

研究コース2（レビュー品質の可視化チーム）

重大欠陥予測手法を活用したレビュー品質評価技法の提案

～既存レビュー記録とプロジェクト特性から第三者がレビュー品質を可視化～

A proposal of evaluation technique of review quality by using a critical defect prediction method

研究員：高橋 喜哉（株式会社日立製作所）
平井 由貴美（株式会社インテック）
福田 秀樹（T I S 株式会社）
横屋 司（ソーバル株式会社）
主査：中谷 一樹（T I S 株式会社）
副主査：上田 裕之（株式会社D T S インサイト）
アドバイザー：安達 賢二（株式会社H B A）

研究概要

重大欠陥を上流工程から発見し対処するために、成果物レビューが行われているが、しばしば後工程に重大欠陥が流出する事態が起きている。そうしたレビューを特定できれば品質強化も効果的にできるが、既存のメトリクスでは、第三者がレビュー記録から重大欠陥の指摘漏れが存在するレビューを検出するのは非常に困難である。そこで筆者らは「先行研究を活用して重大欠陥混入を予測し、その欠陥がレビューによって検出できたかどうかを表すメトリクスによりレビュー品質を可視化する技法」P2DIETを提案する。筆者らが持ち寄った現場のレビュー記録を用いて検証した結果、P2DIETは重大欠陥の指摘漏れ検知に有効であるという結果を得た。

Abstract

A review of the deliverables has been made in order to discover and address serious defects from the upstream process. However, there is a situation where a serious defect is often leaked to the downstream step. It is very difficult for a third party to detect a review in which a significant defect is leaked from a review record though the quality enhancement can be done effectively if such a review is identified. Therefore, the authors propose "a technique to visualize the review quality by using the preceding research to predict the serious defect contamination and to determine whether the defect was detected by a review" P2DIET. As a result of verification using the review records brought by the researchers, the P2DIET was effective in detecting the leakage of a significant defect.

はじめに

1.1 背景

ソフトウェアの品質トラブルはプロジェクトに大きな影響を及ぼす。特に重大欠陥が後工程で見つかり手戻り工数が大きいことはよく知られたことである。これを避けるために上流工程からレビューするなどの品質向上の取り組みが行われている。しかし、レビューを実施していても、重大欠陥の後工程への流出はしばしば見受けられることから、レビューそのものの品質を向上させる必要がある。

レビューの品質を向上させるために、各種レビュー技法や観点が開発されている。しかし多忙な現場に新たな手法を導入することは多大な労力がかかり、なかなか一步を踏み出せないのが実情である。そこで筆者らは、レビューの良し悪しを可視化することからまず

始めるべきだと考えた。

レビューの良し悪しを評価するに際し、各種メトリクス（レビュー工数密度、レビュー指摘密度）が使われている。しかし基準値の許容範囲内に収まっており一見問題ないと判断される成果物や、報告上何ら問題がなさそうに見えるプロジェクトであっても、システムテスト時やリリース直後に重大欠陥を引き起こしてしまうケースがある。

このようなケースを無くすためには、実施されたレビューのうち、品質の悪いレビュー（ここでは、後工程に重大欠陥を流出させてしまうレビューを、品質の悪いレビューと定義する）を検知する必要がある。そのための方法として、第三者が全てのプロジェクトに対してレビューに参加する、設計書などの成果物を直接確認するといった方法も考えられるが、人手もかかるため現実的には困難である。そこで筆者らは品質の悪いレビューを検知できる新たな技法やメトリクスを考案することにした。

1.2 解決すべき課題

レビュー品質を可視化する上で、既存のメトリクス（レビュー工数密度、レビュー指摘密度）には、以下2点の解決すべき課題があると考ええる。

(1) 評価メトリクスに対し当該プロジェクト状況が反映されない

レビュー工数密度やレビュー指摘密度の基準となる値は、あくまで過去のプロジェクトの実績値を統計的に処理して算出したものであり、当該プロジェクトが置かれた状況を端的に反映したメトリクスとはなっていない。プロジェクトの状況の変化や人員の個性は多種多様であるため、過去データに基づくメトリクスよりもそれぞれのプロジェクトの特性に応じたメトリクスを利用する方が、より現実的に品質を評価できるものと思われる。

