設計品質に影響が低い欠陥指摘に時間を費やし、ほかの重大度の高い欠陥検出に費やす時間が減少すれば、期待していたドキュメントの品質向上や生産性向上の効果が得られず時間を無駄にすることになる。限られたレビュー時間で設計仕様書の品質を高めるためには、重大度の高い欠陥を優先して検出する必要がある。

また、経験年数が高くレビュースキルの見込める人は誤字や体裁の誤りのようにとくに知識がなくても指摘できるような欠陥の検出に時間を費やす必要はない。誤字や体裁の

誤りの欠陥を無視した場合、引き継ぎの漏れや処理内容の不整合などの重大な欠陥検出に時間を使うことができると推測できる。

ソフトウェア開発に携わる実務者を被験者としたレビューの実証的試行を実施し軽微な欠陥指摘を抑制が与える影響について以下の Research Question (以下 RQ) に回答する。

RQ1 軽微な指摘を抑制するとにより指摘欠陥種別はどのように変化するか。

RQ2 レビューアの経験年数によって影響の与え方は変化するのか。

試行において被験者には普段通りのレビューと軽微な欠陥の指摘の抑制を依頼したレビューを行う。また、被験者の経験年数によりレビュー結果を比較するため、開発経験年数及びレビュー経験年数をアンケートにより収集した。

## 2. 試行

調査ではシステムの仕様書に含まれる軽微な欠陥指摘を抑制することを依頼した際に、普段のレビューと比較して欠陥指摘の傾向に変化があるかを調査するため、ソフトウェア開発に携わる方を対象に仕様書のレビューを依頼し、指摘された欠陥件数について分類を行った。また、軽微な指摘を抑制することによる影響がレビューアの経験年数によりどのように変化するかを調べるため開発経験年数、レビュー経験年数による分類を行った。対象とするデータについては表2、表3の通りである。また、対象者の開発/レ

<sup>1</sup> 静岡大学情報学部  
Faculty of Informatics, Shizuoka University

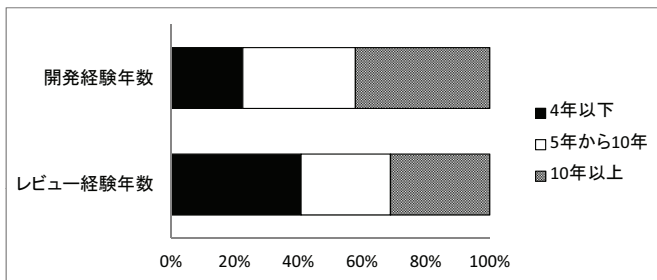


図 1 経験年数の構成比

	ドキュメント A	ドキュメント B
仕様書	基本設計	
対象言語	自然言語	
システム概要	Web アプリケーション	
システム内容	テニスコート予約システム	会議室予約システム
ページ数	3	

表 1 対象ドキュメント

	グループ A	グループ B
人数	97	113
1 回目	ドキュメント A	ドキュメント B
2 回目	ドキュメント B	ドキュメント A

表 2 グループと対象ドキュメント

レビュー経験年数の構成比は図 1 の通りである。本試行ではシステム開発における組み込み、ソフトウェア製品、エンタープライズ、ユーザー企業の情報システムの 4 つの部門に所属する実務者 210 名を対象にレビュー試行を行った。

本試行において軽微な指摘を抑制したことが与える影響について調査するため、表 1 に示す 2 つのドキュメントを用意した。

## 2.1 手順

ドキュメントによる影響を考慮し、ドキュメントをレビューする被験者を A と B の 2 つのグループに分けて試行を行った。グループの人数とドキュメントのレビュー順を示したものが表 2 である。試行では以下の手順に従って試行を行った。また、アンケートでは被験者が検出した欠陥の内容、開発経験年数、レビュー経験年数を記載する質問項目を用意した。

### ● 1 回目

- (1) アンケート用紙を配布する。
- (2) 表 2 で 1 回目に該当するドキュメントを配布する。
- (3) 20 分程時間を設け、個人で欠陥検出を行う。
- (4) 5 人程のメンバで欠陥指摘内容の確認を行う。

