

Inspeção de Qualidade em Descrições de Casos de Uso: uma Avaliação Experimental em um Projeto Real

Paulo Sérgio Medeiros dos Santos e Guilherme Horta Travassos

COPPE/UFRJ, Universidade Federal do Rio de Janeiro, Brasil

Caixa Postal 68511 – CEP 21945-970 – Rio de Janeiro, RJ

{pasemes,ght}@cos.ufrj.br

Abstract. *This article describes an experimental evaluation of a checklist based inspection technique proposed in the literature and its insertion in a real project. The evaluation was conducted using a strategy that combines two experimental study types. First, a controlled study with graduate students aiming at preliminarily evaluate the checklist technique and, then, its observation in a real project through the action research methodology. In both studies the results were favorable to the checklist technique when compared to the ad-hoc approach regarding the number of identified defects – no difference was identified in relation to the time spent variable. A second important result was the observation, only on the real project study, of a significant performance difference among the experienced and inexperienced inspectors. Based on the analysis of the field research actions we have the indication that the experience factor can be preponderant in checklist based inspections.*

Resumo. *Este artigo descreve uma avaliação experimental de uma técnica de inspeção de casos de uso baseada em checklist proposta na literatura e a sua inserção em um projeto de software real. A avaliação foi conduzida por meio de uma estratégia que combina dois tipos de estudos experimentais. Primeiro, um estudo controlado com estudantes de pós-graduação, objetivando avaliar preliminarmente a técnica checklist. Posteriormente, em um projeto real, onde se observou a sua utilização por meio da metodologia da pesquisa-ação. Em ambos os estudos, os resultados foram favoráveis à técnica checklist em relação à abordagem ad-hoc quando considerado o número de defeitos identificados – não foi observada diferença em relação ao tempo gasto. Um segundo resultado importante foi a identificação, apenas no projeto real, de uma diferença significativa de desempenho entre inspetores experientes e inexperientes. Através da análise das ações de pesquisa em campo, tem-se a indicação de que o fator experiência pode ser preponderante em inspeções baseadas em checklist.*

1. Introdução

No contexto do desenvolvimento de software dirigido por modelos de caso de uso (diagrama e descrição), utilizados como base para planejamento do projeto, construção e comunicação com usuários, a qualidade e inteligibilidade destes modelos são fundamentais. Particularmente no escopo do desenvolvimento do software existem interesses distintos relacionados aos papéis desempenhados durante o desenvolvimento. Por exemplo, o gerente do projeto espera que os casos de uso abranjam todos os

requisitos funcionais, enquanto que um projetista tem a expectativa que a terminologia utilizada seja consistente ao longo de todo o documento [Anda e Sjøberg 2002].

Neste sentido, quando de forma inadequada, casos de uso podem levar a construção de funcionalidades incorretas ou necessitar da interação e comunicação entre a equipe para identificar e corrigir, ao longo do processo, omissões e incorreções no artefato. Uma forma de se conseguir identificar possíveis inadequações na descrição dos casos de uso e, conseqüentemente, prevenir estes problemas é a execução de revisões ou inspeções de casos de uso. Inspeções de artefatos de software têm sido utilizadas com sucesso na indústria como meio para evitar retrabalho e melhorar a qualidade do software [Fagan 2001]. Os principais fatores deste sucesso estão relacionados ao relativo baixo custo da sua utilização e a sua capacidade de encontrar defeitos tão logo sejam inseridos ao longo do processo. Além da detecção de defeitos e conseqüente melhoria da qualidade do software, existem benefícios adicionais associados à utilização de inspeções de software. Deles se destaca a possibilidade da integração entre os processos de detecção e prevenção de defeitos.

Diversos fatores podem influenciar na relação custo-eficiência (número de defeitos por unidade de tempo) de uma inspeção e dos tipos de defeitos encontrados. Características relacionadas à experiência do inspetor ou a sua especialidade técnica no processo de desenvolvimento (analista, programador, testador dentre outros) podem afetar o resultado da inspeção. Como conseqüência, a técnica de inspeção *ad-hoc*, onde não há nenhum tipo de controle sobre a intervenção humana, tem a sua produtividade (custo-eficiência) individualizada e não possui nenhuma garantia da cobertura do documento como um todo e dos tipos de defeitos encontrados [Porter e Votta 1998]. Com isto, pode apresentar um custo-eficiência adequado para inspetores com maior experiência, mas não para os inexperientes [Anda e Sjøberg 2002]. No entanto, mesmo com a utilização de inspetores experientes, a cobertura do documento como um todo pode ainda não ser alcançada já que cada inspetor utiliza a “sua técnica” – e, com isto, determinados tipos de defeitos podem ainda permanecer presentes no documento (Figura 1).