(2) 評価メトリクスは指摘数が中心であり、指摘一つ一つの重大度は考慮されない

レビュー指摘密度に代表されるように、既存の評価メトリクスはレビュー指摘数の多寡に注目が集まっており、指摘の重大度を考慮したレビュー品質の評価手段ではないことが多い。また、どのような種類の欠陥が流出しやすいかもあまり評価されていないのが実態である。

これらの課題を解決するために、各プロジェクトはデータを継続的に蓄積して独自の基準値を設定したり、メトリクスの値だけに頼らず、指摘の中身や実際の成果物の内容に踏み込んだ分析を行い、総合的に品質を評価したりするなどの対策を講じながら品質向上を図っている。しかし、第三者（品質保証担当など）の立場では、実際の成果物を入手して内容を確認するには多くの時間がかかり、またそれに協力するプロジェクト側も資料の準備や説明に時間を取られてしまうため、あまり得策とは言えない。この弱点を克服するには、もっと別のより効率的な方法を考える必要がある。そこで筆者らは、上記2点の課題を解決し弱点を克服するための技法に求められる要件を、以下の通りと考えた。

- ・ 第三者でも品質の悪いレビューを端的に可視化できること
- ・ 当該プロジェクトの置かれた状況を極力反映できること
- ・ 重大欠陥の流出防止に有効であること

これらを踏まえて、以下の二つのRQを設定する。

RQ1：新しい技法により、第三者がレビューの品質を可視化できるか

RQ2：新しい技法を取り入れることで、重大欠陥の流出を防止できるか

以降、2章では重大欠陥の流出防止に関する先行研究の調査結果を示し、3章では筆者らが提案する評価技法を示す。4章で提案技法に対する実験と評価考察を行い、5章で今後の課題と展望、および全体のまとめを示す。

2. 先行研究

課題の解決策として付録1（先行研究の比較結果）に掲げる10件の先行研究^{[1]~[10]}を調

査した。対象となる先行研究は、重大欠陥の検出を目的としたものを選定した。調査した結果、先行研究はそれぞれ重大欠陥の流出防止観点を導き出しているが、いずれもレビュー自体のやり方や観点の提示によってレビュー行為そのものを直接的に強化する手法であることが分かった。今回提案する技法は、レビュー行為そのものの強化ではなく、レビュー品質の可視化により、重大欠陥の流出防止を図ろうとするものである。可視化の軸となる重大欠陥の候補には、先行研究の手法による観点や重大欠陥の種類が有効であることから、これらを組み込んで活用することとした。

3. 提案

3.1 提案技法の前提

技法の適用に当たり、筆者らは以下の前提を置くことにした。

(1) レビューの定着度合い

今回提案する技法は、重大欠陥の候補をレビューでの指摘と突き合わせて分析するため、レビューが実施され、指摘内容が記録されている必要がある。

(2) 技法の活用タイミング

できるだけ上流工程で品質強化を図った方が、下流工程での手戻りを防ぎ、欠陥対処工数を圧縮できることは一般的に言われていることである。本技法もその観点から、レビューを実施したその工程で評価を行い、成果物品質の再点検をするべきレビューがあるかを確認するといった活用方法を想定している。今回の実験で筆者らは基本設計のレビューを対象としたが、より上位の要件定義、より下位の詳細設計においても活用できる技法であると考えている。

(3) 第三者の立ち位置

プロジェクトにとっての第三者は、全社の品質保証部門、事業部や現場部門付きの品質保証担当、当該プロジェクトと異なる有識者など様々存在する。本提案においては、当該プロジェクトの状況を反映した評価を実施するため、プロジェクトの状況が比較的容易に把握でき、客観的な視点を持つ立場である者（事業部や現場部門付きの品質保証担当など）を想定している。

3.2 提案する技法(P2DIET)

1.2 章の課題および 2 章の先行研究の調査を踏まえ、筆者らはレビュー品質評価技法 P2DIET (by Using Pre-Predicted Defect Indicators, Evaluation Technique of the review quality: 重大欠陥予測手法を活用したレビュー品質評価技法) を提案する。P2DIET は、予測された重大欠陥が、レビューにて適切に検出できているかをレビュー記録との照合によって可視化し評価する技法である。

