

指摘内容と工程内での検出時期によるレビュー指摘の分析

阿形 尚季¹ 森崎 修司¹

概要：レビュー実施時期の最適化とレビューアの割当ての効率化を目的として、実務者によるレビュー指摘結果を指摘内容と検出時期によって分類した。指摘内容の分類は、ソフトウェアが実現する機能を利用する利用者や他システムとの連携をはじめとした外界との不整合、記述の漏れ、誤り、体裁や表記のミスの4つであり、これらとレビューアの経験に関係があるかを調べた。検出時期は欠陥の混入時点ですぐに検出できるかどうかを基準とし、記載とともに検出できる、関連する他の部分が記載されれば検出できる、全てが記載されなければ検出できない、の3つであり、これらの割合とレビューアの経験に関係があるかを調べた。その結果、外界の情報が影響する欠陥は、経験年数の長いレビューアから指摘されることが多かった。また、序盤、終盤に検出できる欠陥は開発の経験年数との関連があることを示す結果が得られた。

キーワード：欠陥分類、欠陥分析、レビューアの経験

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。[2] ソフトウェアレビューは早期段階での欠陥の発見によって欠陥修正のコストを低減する。網羅的な欠陥指摘の方法や手順が数多く提案されている[1][4][5]が、欠陥指摘の時期やそれを意識した指摘内容の分類は少ない。欠陥指摘の時期が明らかになれば、特定の欠陥が同一ドキュメントに多数現れることを防げたり、工程終盤に長時間のレビューを実施することを防げたりする可能性がある。また、指摘内容の分類により、レビューの実施時期に応じて、レビューに参加すべきレビューアの選定基準として活用できることが期待される。

標準的なソフトウェアレビューは開発工程の終盤に実施される。例えば、基本設計書のレビューは、基本設計工程の間に混入した欠陥を検出し修正するために、基本設計の最後の段階において実施される。しかしながら、基本設計書に記入した内容が誤っている場合など、混入してすぐに検出できる欠陥もある。一方で、基本設計書内のすべての記述内容を検討しないと検出できない欠陥もある。「エラーコードの定義と使われ方の矛盾」、「アルゴリズムの誤り」、「誤入力」は混入してすぐに検出できる欠陥の例である。また、「トランザクション処理中の処理速度の低下」、「機能定義の不足」などは基本設計書内のすべての記述内容を検討しないと検出できない欠陥の例である。

レビューに参加するレビューアの選定基準を意識した指摘内容が明らかになれば、参加すべきレビューアが明らかになることが期待される。指摘内容の分類として「別システムからの情報が必要」のような外部の情報を熟知しているレビューア、「予約のシステムなのに予約の確認ができない」というようなレビュー対象を熟知しているレビューア、「機能間で共通であるべき数値が異なっている」というようなレビュー対象の作成に携わったレビューア、「メッセージにおいて「エェです」の誤字」というようなレビュー対象のことを十分に知らない場合でも検出できる欠陥といった分類が考えられる。

本稿では、レビュー指摘を検出可能時期と指摘内容により分類することにより、レビュー実施の効率化につながるか実証的に評価する。検出可能時期や指摘内容の分類が明確になれば、レビュー対象の作成が完了していない時点でのレビュー実施やその際に必要となるレビューアの選定が可能となる。本稿では、実務者によるレビュー指摘結果を、検出可能時期と指摘内容で分類する。また、実務者の経験年数等により傾向がかわるかどうかを確認する。

2. 指摘された欠陥の分類

レビューで指摘される欠陥は観点を決定することで、様々な分類をすることが可能である。レビュー実施者の属性とその観点の間に関連性がある場合は、レビューを実施する際にレビューを行う観点を設定し、指摘された欠陥を分析することでレビューの品質の向上、効率化が見込まれる。そのためにはレビュー実施者の属性と指摘される欠陥の関

¹ 静岡大学情報学部
Faculty of Informatics, Shizuoka University

連性を検出し、関連のある欠陥ごとに分類することが重要となる。今回は前項で述べたように「開発、レビュー経験年数と欠陥検出可能時期」、「開発、レビュー経験年数と指摘内容」の2つの観点より欠陥の分類を行った。