### ● 2 回目

- (1) 軽微な指摘を抑制したレビューを行うことを依頼する。
- (2) 表 2 で 2 回目に該当するドキュメントを配布する。
- (3) 20 分程時間を設け、個人で欠陥検出を行う。
- (4) 5 人程のメンバで欠陥指摘内容の確認を行う。

欠陥分類	概要	
I-i	抜け漏れ	記述や実装が漏れ
I-ii	機能提案	実装が望ましい機能
II	曖昧さ	何通りかに解釈できるもの
III	誤り	記述されている情報が誤っている
IV	誤字脱字	誤変換や体裁の誤り

表 4 指摘内容に着目したカテゴリ分類

## 2.2 欠陥種別の分類

レビュー手法の中に欠陥分類ごとにレビューアが要求仕様書を検査する手法 [7] がある。ここでは「漏れ、曖昧さ、誤り」の 3 種の着眼点から欠陥を検出する方法が記載されている。「漏れ」とは仕様書内に記述や実装が漏れていることである。曖昧さは何通りかに解釈でき解釈によっては誤って実装される恐れのある表現である。「誤り」は数値の誤りなど、整合性がとれていないなど記述されている情報が誤っている表記のことを指す。この 3 種の欠陥種別の分類に加え、軽微な欠陥としてシステム構築にあまり影響のない誤字脱字を含めた 4 種の欠陥種別に分類した。その詳細を表 3 に示す。

## 3. 結果

### 3.1 指摘された欠陥の分類

表 3 で分類した欠陥種別について、カテゴリ I の欠陥に機能提案のような指摘がお多かったため、カテゴリ I-ii として分類した。そのような指摘として「使用者に対しメール通達するなどの仕組みも必要ではないか」といった指摘があった。欠陥分類を表 4 に示す。

- カテゴリ I-i に分類される欠陥は機能の抜け、操作手順の漏れなどのシステム内の抜け、漏れを含む指摘が該当する。
- カテゴリ I-ii に分類される欠陥はなくても問題ないが追加されることが望ましい機能として提案されたものである。
- カテゴリ II に分類される欠陥は解釈する人によって解釈の仕方が異なるような表現が該当する。たとえば「画面が表示される」という文章だけではどの画面を表示すればよいかについてさまざまな解釈が生まれてしまう。またシステム内だけでなくシステム外の情報も考慮する必要がある欠陥もこの分類に該当する。
- カテゴリ III に分類される欠陥は仕様書でモジュールごとに定義している最大文字数が異なっている、設定可能な数値の不整合などの欠陥が該当しモジュール間で共通であるべきものが不整合の状態である欠陥が分類される。
- カテゴリ IV に分類される欠陥は誤字脱字である。

カテゴリ	欠陥分類	概要	具体例
I	漏れ	記述や実装が漏れ	ログイン画面がない
II	曖昧さ	何通りかに解釈できるもの	A の場合 B または C
III	誤り	記述されている情報が誤っている	48 都道府県
IV	誤字脱字	誤変換や体裁の誤り	「しま s」ではなく「します」

表 3 欠陥種別の分類

分類	ドキュメント A		ドキュメント B	
	総件数	重複を除く	総件数	重複を除く
I-i	732 件	55 件	685 件	55 件
I-ii	98 件	13 件	87 件	12 件
II	361 件	47 件	287 件	43 件
III	258 件	4 件	148 件	6 件
IV	144 件	7 件	99 件	12 件
総数	1593 件	125 件	1306 件	135 件

表 5 各カテゴリに分類される指摘欠陥件数

	I-i	I-ii	II	III	IV
グループ A	0.0061*	0.0058*	0.9989	0.0000*	0.0000 *
グループ B	0.0000*	0.0028*	0.3508	0.3916	0.0000 *