以下に P2DIET の特徴を示す。

(1) 当該プロジェクトの特性を考慮

過去の統計データから導き出されたメトリクスではなく、当該プロジェクトの状況や対象成果物の特徴を評価することによって重大欠陥を予測することから、より納得性の高い評価を行うことができる。

(2) 客観的な第三者による評価

当該プロジェクトや対象成果物、レビュー記録を分析することによって評価する技法であるため、プロジェクトの当事者でなくても評価することができる。

(3) 重大欠陥の予測手法を組み合わせ活用

重大欠陥を予測する手法は先行研究でいくつか提示されている。今回提案する技法は重大欠陥の予測手法には依存しないことから、自由に手法を入れ替えたり、複数の手法を組み合わせたりすることが可能である。

(4) 評価のための新たな指標値を設定

レビューの品質を可視化するメトリクスとして、予測される重大欠陥をレビューでどれだけ検出できたかを示す、「予測重大欠陥レビュー検出率」を設定する。算出式は以下の通り。

$$\text{予測重大欠陥レビュー検出率(\%)} = \text{検出した重大欠陥項目数} \div \text{予測重大欠陥項目数} \times 100 \quad \dots\text{式①}$$

このメトリクスの値が高いほど、重大欠陥を阻止できた品質の高いレビューであると言えることができ、値が低いほど、重大欠陥の見逃しの可能性が高い、手を打つべきレビューであると言えることができる。

3.3 P2DIET の適用手順

P2DIET の手順を図 1 に示す。手順の詳細は(1)～(4)の通り。

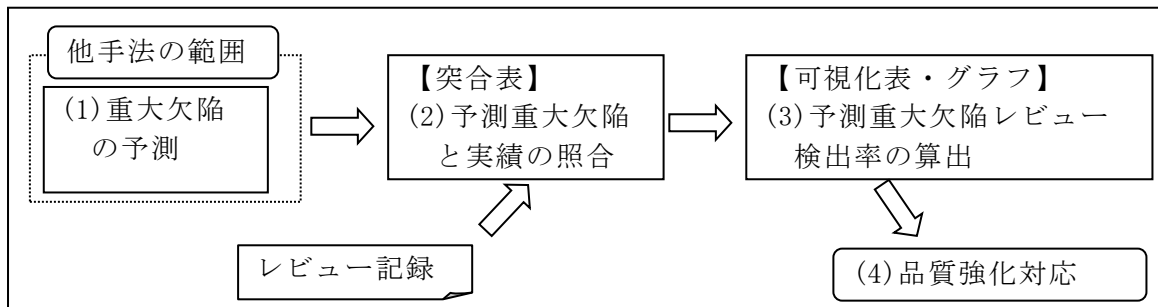


図 1 P2DIET の実施フロー

(1) 重大欠陥の予測

(1)-1 重大欠陥予測手法の選定

先行研究^{[1]~[10]}で提案されたものの中から、付録 1. に挙げた下記の 6 つの観点を 3 段階で点数化し、総合得点が高いものを選出する。

- ・準備の容易さ
- ・特別な知識やスキルの必要性
- ・予測できる重大欠陥の信頼性
- ・プロジェクトの特性の反映度
- ・実施の容易さ
- ・結果分析の容易さ

付録 1. で比較検討した通り、総合得点が最も高かったのは D2BOCs 法^[1]であった。D2BOCs 法は、最も予測される重大欠陥種類が具体的であり、対象成果物に潜む重大欠陥を論理的に抽出できるという特徴がある。本論文では D2BOCs 法を採用しているが、他の手法であっても重大欠陥を導け、プロジェクトの置かれた状況を反映できる手法であれば適用は可能である。

(1)-2 重大欠陥予測手法を用いた重大欠陥の予測

本論文で採用した D2BOCs 法では、成果物作成者の置かれた状況に関する質問項目と、成果物そのものの記載の特徴をチェックシートによりチェックし、作成者が陥っていたと推測される認知バイアスを特定することにより、混入した可能性が高い重大欠陥種類を予測する。