2.1 検出可能時期による分類

この分類では指摘された欠陥がどの検出時期のレビューにおいて指摘可能となりうるのかの観点より分類を行う。この分類では指摘された欠陥に現時点のレビューよりも早期の段階で検出できる欠陥が含まれていた場合はレビューをより早い段階から実施することで開発工程における手戻りを減らすことを目的としてこの分類を設定した。

この目的を達成するために指摘された欠陥を3種類に分類し、その基準を示したものが表1である。早期に検出可能な欠陥は仕様書に記述した段階で検出できるものであり、メッセージの誤字脱字など、その機能、モジュールの中で完結しているため、仕様書の作成途中で検出できるため、早期のレビューで検出できるものとして分類した。この欠陥が仕様書の多くを占めるのであればレビューの早期実施によって開発コストの低減が見込まれる。

次に検出可能時期が中期のものについては機能の抜けなどが該当し、仕様書を作成している途中ではまだ仕様の検討中などで機能を記載できないのかそれとも必要にもかかわらず書き漏らしているのかが明らかでないため、仕様書の特定の機能群が作成完了した段階で検出可能なものと定義する。

最後に検出可能時期が終期のものについてはシステムの要件の誤り、処理速度の想定が不適切などのシステムの全容が確定しなければ指摘できない欠陥が該当する。そのため、これらの欠陥はシステムの全容が確定するため早期のレビューによって検出することは困難であると考え。そのため、この分類に該当する欠陥がレビューで指摘される事が多い場合はレビューの早期実施による手戻り軽減を見込むことは困難である。

なお、今回の調査においてレビューにて指摘された使い勝手に対する提案などは欠陥ではないため集計対象から除外した。

2.2 指摘内容による分類

この分類では指摘された欠陥の内容を4つに分類する。この分類では前節の分類において検出可能時期に着目したのに対して指摘された内容に重点を置いて分類を行った。その分類結果をまとめたものが図11である。

- カテゴリ A に分類した欠陥については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が該当する。
- カテゴリ B に分類した欠陥は機能の抜け、操作手順の

欠陥か?	検出可能な段階	分類基準	具体例
欠陥	序盤	仕様書に記述した段階で誤りを検出できる	誤字脱字
	中盤	モジュールが完成した段階でのレビューで検出できる	機能の抜け
	終盤	システムが完成した段階で検出できる	要件の誤り
欠陥ではない	—	その指摘された部分はシステムの動作に影響しない	使いやすさに関する提案など

表 1 欠陥の分類

Table 1 Category of defects

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
設計工程	基本設計
ページ数	3 ページ

表 2 対象ドキュメント A のデータ

Table 2 The outline of "document A"

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	勉強会予約システム
設計工程	基本設計
ページ数	3 ページ

表 3 対象ドキュメント B のデータ

Table 3 The outline of "document B"

漏れなどのシステム内の抜け、誤りを含む指摘が該当する。

- カテゴリ C に分類した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 D 分類される欠陥は誤字脱字であり、仕様書内の文章の誤字脱字、システムのメッセージでの誤字脱字などが含まれる。

3. 実験

調査ではシステムの仕様書に含まれる欠陥を検出する傾向を調査するため、ソフトウェア開発実務者に仕様書のレビューを依頼し、指摘された欠陥を分類した。

(1) 対象とデータ

対象とするデータについては表2、表3、図3の通りである。また、対象者の開発、レビュー経験年数構成比は図1、図2の通りである。仕様書に定義されている機能の例を表5にまとめた。

(2) 調査手順

所属	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部門
対象人数	77人
欠陥検出件数	1124件

表 4 対象者のプロフィール
 Table 4 Inspectors's profile

テニスコートの予約システム		
機能大分類	画面名称	機能小分類
予約機能	予約画面	
	空き状況確認画面	
...