\* $p < 0.05$

表 6 欠陥種別ごとの指摘件数

### 3.2 RQ1. 軽微な指摘を抑制するとにより指摘欠陥種別はどのように変化するのか。

表 4 で分類した指摘欠陥種別について指摘件数の変化を分析した。表 5 は全体の指摘件数を表したものである。重複を除いたものであり「コートの面数が違う」や「コートの面数が 10 面ではなく 6 面になっている」のように同じ意味を持つ指摘をまとめ集計を行っている。図 2 は全体の指摘件数について、欠陥種別ごとの指摘件数の割合を求めたものである。得られた結果から指摘欠陥種別について、帰無仮説を 1 回目と 2 回目の指摘件数に差がないとした場合の Wilcoxon の U 検定の結果 ( $p$  値) を 6 に示す。  $p$  値が小さいほど 1 回目と 2 回目の指摘件数の平均値に差があると言える。図 2 からそれぞれのグループについて、カテゴリ I-i, カテゴリ I-ii に分類される欠陥の指摘件数の割合は増加したと判断できる。6 より有意水準 5% で統計的有意差があることから軽微な指摘を抑制を依頼することでカテゴリ I-i, カテゴリ I-ii の指摘件数が増加すると言える。カテゴリ IV に分類される欠陥についてはどちらのグループも有意水準 5% で減少する結果が得られた。グループ B では見られなかったが、グループ A ではカテゴリ III に分類される欠陥について有意水準 5% で指摘件数が減少する結果も得られた。

### 3.3 レビューアの経験年数によって影響の与え方は変化するのか。

軽微な欠陥指摘を抑制することによる指摘欠陥の変化は経験年数による影響があるか調査した。表 7, 表 8 は指摘欠陥種別ごとに指摘件数の変化について、帰無仮説を 1 回

	4 年以下	5 年から 9 年	10 年以上
I-i	0.3346	0.1855	0.0204
I-ii	0.4398	0.3754	0.0098*
II-i	0.1189	0.1373	0.2299
II-ii	0.9851	0.9391	0.8957
III	0.0033*	0.0003*	0.0000*
IV	0.0049*	0.0000*	0.0000*

\* $p < 0.05$

表 7 グループ A に関する指摘件数

	4 年以下	5 年から 9 年	10 年以上
I-i	0.0013*	0.32264	0.0002*
I-ii	0.1300	0.1840	0.0267*
II-i	0.3569	0.3555	0.9941
II-ii	0.0762	0.5615	0.1856
III	0.5687	0.1206	0.6922
IV	0.0000*	0.0006*	0.0000*

\* $p < 0.05$

表 8 グループ B に関する指摘件数

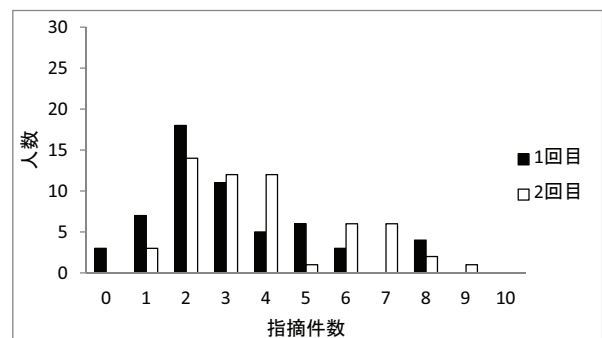


図 3 グループ A における経験年数 10 年以上の場合のカテゴリ I-i に関する指摘傾向の変化

目と 2 回目の指摘件数に差がないとした場合の Wilcoxon の U 検定の結果 ( $p$  値) を示したものである。

表 7, 表 8 の結果から有意水準 5% 未満で 1 回目と 2 回目の欠陥指摘件数に差がみられたデータについて、1 人あたりの指摘件数が 1 回目と 2 回目とどのように変化したかを図 3 から図 13 のグラフにまとめた。