(2) 予測重大欠陥と実績の照合

(1)で予測した重大欠陥と、レビュー記録から読み取れる実際に検出できた重大欠陥を並べてマークし、突合表を作成する。

(3) 予測重大欠陥レビュー検出率の算出

式①にて予測重大欠陥レビュー検出率を導く。予測数と実績数、および検出率をレビューごとに並べ、可視化表を作成する。可視化表をグラフ化し、各々のレビューの品質を可視化する。(図 2 に突合表、可視化表・グラフのイメージを示す)

(4) 品質強化対応 (予測重大欠陥レビュー検出率の低い成果物レビューの再点検)

予測重大欠陥レビュー検出率が低いものは、欠陥の見逃しがあった可能性が高いことから、対象成果物そのものの再点検、および下流工程のテストで見逃し対象の欠陥が検出されていないかの確認を行う。全てのレビュー観点で総点検をする必要はなく、予測と実績が乖離していた重大欠陥項目に焦点を当てて実施すれば良い。

欠陥の傾向	重大欠陥種類	レビュー指摘予実	レビュー	
			予測	実績
欠陥 (対応されず)	C-03	必要な記載がない の条件が漏れる	★	○
欠陥 (考慮されず)	C-04	例外ケースの考慮が漏れる		
	C-05	セキュリティ面の考慮が漏れる		
	C-06	類似した機能において、機能独自の仕様が記載されていない	★	
	C-07	非連携部分の情報が欠落する		
	C-08	影響範囲の対応が抜ける		
	C-09	性能に関する考慮がされない		
矛盾	C-10	InputとOutputの内容に齟齬がある	★	○
	C-11	機能に関する影響範囲間の辻褄が合っていない	★	
未対応	C-12	課題修正箇所に関連する箇所の変更対応がされない		
	C-13	類似機能の処理が流用元のままと なっている	★	○
読解困難	C-14	用語の使い方を間違えている・意味 を取り違えている	★	
	C-15	誤装袋の誘発	★	
重大欠陥種類レビュー検出率			7	3
			43%	

レビュー	予測	実績	検出率
A	7	3	43%
B	4	4	44%
C	2	2	22%
D	3	3	30%
E	9	2	22%
F	9	6	67%
G	9	3	33%
H	9	2	22%
計	71	25	35%

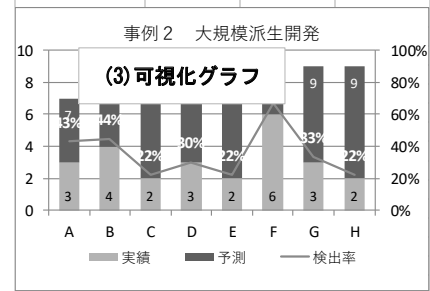


図2 突合表、可視化表・グラフ (イメージ)

4. 実験と評価

4.1 実験手順

3章で提案したP2DIETの有効性を確かめるために、3.3項の手順に沿って実験を行った。

4.2 実験結果

筆者らの分析対象プロジェクトの特性を表1に、実験結果(可視化グラフ)を図3に記す。(詳細は付録2.研究員の実験データ詳細を参照されたい)

表1 分析対象プロジェクトの特性

事例	対象業種	規模	開発スタイル・特徴
事例1	金融・カード系	大規模(200人月超)	短納期新規開発(横展開)
事例2	製造業系	大規模	派生開発
事例3	交通系	大規模	派生開発
事例4	公共系	小規模	保守, 新規開発(横展開), 新規開発