表 5 仕様書で定義された機能の一部
 Table 5 A part of function defined by specifications

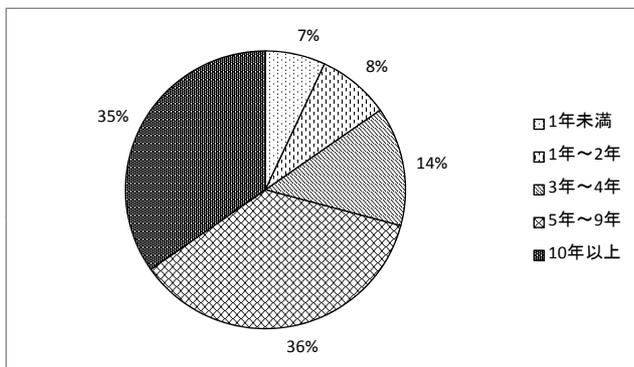


図 1 アンケート対象者の開発経験年数の構成比

Fig. 1 Percentage of development experience years

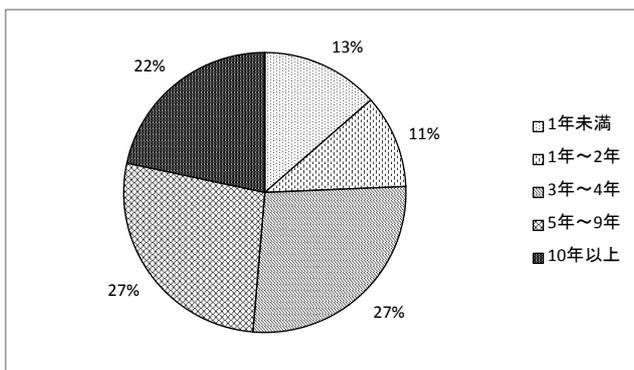


図 2 アンケート対象者のレビュー経験年数の構成比

Fig. 2 Percentage of inspection experience years

以下の手順に従って調査を行った。また、アンケートでは開発経験年数とレビュー経験年数が検出する欠陥の種別に影響に着目して設問を作成した。

- 2社の社内システムに用いる Web アプリケーションの基本設計段階の仕様書を配布する。なお、この仕様書には様々な欠陥が存在している。

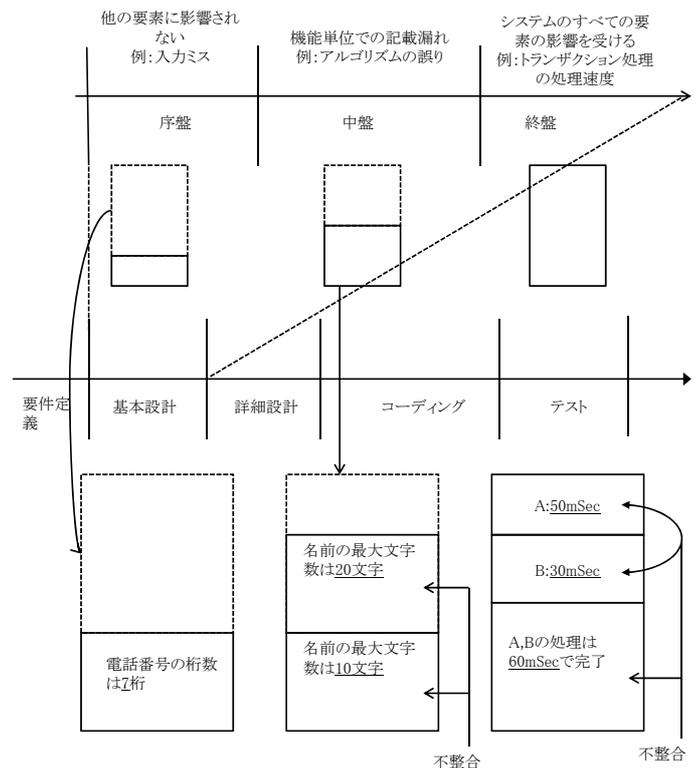


図 3 対象とした仕様書と検出される欠陥の例

Fig. 3 Example of detected defects

4. 調査結果

(a) 対象者のプロフィール

レビューアのプロフィールを表 4 にまとめた。なお、図 4, 図 5 のグラフにおける凡例は表 9 の通りである。

(b) 検出時期による分類

表 7, 表 8 の結果より、開発経験年数、レビュー経験年数共に経験年数が増加するにつれて検出する欠陥の種別の偏りが減少していくことが示されている。また、開発件数、レビュー年数が少ない場合は終期に検出できる欠陥の割合が多いことがこの表より読み取れる。

さらに、図 4 より、開発経験年数が増加するに従って、最大値と最小値の差が減少し、これは開発経験の増加と欠陥検出の能力の個人差の減少に関連があることを示している。また表 7 のデータからも開発経験年数の増加に伴って検出する欠陥の傾向のばらつきが少なくなっていることが示されている。