#### 3.3.1 グループ A における指摘傾向の変化

図 3 から開発経験が 10 年以上の場合、2 件以下の指摘件数が減少し 3 件以上指摘件数が増加する結果となった。

図 4 から開発経験年数が 10 年を超える場合にカテゴリ II-ii の欠陥指摘数が 0 件の人数が減少し、1 件あるいは 2 件の指摘をする人数が増加した。

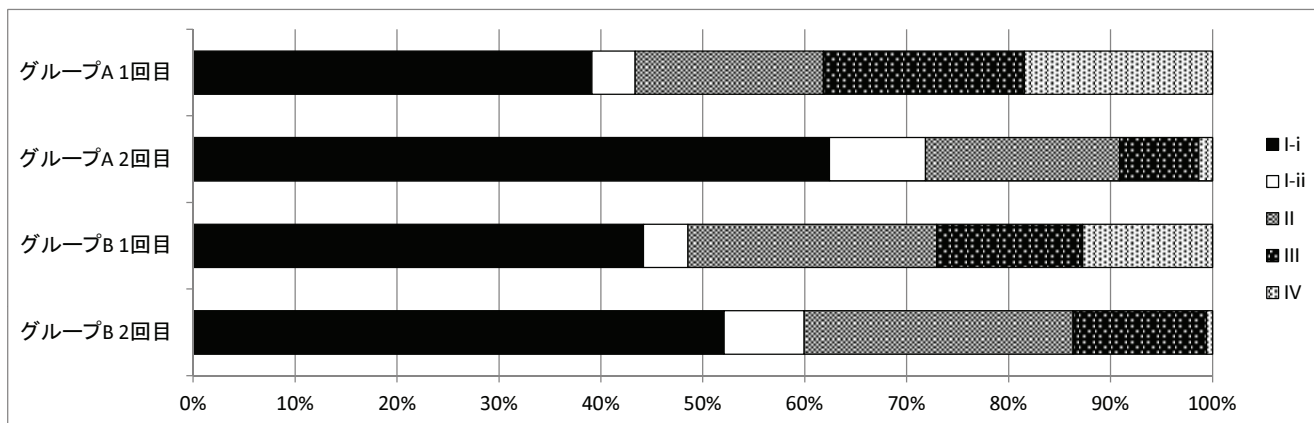


図 2 各グループにおける欠陥指摘件数の推移

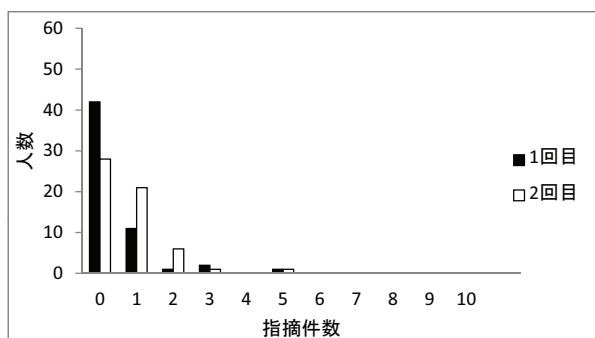


図 4 グループ A における経験年数 10 年以上の場合のカテゴリ II-ii に関する指摘傾向の変化

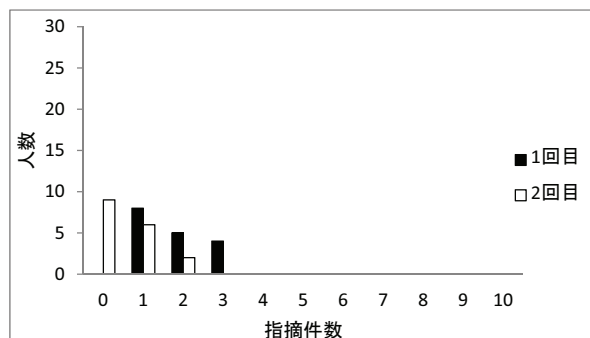


図 6 グループ A における経験年数 5 年から 9 年の場合のカテゴリ III に関する指摘傾向の変化

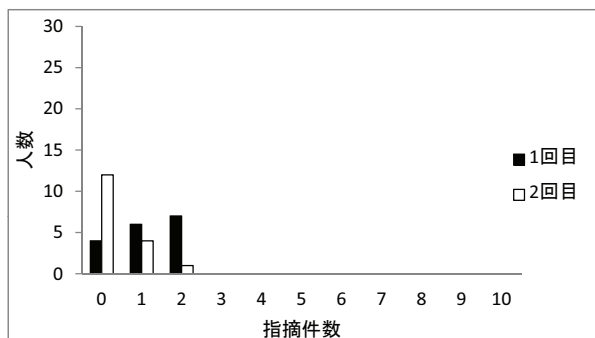


図 5 グループ A における経験年数が 4 年以下の場合のカテゴリ III に関する指摘傾向の変化