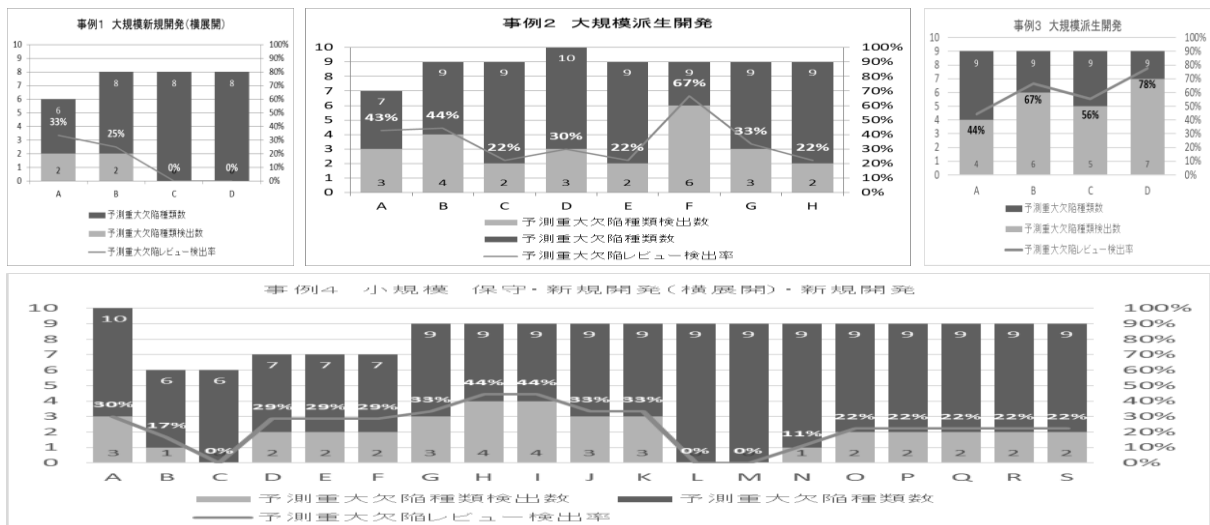


図3 分析対象プロジェクト事例の可視化グラフ

4.2.1 予測重大欠陥レビュー検出率の傾向

(1) 事例 1

この事例では予測重大欠陥レビュー検出率が全体的に低かった。プロジェクトの特性として大規模かつ短納期開発であることから、他社資産を活用することで効率化を図ったが、レビューは有識者の人数が限られていたため手分けして実施したことから、レビューにより指摘レベルにバラつきが見られた。また、成果物作成者が比較的若手で基本設計からアサインされたことから、要件定義の内容やプロジェクト自体の前提への理解が不十分で例外処理の作り込みが十分でなかった。後続工程の品質を調査してみると、結合テスト工程にて、データバリエーションに起因する例外データの考慮漏れが発見されており、重大欠陥の流出が認められた。

(2) 事例 2

この事例も予測重大欠陥レビュー検出率は全体的に低いですが、予測重大欠陥レビュー検出率が高いレビューも含めて、後工程で多数の欠陥が検出されている。本事例ではレビュー品質の良し悪しと重大欠陥の流出との因果関係は認められなかった。

(3) 事例 3

この事例では全てのレビューについて追跡調査を実施したところ、業務 D と比較して予測重大欠陥レビュー検出率が低い業務 A・B・C について、重大欠陥の流出が認められた。流出した欠陥の種類としては、未経験箇所の条件漏れ、例外ケースの考慮漏れ、類似機能の独自仕様部分反映漏れ、上位設計書との仕様齟齬、機能間の整合性不備などが挙げられる。

(4) 事例 4

この事例も予測重大欠陥レビュー検出率が全体的に低めである。再点検を指示したが新たな欠陥は検出されなかった。業務 A は保守、B～C、D～F はパッケージシステムの横展開プロジェクトであり、帳票のレイアウト変更など軽微な要件の設計書レビューであるため、成果物の品質自体が良く、予測重大欠陥レビュー検出率は低いものの問題はなかった。業務 G～S は類似機能が多いため最初は指摘が多く出たものの、指摘に対応した設計書を流用した設計書では指摘が少なくなった。ただし、業務 J のみ、例外ケースの記述漏れをレビューで指摘できず、後続工程への流出が認められた。

4.2.2 考察

1.2 章で提示した RQ に対し、実験結果を踏まえて考察を示す。

(1) RQ1：新しい技法により、第三者がレビューの品質を可視化できるか

実験結果から、P2DIET により予測重大欠陥レビュー検出率が低いレビューを可視化することができ、全ての事例で実際に後続工程に欠陥が流出している、品質の悪いレビューが含まれていることが確認できた。従って、P2DIET はレビュー品質の可視化に有効であるといえる。ただし、小規模プロジェクトの場合や成果物品質が元々高いプロジェクトの場合は、そもそも混入する欠陥が少ないため、本技法のメトリクスだけでは顕著な差が出にくいことも分かった。