次に、中期に発見できる欠陥に着目すると欠陥の平均検出件数は同じ位置にあることより、中期に発見できる欠陥については経験年数は大きく影響しないことがこのグラフより明らかとなった。

早期に発見できる欠陥については平均値は開発経験年数の増加に従って減少していく結果となっている。開発経験年数が少ない対象者について注目すると、25%~75%のばらつきが少なく、個人差が少ないものと考えられる。

検出フェーズ	件数
早期	329(29%)
中期	366(33%)
終期	429(38%)

表 6 検出段階と検出された欠陥の割合

Table 6 Ratio of detection stages and detected defects

	早期	中期	終期
1年未満	31%	25%	45%
1年～2年	11%	39%	51%
3年～4年	33%	26%	41%
5年～9年	28%	38%	35%
10年以上	32%	32%	36%

表 7 開発経験年数と検出した欠陥の件数の割合

Table 7 Ratio of the number of years that experienced development and detected defects

	早期	中期	終期
1年未満	19%	30%	51%
1年～2年	24%	29%	47%
3年～4年	33%	34%	33%
5年～9年	35%	35%	30%
10年以上	37%	33%	30%

表 8 レビュー経験年数と検出した欠陥の件数の割合

Table 8 Ratio of the number of years that experienced inspection and detected defects

レビュー年数とは種別ごとの欠陥の検出件数を示したグラフである図 5 から、レビュー経験年数が 10 年を超える熟練者は終期の欠陥の検出には個人差があるが、初期に検出可能な欠陥については他の種別の欠陥と比較してばらつきが少ないことが読み取れる。次に、早期に発見できる欠陥とレビュー経験年数の関係について着目すると、平均は年数に影響していないことが示されている。

担当分野と検出する欠陥の種別の関連については図 6 より、エンタープライズ並びにソフトウェア製品を担当している担当者は終期に検出可能な欠陥を多く指摘し、組み込みを担当している担当者からの指摘では中期に検出できる欠陥を多く指摘していることが読み取れる。また、ソフトウェア製品を担当している対象者の検出した欠陥の平均を種別ごとに着目すると初期に検出できる欠陥の指摘が突出している。

(c) 欠陥の内容に着目した分類

前節では欠陥に対して検出可能な時期という観点においてカテゴリ分けを試みた。本節では欠陥を欠陥の内容に着目して分類を行うことで、開発経験年数、レビュー経験年数と検出する欠陥種別の間に関係性を見出すことが可能であるのか調査した。なおここでは検出時期での分析の際に利用したデータの中から 31 人分のデータを抽出したものをデータとして利用した。新たなカテゴリ分類については

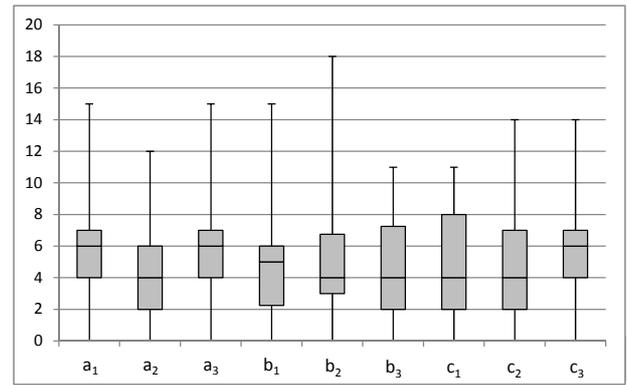


図 4 開発経験年数と検出された欠陥件数

Fig. 4 The number of years that experienced development and the number of detected defects

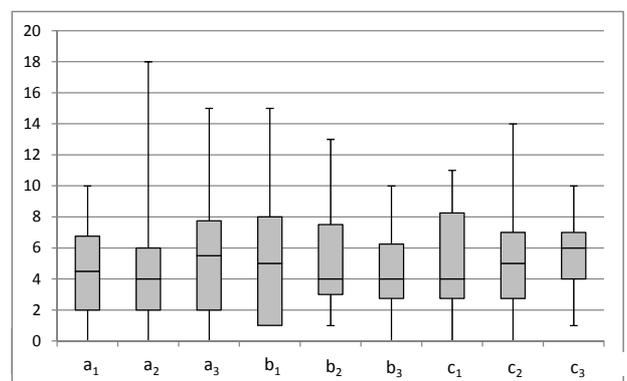


図 5 レビュー年数と検出された欠陥件数

Fig. 5 The number of years that experienced inspection and the number of detected defects