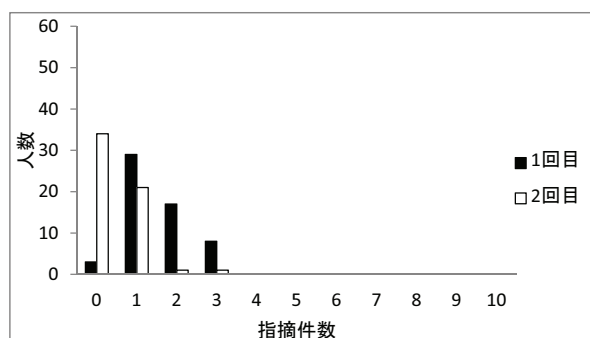


図 7 グループ A における経験年数 10 年以上の場合のカテゴリ III に関する指摘傾向の変化

図 5, 図 6, 図 7 はカテゴリ III の欠陥指摘数についてのグラフである。カテゴリ III の指摘件数について 1 回目と比較し 2 回目では経験年数にかかわらず減少し、指摘件数が 0 件である場合が増加した。

図 8, 図 9, 図 10 はカテゴリ IV の欠陥指摘数についてのグラフである。図 8, 図 9, 図 10 から経験年数に限らず指摘件数が 0 件である場合が増加したことがわかる。一方で、僅かではあるが誤字脱字の指摘をする人もみられた。

### 3.3.2 グループ B における指摘傾向の変化

図 11, 図 12 より経験年数が 4 年以下また 10 年以上のどちらも、指摘件数が 3 件以上である人数が増加し、1 人当たりの指摘数が 1 件ほど増加する結果が得られた。また、指摘件数が 1 件である人数は 0 人であったが、指摘件数が