(2) RQ2：新しい技法を取り入れることで、重大欠陥の流出を防止できるか

実験結果から、全ての事例において、予測重大欠陥のうちレビューで指摘できず後続工程に流出したものが認められた。該当のプロジェクトは後続工程まで進んだプロジェクトであったため設計工程内での対応は取れなかったが、仮に設計段階でレビューの再点検を指示していれば重大欠陥が発見できたと言え、本技法で重大欠陥の流出阻止ができる可能性を示している。また、事例 2 では予測されていない重大欠陥種類が後続工程に流出している事象が見受けられた。このことは、重大欠陥種類の予測自体にも改善の余地があることを示している。

なお、予測重大欠陥レビュー検出率が低い場合、「レビューの品質が悪くて指摘し切れ

なかった」または「成果物の品質が良くてレビューで指摘が出なかった」の両方の状況が考えられる。再点検の結果、レビューの質が悪ければ漏れていた観点での欠陥が見つかり修正ができ、成果物の品質が元から高ければそれはそれで良い。いずれにせよ本技法を用いて予測重大欠陥レビュー検出率の低いレビューの再点検をすることが、成果物の品質を高めることに繋がる。

5. おわりに

5.1. 今後の課題と展望

本研究における今後の課題は以下の通りである。

(1) 重大欠陥の予測精度

本技法は第三者が成果物作成者の置かれた状況や、成果物そのものに基づいて重大欠陥の混入を予測し、レビュー記録にてその欠陥を指摘できているかを可視化する技法であることから、第三者がいかに対象成果物の作者の状況や成果物そのものを正確に把握できるかにより、予測精度にバラツキの生じる可能性がある。従って、実際の適用時にはできる限り現場の状況を知り得る立場の第三者にて評価することが望ましい。

(2) 他の重大欠陥予測手法の適用可能性

本技法の実験と評価では D2BOCs 法^[1]を用いた重大欠陥予測を採用したが、他の重大欠陥予測も活用できる可能性がある。どの手法を用いるのが、より精度の高い重大欠陥予測ができるかについては、比較検証を継続する必要がある。

(3) 検証範囲の特殊性

本技法は筆者らの持ち寄ったデータにおいては確からしさを実証できた事例があるが、他の領域や成果物においては必ずしも当てはまるとは限らない。今後の追実験や他案件への適用を通じ、検証を継続していく必要がある。

(4) 閾値の妥当性

筆者らは調査したレビュー記録間で、予測重大欠陥レビュー検出率の顕著な差異を見出し、閾値を設定することも目論んでいたが、明確な閾値の設定には至らなかった。さらに顕著な差異を識別できる新たな条件がないか、検証データを蓄積していく必要がある。

5.2. まとめ

一般的に利用されているレビュー品質を評価するメトリクスの基準値は、過去の類似プロジェクトでの実績を基に算出されている場合が多い。本研究では対象とするプロジェクトの状況や特性を考慮した上で、レビューが効果的に実施されていたかを評価する新たなメトリクス「予測重大欠陥レビュー検出率」とその評価技法（P2DIET）を提案した。

P2DIET は混入が予測される重大欠陥とレビュー記録の内容を照合することにより、レビューにおける重大欠陥の流出可能性を可視化する技法である。プロジェクト外の第三者でも評価可能であることから、開発現場への追加負担なく実装が可能である。本技法の実験と評価では、混入が予測される重大欠陥の種類を抽出する手法として、10 件の先行研究^{[1]~[10]}のうち D2BOCs 法^[1]を選択し、重大欠陥種類の予測を行ってレビュー品質の可視化と重大欠陥の流出を観測できた。

開発中プロジェクトの次工程への移行あるいは出荷時におけるソフトウェアの品質判定で本技法を適用することにより、重大欠陥の指摘漏れを起こしたレビューを再点検し、必要に応じて品質強化策を打つことで、成果物品質の向上が期待できる。

なお、本研究では、筆者らの所属する企業の事例に対する評価までで、実際の現場への導入までは試行できていない。従って、現時点ではあくまでも「可視化すれば再点検の必要性をより客観的に示せ、現場の理解も得られるだろう」と推測できるレベルでしかない。