記号	経験年数	検出フェーズ
a ₁	1年～4年	早期
a ₂		中期
a ₃		終期
b ₁	5年～9年	早期
b ₂		中期
b ₃		終期
c ₁	10年以上	早期
c ₂		中期
c ₃		終期

表 9 図 4, 図 5 にてあげたグラフの凡例

Table 9 The legend of fig.4 and fig.5

所属	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部門
対象人数	31人
欠陥検出件数	508件

表 10 内容に着目した観点による分析の際の対象者のプロフィール

Table 10 The profile in the case of analysis by the contents of indication

表 11, 対象としたデータは表 10 の通りである。

まずはじめに開発経験年数と検出する欠陥種別の間にお

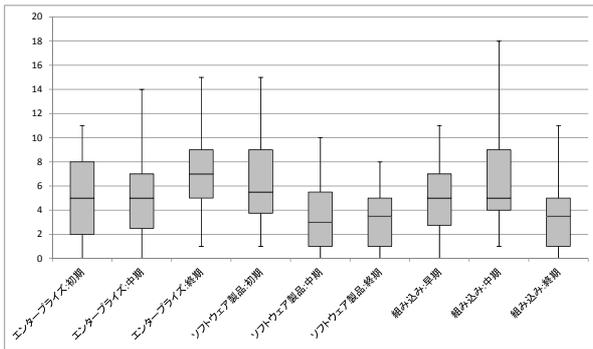


図 6 対象者の開発担当分野と検出された欠陥件数

Fig. 6 The field of development and the number of detected defects

ける関連性について調査した結果を図 7, 図 8, 図 9, 図 10 にまとめた。

カテゴリ A の欠陥については開発経験年数が 4 年未満の対象者の欠陥検出数においては、最大は少ないが中央値は最も高いという結果となった(図 7)。カテゴリ B の欠陥については最大値、中央値より、総合的に 4 年未満の欠陥検出率も最も高い結果となった。カテゴリ C の欠陥については 4 年未満と 10 年以上の経験年数において最大、最小、平均は同様の数値であるが、10 年以上の場合は 4 年未満より検出件数にばらつきが大きい結果となった。カテゴリ D の欠陥においては 4 年未満よりも 10 年以上の経験年数のほうが中央値も高く、検出件数のばらつきはカテゴリ C とは対照的に少ない結果となった。

次にレビュー経験年数と検出する欠陥種別の間における関連性についても分析を行ない、図 11, 図 12, 図 13, 図 14 にまとめた。

カテゴリ A の欠陥については開発年数の場合と同様の傾向が見られ、4 年未満の欠陥検出数が最も高い結果となった。カテゴリ B については、検出数の中央値は 4 年未満が高い結果となり、最大検出数は 10 年以上が高い結果となった。カテゴリ C については 4 年未満と 10 年以上では検出数の最大、最小は変わらず、10 年以上の対象者のほうが検出数のばらつきが大きい結果となった。カテゴリ D の欠陥については、10 年以上の対象者が検出数の中央値が最も高く、検出件数も高い位置に分布している結果となった。このカテゴリについては欠陥の検出数とレビューの経験年数は比例関係にあり、経験年数の増加によって検出件数も高い位置に移動している結果となっている。また、最大値、最小値についてはどの区分においても同じ数値である。

5. 考察

5.1 検出時期

検出時期と開発経験年数で分類した(図 4)の中で検出可

カテゴリ	カテゴリ概要	具体例
A	ソフトウェアと外界の情報を検討した欠陥	ユーザーの持つ情報との不整合
B	機能、操作手順の抜け、漏れ	必要なボタンが定義されていない
C	誤り	仕様書内で最大文字数の不整合
D	誤字	誤字