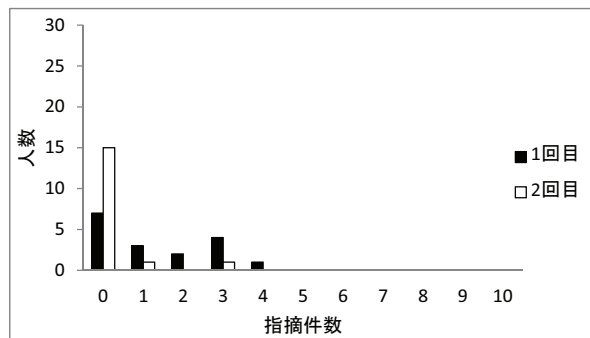


図 8 グループ A における経験年数が 4 年以下の場合のカテゴリ IV に関する指摘傾向の変化

0 件である人も少数みられた。

図 13 の結果から指摘件数が 0 件である人数が減少し、1

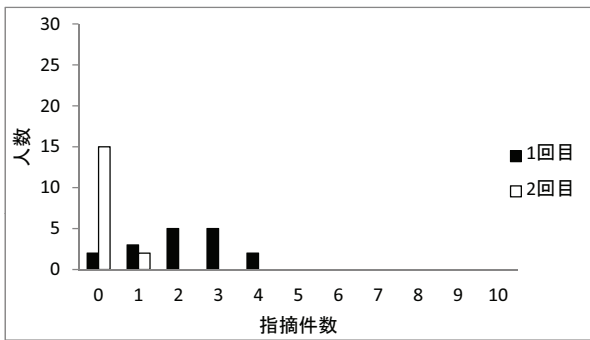


図 9 グループ A における経験年数 5 年から 9 年の場合のカテゴリ IV に関する指摘傾向の変化

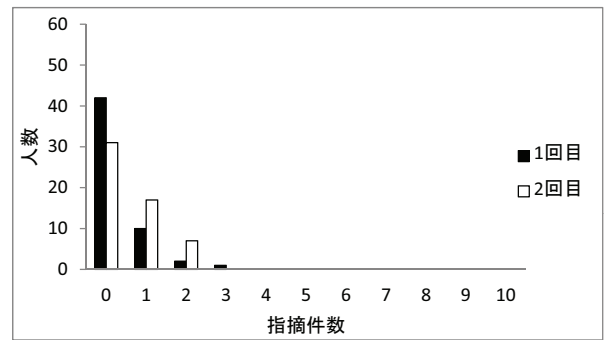


図 13 グループ B における経験年数 10 年の場合のカテゴリ I-ii に関する指摘傾向の変化

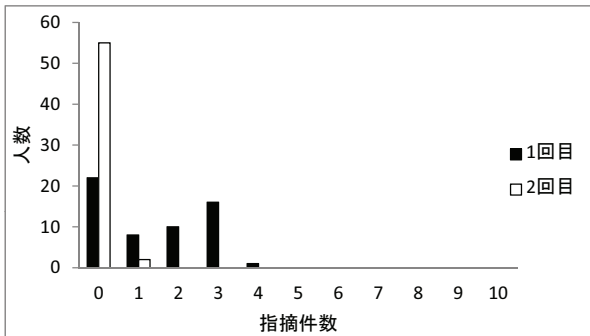


図 10 グループ A における経験年数 10 年以上の場合のカテゴリ IV に関する指摘傾向の変化

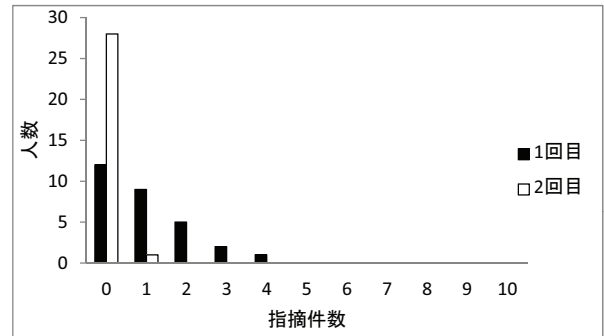


図 14 グループ B における経験年数 4 年以下の場合のカテゴリ IV に関する指摘傾向の変化

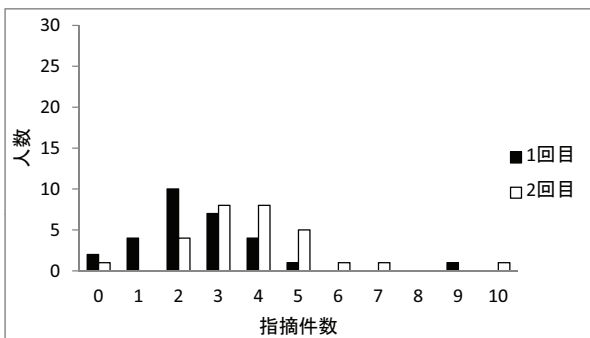


図 11 グループ B における経験年数 4 年以下の場合のカテゴリ I-i に関する指摘傾向の変化

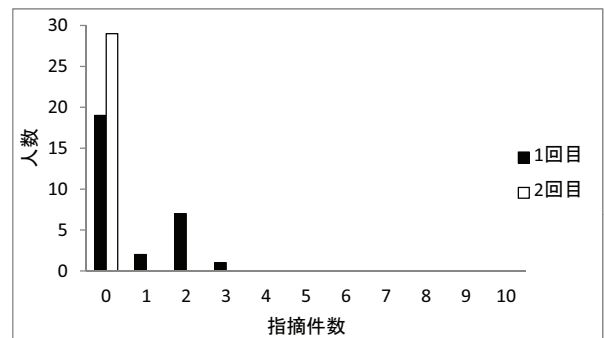


図 15 グループ B における経験年数 5 年から 9 年の場合のカテゴリ IV に関する指摘傾向の変化

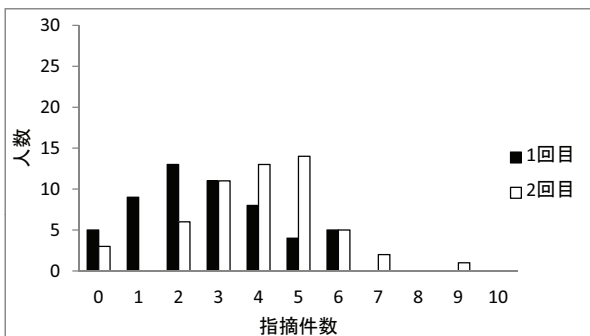


図 12 グループ B における経験年数 10 年以上の場合のカテゴリ I-i に関する指摘傾向の変化

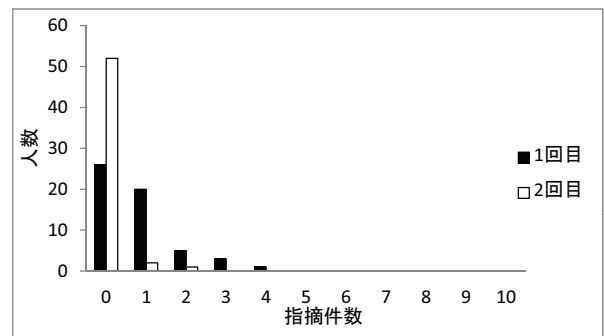


図 16 グループ B における経験年数 10 年の場合のカテゴリ IV に関する指摘傾向の変化

件または 2 件の欠陥を指摘する人数が増加する結果が得られた。

図 14, 図 15, 図 16 はカテゴリ IV の欠陥指摘数についてのグラフである.. 図から経験年数がいずれの場合も指摘件数が 0 件である人数が増加したことがわかる. 一方で,