予測結果をレビューアに事前に渡してレビュー自体の検出力を直接的に向上する手段もあるが、導入のハードルが高いことから、まずは本技法により強化すべき（再点検すべき）レビューを絞り込むことから始めるのが、しらみ潰しでなく、効率的かつ少ない現場負荷で品質強化の効果を出すことに繋がると考えられる。本技法を活用しレビューの再点検を繰り返していくことで、品質の悪いレビューを低減していくことが期待できる。

参考文献

[1] 久禮 尚, 小林 享, 櫻井 瑞穂, 原田 和典, 湯川 健, 中谷 一樹, 上田 裕之, 安達 賢二, 「作成者の認知バイアスに着目したレビュー手法の提案」, SQiP 研究会第 33 年度 (2017 年度) 研究コース 2 (思考チーム), 2017

[2] 川合 大之, 奥山 剛, 上野 直樹, 會見 知史, 菅野 良太, 細川 宣啓, 永田 敦, 藤原 雅明, 森崎 修司, 「検出難易度の高い欠陥を検出するレビュー観点の提案」, SQiP 研究会第 27 年度 (2011 年度) 研究コース 3 (第 2 グループ), 2011

[3] 山口 友紀, 豊泉 大介, 吉田 憲人, 佐々木 明, 外山 泰久, 木村 敏康, 細川 宣啓, 永田 敦, 藤原 雅明, 森崎 修司, 「重大欠陥を効率よく検出するレビュー手法の提案と有効性の実験報告」, SQiP 研究会第 28 年度 (2012 年度) 研究コース 3 (陸グループ), 2012

[4] 小田切 勇人, 近藤 忍, 藤崎 祐美子, 細川 宣啓, 永田 敦, 藤原 雅明, 森崎 修司, 「ビジネスリスクに直結するレビューポイント導出方法の提案」, SQiP 研究会第 28 年度 (2012 年度) 研究コース 3 (海グループ), 2012

[5] 芦田 直之, 篠崎 悦郎, 仁藤 千博, 細川 宣啓, 永田 敦, 藤原 雅明, 森崎 修司, 「重大欠陥検出に集中するためのレビューポイントの導出方法の提案」 SQiP 研究会第 29 年度 (2013 年度) 第 3 分科会 (0 チーム), 2013

[6] 池戸 春樹, 高橋 信弘, 田中 賢太郎, 塚本 悠仁, 山本 千絵, 中谷 一樹, 原 佑貴子, 上田 裕之, 森崎 修司 「重大欠陥を早期是正するレビュー手法 3 分割レビューの提案」, SQiP 研究会第 30 年度 (2014 年度) 研究コース 3 (チーム PTR), 2014

[7] 加賀 讓, 北野 宗之, 白幡 千香, 安田 聡美, 中谷 一樹, 上田 裕之, 原 佑貴子, 「SBR 法 (ステルスベースドレビュー手法) の提案」, SQiP 研究会第 31 年度 (2015 年度) 研究コース 3 (餡餅チーム), 2014

[8] 諏訪 博紀, 中谷 一樹, 田邊 哉好, 末次 努, 森崎 一邦, 細川 宣啓, 永田 敦, 森崎 修司, 「間接的メトリクスを用いて欠陥予測を行うレビュー方法の提案」, SQiP 研究会第 26 年度 (2010 年度) 第 3 分科会 (A チーム), 2010

[9] 上田 裕之, 高橋 功, 高橋 実雄, 中谷 一樹, 細川 宣啓, 永田 敦, 森崎 修司, 「HDR 法: 仮説駆動型レビュー手法の提案 - HDR 法の実践による生産性と品質の同時向上 -」, SQiP 研究会第 28 年度 (2012 年度) 第 3 分科会 (A チーム), 2012

[10] 佐津 共淑, 高橋 信弘, 高橋 美和子, 延原 敦, 中谷 一樹, 上田 裕之, 原 佑貴子, 「欠陥パターンと検出テクニックを用いたレビュー手法の提案とそのトレーニング教材の有効性評価の報告」, SQiP 研究会第 31 年度 (2015 年度) 第 3 分科会 (チョコチーム), 2015