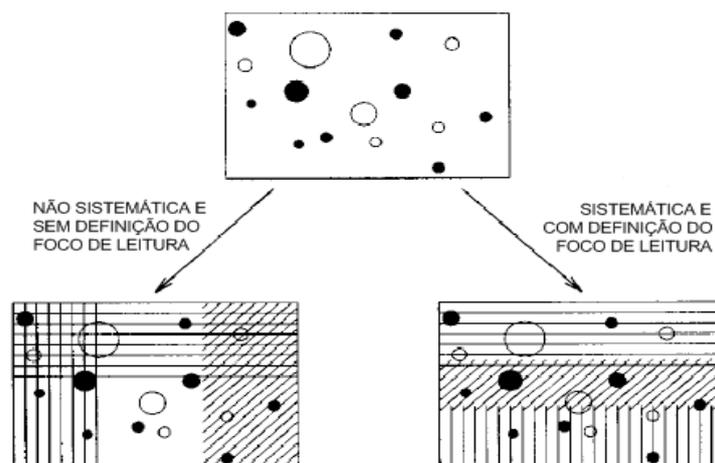


Figura 1. Cobertura dos defeitos em uma inspeção não sistemática e sistemática – círculos e pontos representam defeitos e áreas tracejadas a cobertura de diferentes técnicas/inspetores; adaptado de Porter e Votta (1998)

Existem diferentes técnicas de inspeção que permitem aprimorar as técnicas *ad-hoc* reduzindo a influência do fator humano nos resultados de uma inspeção. Estas téc-

nicas podem ser classificadas em *checklist* e técnicas de leitura. Ainda que não seja um tipo de técnica de inspeção estritamente sistemática, um *checklist*, permite ao inspetor seguir uma lista de itens com características a serem revisadas, sendo a cobertura do documento inspecionado fortemente influenciada pelos itens do *checklist* [Shull *et al.* 2000]. Técnicas de leitura, por sua vez, além de especificarem o que deve ser revisado em um documento, instruem como a leitura deve ser realizada [Shull *et al.* 2000], possuindo assim um caráter mais sistemático. De qualquer maneira, independentemente da técnica de inspeção, um dos principais aspectos na utilização de técnicas de inspeção é observar a sua complementaridade com outras técnicas [Maldonado *et al.* 2006].

O projeto onde este estudo foi conduzido utiliza, desde o seu início, a técnica de inspeção *ad-hoc* para modelos de casos de uso. Além disto, utiliza modelos de caso de uso para representar os requisitos funcionais do software. Ainda que atividades de inspeção de software fizessem parte do processo de desenvolvimento do projeto, a equipe enfrentava dificuldades na compreensão dos requisitos do software. Neste sentido, esta pesquisa investiga a utilização de técnicas de inspeção mais aprimoradas com objetivo de aperfeiçoar o processo de detecção de defeitos e melhorar a compreensibilidade dos modelos de casos de uso do projeto.

2. Metodologia de Pesquisa

O método de pesquisa utilizado neste trabalho é a pesquisa-ação. A pesquisa-ação tem a sua origem associada às primeiras práticas realizadas por Lewin na década de 1940 em psicoterapia. Atualmente, é utilizada em outras áreas como educação e negócios, mas tem sido pouco aproveitada em Engenharia de Software [Santos e Travassos 2009] ainda que possua características adequadas para a área [Santos e Travassos 2008].

A pesquisa-ação é definida como um tipo de pesquisa social com enfoque pragmático a qual é concebida e realizada em estreita associação a uma ação ou com a resolução de um problema coletivo e no qual os pesquisadores e os participantes, representativos da situação ou do problema, estão envolvidos de modo cooperativo [Susman e Evered 1978]. Embora não seja compatível com a vertente de estudos controlados e com os pressupostos do experimentalismo (neutralidade do observador, isolamento de variáveis, etc), a pesquisa-ação não deixa de ser uma forma de experimentação em situação real (*in vivo*) na qual pesquisadores intervêm conscientemente [Checkland e Holwell 1998]. Nela, os envolvidos não são reduzidos a meros participantes e desempenham um papel ativo. E a partir da observação e avaliação das transformações realizadas, e também pela evidenciação dos obstáculos encontrados durante o processo, há um ganho de informação a ser captado e restituído como elemento de conhecimento. Por isto, para o estudo descrito neste trabalho optou-se pela pesquisa-ação, pois se tratava de um projeto real onde os pesquisadores participavam da equipe de desenvolvimento.