僅かではあるが誤字脱字の指摘をする人もみられた.

## 4. 考察

### 4.1 RQ1. 指摘欠陥種別と欠陥の指摘数

軽微な欠陥指摘の抑制を依頼した状態でレビューを行う



ことにより抜け漏れの指摘件数が増加することがわかった。1人当たりの平均指摘数を求めるとどちらのグループも1件ほどの指摘件数が増加していた。ドキュメントレビューの段階で抜けや漏れといった欠陥が指摘された場合、テスト工程で本来必要な作業が抜けていることが判明するといった状態を回避することができる。この結果からレビューアに対し軽微な欠陥指摘の抑制を依頼することによりドキュメントに含まれる欠陥検出件数が増加し、修正によるドキュメント品質向上の可能性が高いことが示された。

一方で、誤りの指摘が減少する結果となった。これはこの種の欠陥が軽微なものであると判断された、あるいはそもそも誤りに関しての指摘を見つけることができなかった可能性がある。誤りに分類される欠陥について行為が正しくても知識が間違っていれば、要求に欠陥が生じる可能性がある。そのため誤りのカテゴリに分類される欠陥は指摘する必要がある。誤りの欠陥指摘件数が低いと判断された場合には、誤りの観点を基に再度レビューを実施し誤りに分類される欠陥を検出する必要があると考えられる。

#### 4.2 RQ2. 開発経験年数による影響

表7, 表8および図3から図13の結果から開発経験年数により差がある結果が得られた。開発経験年数が10年以上の場合にグループA, グループBのいずれも抜け漏れと提案の指摘件数が増加する結果が得られた。このことから開発経験年数が10年を超えるレビューアについて、軽微な欠陥指摘の抑制を依頼することで重大な欠陥の検出件数の増加が期待できる。また、グループBでは開発経験年数が4年未満の場合には抜け漏れの指摘件数が増加する結果が得られた。この結果から経験年数が低いレビューアに対しても指摘件数が増加する効果が得られることが考えられる。

一方で、グループAの誤りの指摘に関しては開発経験年数に関係なく減少するという結果が得られた。この結果から経験年数に関わらず誤りの指摘を軽微なものとして判断してしまう傾向があると推測される。軽微な欠陥指摘の抑制したレビューを行う際には、局所的な欠陥を指摘しつつも軽微な指摘を抑制してもらう必要がある。また誤字脱字について全てのレビューアが指摘しないという状態にはならなかった。