表 11 指摘内容に着目したカテゴリ分類

Table 11 Classification by indication contents

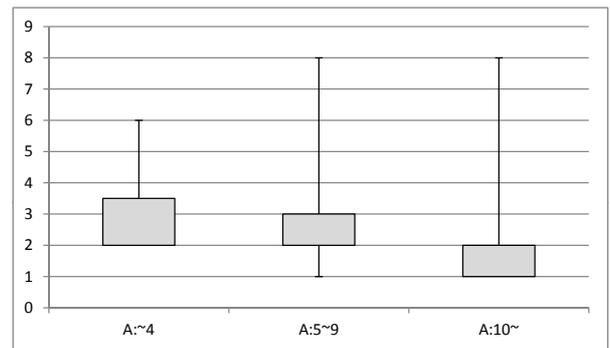


図 7 カテゴリ A に分類される欠陥と開発経験年数の関連

Fig. 7 Relation of the defect classified into the category A, and the number of years that experienced development

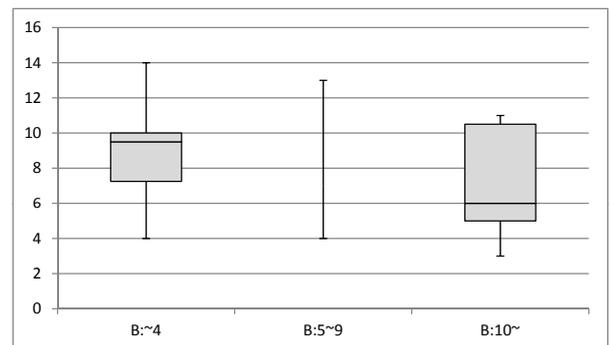


図 8 カテゴリ B に分類される欠陥と開発経験年数の関連

Fig. 8 Relation of the defect classified into the category B, and the number of years that experienced development

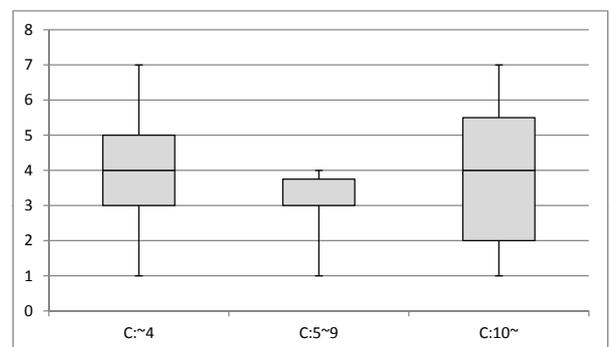


図 9 カテゴリ C に分類される欠陥と開発経験年数の関連

Fig. 9 Relation of the defect classified into the category C, and the number of years that experienced development

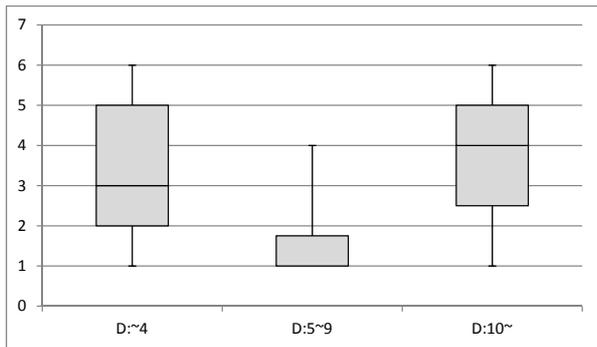


図 10 カテゴリ D に分類される欠陥と開発経験年数の関連
Fig. 10 Relation of the defect classified into the category D, and the number of years that experienced development

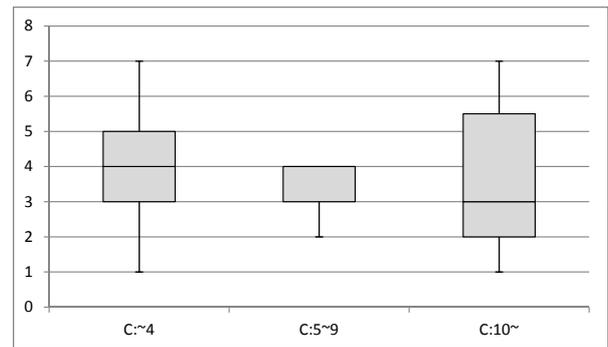


図 13 カテゴリ C に分類される欠陥とレビュー経験年数の関連
Fig. 13 Relation of the defect classified into the category C, and the number of years that experienced inspection