Em termos de processo, a pesquisa-ação normalmente contempla as seguintes etapas [Baskerville 1999] (que serão descritas nas próximas seções):

- (1)**Diagnóstico:** consiste em descobrir o campo de pesquisa, os interessados e suas expectativas sob uma perspectiva sistêmica. Nesta etapa, existe ainda a definição do tema de pesquisa o qual é representado através da designação do problema prático e da área do conhecimento a ser abordada (seção 3).
- (2)**Planejamento:** onde são definidas ações para o quadro diagnosticado (seção 6).
- (3)**Tomada da ação:** corresponde a implantação das ações planejadas (seção 7).

- (4)**Avaliação:** etapa onde se realiza a análise dos efeitos das ações frente ao apoio teórico utilizado como ponto de partida para a definição das ações (seções 8).
- (5)**Aprendizagem:** envolve a circulação da avaliação dos resultados entre os participantes e setores da organização. Esta etapa não está contemplada no escopo deste artigo.

3. Contexto da Pesquisa

O projeto no qual esta pesquisa foi conduzida visa o desenvolvimento de um novo sistema de informação Web para gerenciamento das atividades de uma Fundação de Apoio a projetos de Ciência e Tecnologia. Trata-se de um projeto de grande porte que envolve diferentes setores da organização, como: recursos humanos, financeiro, contabilidade, gerenciamento de projetos e protocolo. O projeto pôde ser modularizado e desenvolvido seguindo um ciclo de vida incremental utilizando tecnologia Java e JavaServer Faces (JSF). Além disto, a Fundação possui um modelo de referência de processo baseado nos níveis G a C do MPS.br [SOFTEX 2007].

No momento deste estudo (março, 2009), a equipe do projeto estava organizada de forma distribuída e possuía três projetistas e dois desenvolvedores no local da Fundação (equipe L) e mais seis desenvolvedores remotos em uma cidade distante 200km (equipe R) da sede da Fundação. A experiência média com desenvolvimento da equipe L era de cerca de 7 anos (variando entre 3 e 15 anos), e da equipe R cerca 2,5 anos (variando entre 1,5 e 4 anos).

Para cada módulo do sistema, as atividades de desenvolvimento são desempenhadas em cascata: análise, projeto, codificação e testes. Em todas estas atividades modelos de casos de uso são utilizados como base. Na etapa de projeto, modelos de alto nível são construídos e diagramas de classes e de sequência são elaborados a partir do modelo de caso de uso, todos usando notação UML. Durante a codificação, o modelo de casos de uso é utilizado para implementar os elementos da conversação entre o sistema e o usuário, tais como navegação, validação de formulários e apresentação de dados – a interface gráfica do sistema já é construída durante a etapa de análise por meio da utilização de prototipação. Testes estruturais e funcionais são extraídos do modelo de casos de uso [Dias Neto *et al.* 2007].

No momento desta pesquisa, o projeto contava com três módulos desenvolvidos (módulo de solicitações, gestão de usuários e protocolo) e um parcialmente construído (módulo de acompanhamento de projetos – MAP) com a conclusão de duas iterações, sendo que toda a especificação do módulo já havia sido inspecionada. Um quinto módulo, o módulo financeiro (MFI), já possuía duas iterações especificadas, sendo que a primeira encontrava-se inspecionada.

Em uma análise quantitativa sobre a utilização de inspeção nos três primeiros módulos já desenvolvidos [Kalinowski *et al.* 2007], pôde-se verificar que as inspeções permitiram identificar um grande número de defeitos. Observou-se ainda uma redução contínua da densidade de defeitos (defeitos por página) entre as diferentes inspeções dos respectivos módulos – a densidade passou de 1,4 para 0,84 até chegar a 0,58 no terceiro módulo. Os dados das inspeções do MAP e MFI ainda não haviam sido consolidados. Dois fatores importantes podem ser atribuídos a este decréscimo; a mudança do processo de elicitação de requisitos que passou a considerar também nas entrevistas

pessoas com perfil gerencial e não só operacional, e o aprendizado adquirido pela equipe com relação aos defeitos presentes nos modelos de casos de uso [Kalinowski *et al.* 2007].

No entanto, mesmo com o bom histórico da utilização de inspeções nestes módulos do projeto, durante a construção do MAP as equipes de projetistas e desenvolvedores enfrentaram dificuldades no entendimento dos modelos de casos de uso, exigindo uma constante interação com os analistas para prosseguir com o desenvolvimento. A percepção da equipe em relação a esta situação foi a de que o domínio havia tornado-se maior e mais complexo e, associado a isto, a forma como as descrições dos casos de uso vinham sendo elaboradas estavam inadequadas para representar esta crescente complexidade. Por exemplo, cálculos financeiros eram apresentados de forma escrita – “para calcular o percentual da taxa X multiplique o valor orçado para o projeto pelo percentual de administração de projetos” – em vez do uso de uma fórmula algébrica.