#### 5. まとめ

ドキュメントレビューにおいて軽微な欠陥を指摘を抑制することで、全体の欠陥の指摘件数にどのような影響があるのか調査した。試行では2つのグループを用意し、それぞれのグループで指示のない状態と、軽微な欠陥指摘の抑制を依頼した状態の2通りのレビューを実施した。試行で得られたデータから指摘内容と開発経験年数をもとに分類

し指摘欠陥の分析を行った。指摘内容の分類は、記述の抜け漏れ、望ましい機能の提案、曖昧さ、仕様の誤り、誤字脱字の5種である。

指摘内容の分類からは抜け漏れの指摘件数及び提案の指摘件数が増加することがわかった。この結果は軽微な欠陥指摘を抑制を依頼することによりレビュー品質が向上する効果を得られることを示している。また、経験年数の分類は4年以下, 5年から9年, 10年以上の3種について分析を行ったところ、経験年数毎の分類では開発経験年数が10年以上の場合について抜け漏れと提案の指摘件数が増加する傾向が得られた。経験年数が豊富な方をレビューアとして選定することにより、時間が限られた状態、新規プロジェクトのように詳細な観点を設定するためのデータが揃っていない場合にレビュー効率の向上が期待できる。

一方で、誤りの指摘件数については減少する結果が得られた。誤りの指摘を軽微なものとして判断する可能性が推測されるので、誤りに関しての説明を入れ指摘が抑制されないように注意した状態で試行することが今後の課題の1つとして費やすことのできる。今後はレビューアのより詳細な選定基準を設けることを目的とし、レビュー経験年数や職種を考慮した分類を設定し、分析を続けていく予定である。

#### 謝辞

試行にご協力いただいた実務者の御方々に深く御礼申し上げます。本研究は文部科学省科学研究補助費(基盤研究B: 課題番号23300009)による助成を受けた。

#### 参考文献

- [1] V. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sorumgard, M. Zelkowitz: The Empirical Investigation of Perspective-based Reading, *Journal of Empirical Software Engineering*, vol.2, no.1, pp.133-164(2006)
- [2] M. E. Fagan: Design and Code Inspection to Reduce Errors in Program Development, *IBM Systems Journal*, vol.15, no. 3, pp. 182-211(1976)
- [3] Gilb T. "Agile Specification Quality Control: Shifting Emphasis from Cleanup to Sampling, Measurement, Motivation and Prevention of Defects," In *Proceedings of 15th Annual International Council on Systems Engineering* (2005)
- [4] KE Wieggers: Seven Truths About Peer Reviews, *Cutter IT Journal*, July 2002
- [5] T. Thelin, P. Runeson, B. Regnell: Usage-based Reading: An Experiment to Guide Reviewers with Use Cases, *Information and Software Technology*, vol. 43, no. 15, pp.925-938(2001)
- [6] Barry w. Boehm, "Software Engineering Economics", Prentice-Hall, 1981.
- [7] A. Porter, L. Votta and V. Basili, "Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment," *IEEE Transactions on Software Engineering*, Vol. 21, No. 6, pp. 563-575 (1995)
- [8] Mika V. Mantyla and Casper Lassenius. "What types

of defects are really discovered in code reviews? ” IEEE Transactions on Software Engineering,430.448, 2009.

- [9] 野中 誠, “設計・ソースコードを対象とした個人レビュー手法の比較試行” 情報処理学会研究報告. ソフトウェア工学研究会報告 2004, 25-31, 2004
- [10] 田口 雅裕, “ソースコードの読解方法とレビューアの経験が欠陥検出の効果と効率に与える影響の分析”, 奈良先端科学技術大学院大学, 2012
- [11] 酒井 郁子, 舘 伸幸, “強いプログラムを作るテクニックを学ぶ 第9回 分析モデルをレビューする”, インターフェース, vol.37, no.9, pp.153-158, 2011.