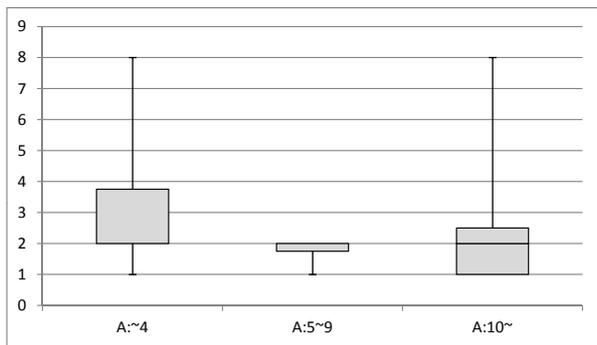


図 11 カテゴリ A に分類される欠陥とレビュー経験年数の関連
Fig. 11 Relation of the defect classified into the category A, and the number of years that experienced inspection

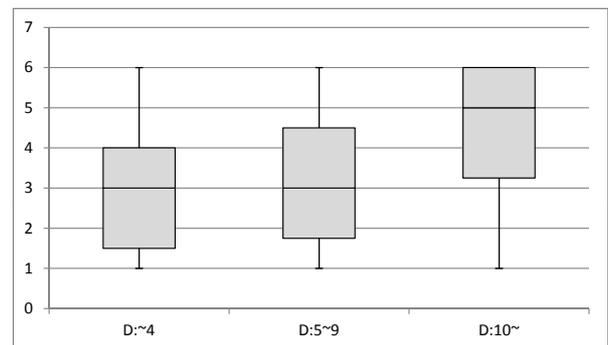


図 14 カテゴリ D に分類される欠陥とレビュー経験年数の関連
Fig. 14 Relation of the defect classified into the category D, and the number of years that experienced inspection

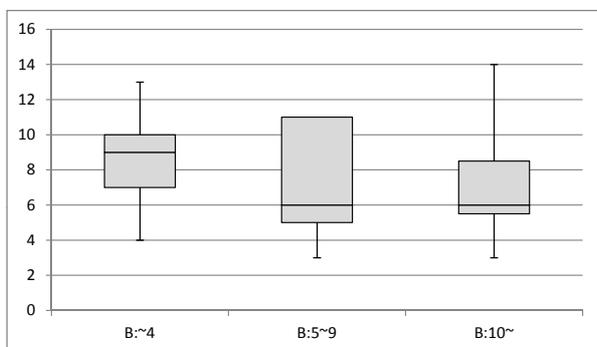


図 12 カテゴリ B に分類される欠陥とレビュー経験年数の関連
Fig. 12 Relation of the defect classified into the category B, and the number of years that experienced inspection

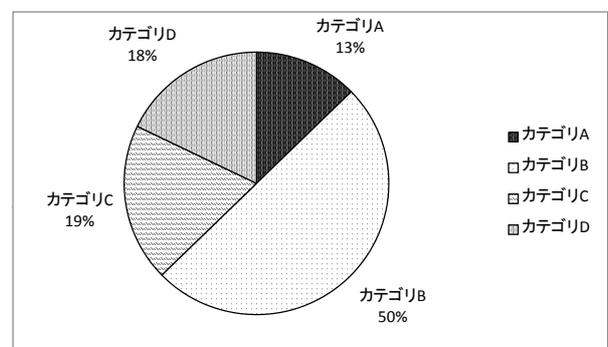


図 15 開発経験年数が 5 年未満の対象者の欠陥検出種別の割合
Fig. 15 Ratio of defect detection classification with less than five years in development years of experience

能フェーズが早期のものは経験年数が増加するにつれて分布のばらつきが大きくなり、平均も減少しているが、これは経験が増加することで対象者の中で欠陥と判定するための基準が確立され、その基準に差が開いていくため、誤字脱字を欠陥として指摘する場合の基準が対象者ごとによって異なることとなりそれがこのばらつきの増加に現れているものと考えられる。

検出された欠陥の内、早期、中期に検出可能な欠陥の割合(図 6)に着目すると 62%と半分以上を占めており、レビューをより早い段階でも実施する [3] ことで欠陥を早期

に修正できる可能性が高いことが示されている。

5.2 指摘内容

開発経験年数、レビュー経験年数の増加にともなって検出可能な欠陥の種別が「序盤に検出可能な欠陥」、「誤字、脱字」から「後期に検出可能な欠陥」、「外界の影響を受けられる欠陥」に変化していくというわけではなかった。

検出時期による分類ではなく、開発経験年数ごとの一人あたりの検出件数をグラフ化したものが図 15、図 16、図 17 であるが、このグラフでも開発経験年数によって検出

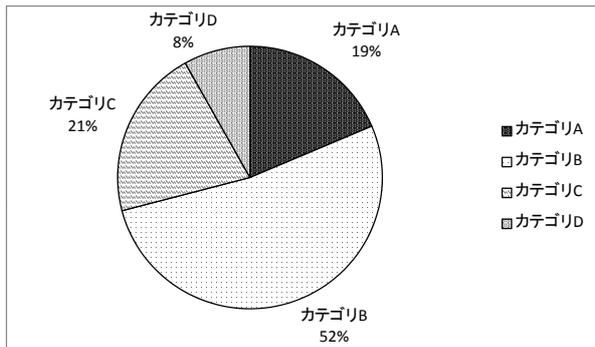


図 16 開発経験年数が 5 年以上 10 年未満の対象者の欠陥検出種別の割合

Fig. 16 Ratio of defect detection classification with five years or more to less than ten years in development years of experience

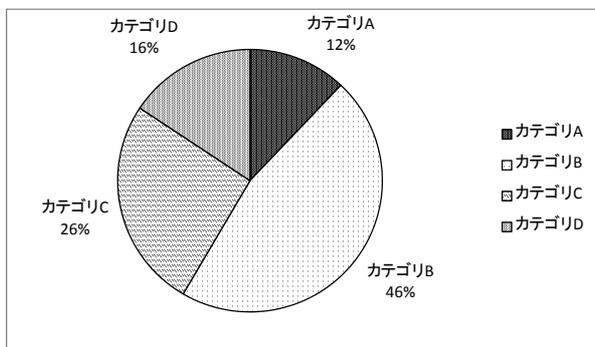


図 17 開発経験年数が 10 年以上の対象者の欠陥検出種別の割合

Fig. 17 Ratio of defect detection classification with ten years or more in development years of experience

する欠陥が外界に関係するものに変化する結果とはならなかった。

新たにカテゴリを定義し、開発年数と検出する欠陥のカテゴリの関係性をグラフにした中でも、特にカテゴリ A として分類した「ソフトウェアと外界の情報を考慮した欠陥」について、中央値と最大値の差が経験年数の増加とともに非常に大きなものになっているが、これはその対象とした実務者のスキルが大きく影響するために、最初はあまりスキルに差がないために欠陥の検出量の差も大きくないが、経験を積みスキルに差が出ることで欠陥を指摘する際にも影響を与えたのではないかと考える。

本研究で用いた仕様書は画面イメージを含めて A4 で 3 ページと小規模なものであるため、大規模な仕様書となった場合に、同じ傾向となるかどうかは調査と議論を積み重ねる必要がある。

6. おわりに

本稿では、レビューの実施時期や参加メンバーの選定に役立てることを目的として、検出可能時期と指摘内容による分類を検討し、実務者による欠陥指摘を分類し、傾向を分析した。指摘内容の分類は、ソフトウェアが実現する機能

を利用する利用者や他システムとの連携をはじめとした外界との不整合、記述の漏れ、誤り、体裁や表記のミスの 4 つである。検出可能時期は欠陥の混入時点ですぐに検出できるかどうかを基準とし、記載とともに検出できる、関連する他の部分が記載されれば検出できる、全てが記載されなければ検出できない、の 3 つであり、これらの割合とレビューアの経験に関係があるかを調べた。

開発経験、レビュー経験、開発分野の異なる実務者によるレビュー指摘内容を分類した結果、外界の情報が影響する欠陥では、レビューアのスキルが大きく影響する結果が得られた。また、序盤、終盤に検出できる欠陥は開発の経験年数との関連があることを示す結果が得られた。

今後は分析対象のデータをさらに増やして傾向を検討する予定である。今回の指摘内容は Web システムの仕様書をレビューした結果であるが、他のシステムについても同様の傾向があるか調べることも今後の課題の 1 つである。

謝辞 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] A. Porter, L. Votta, 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)
- [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)