Diante das dificuldades no entendimento dos casos de uso do MAP e do histórico de bons resultados da utilização de inspeções no projeto surge, então, a questão do porquê os problemas relacionados à compreensão dos modelos não foram capturados anteriormente. Devido às limitações das inspeções ad-hoc, esta pesquisa busca viabilizar e avaliar o uso de técnicas mais controladas com o objetivo de melhorar a qualidade dos modelos dos casos de uso visando, principalmente, a compreensibilidade e inteligibilidade dos artefatos.

4. Trabalhos Relacionados

Em um estudo de pesquisa-ação, os trabalhos relacionados devem ser úteis à pesquisa no sentido de indicarem como o estudo deve ser conduzido. Para o contexto desta pesquisa, o trabalho de Deboni e Gregolin (2008), que corresponde à pesquisa realizada por Gregolin (2007), mostrou-se adequado.

Gregolin (2007) define um modelo de qualidade para diagrama e descrição de casos de uso, onde são definidos atributos de qualidade e regras de elaboração. A partir deste modelo, *checklists* foram elaborados para verificar se os diagramas e descrições do caso de uso inspecionado atendem aos modelos de qualidade propostos.

O modelo de qualidade definido por Gregolin (2007) baseia-se em uma ampla pesquisa bibliográfica sobre boas práticas e diretrizes de elaboração de modelo de caso de uso, além de três modelos de qualidade de especificação de software [IEEE Std 830 1998, Davis *et al.* 1993, Fabbrini *et al.* 2001]. O modelo do Institute of Electrical and Electronics Engineers [IEEE Std 830 1998], em suas práticas recomendadas para uma boa especificação de requisitos de software, define como características a correção, não ambigüidade, completeza, consistência, classificação por importância ou estabilidade, verificabilidade, modificabilidade e rastreabilidade. Em Davis *et al.* (1993) são definidos 18 atributos de qualidade muitos dos quais são análogos aos do modelo do IEEE Std 830 (1998), mas também apresentam preocupações diferentes incluindo, por exemplo, compreensibilidade e independência de ferramenta. O terceiro modelo de qualidade, de Fabbrini *et al.* (2001), é fundamentado em técnicas lingüísticas e visa, essencialmente, a testabilidade, consistência e compreensibilidade. Entre os indicadores de defeitos rela-

cionados a estes atributos de qualidade de Fabrini *et al.* (2001) estão a imprecisão, subjetividade, presença de termos implícitos e média de palavras por sentença.

Por meio de um comparativo entre estes modelos, Gregolin (2007) extrai atributos de qualidade e características que poderiam servir de regra de elaboração de um modelo de caso de uso e que poderiam ser verificáveis através de um processo de inspeção utilizando *checklists*. Voltando ao contexto do projeto apresentado anteriormente, as características da proposta de inspeção de Gregolin (2007) apresentaram-se como uma alternativa interessante para os problemas diagnosticados – principalmente em relação ao entendimento dos casos de uso. De fato, Anda e Sjöberg (2002) já haviam mostrado que diretrizes para elaboração de casos de uso são importantes para as descrições em termos do nível de detalhe utilizado, realismo (sequência lógica e completa) e consistência (uso correto de vocabulário). Apresentaram como resultado que o uso de diretrizes é significativamente melhor para construir, e posteriormente compreender, casos de uso daqueles que não são elaborados com a utilização de diretrizes. Seguindo este raciocínio, esperávamos que a utilização da proposta de Gregolin (2007) no contexto do projeto pudesse apresentar ganhos frente à inspeção *ad-hoc*, já que o seu foco está nos atributos de qualidade fundamentados nas diretrizes de elaboração de casos de uso dos três modelos de qualidade apresentados. Os atributos de qualidade são: compreensibilidade, completeza, precisão, não ambigüidade, consistência, independência de ferramenta e de interface, rastreabilidade, objetividade e representação de valor. Cada atributo de qualidade possui associado uma ou mais questões do *checklist*. O detalhamento do *checklist* pode ser obtido em Gregolin (2007).

Para melhor compreender e avaliar a decisão da escolha da tecnologia *checklist* de Gregolin (2007) antes de aplicá-la em um projeto real (Seção 6), entendeu-se necessária a condução de um estudo controlado o qual é descrito na seção a seguir.

5. Estudo Controlado Inicial

Três objetivos principais foram traçados à execução do estudo controlado: (1) obter conhecimento e proficiência sobre a tecnologia que seria utilizada no projeto, no caso o *checklist*, (2) observar como a tecnologia se comportaria frente à técnica *ad-hoc* em um ambiente controlado, capturando indícios de seus benefícios e limitações, e (3) verificar a necessidade de alguma melhoria ou ajuste à tecnologia antes de aplicá-la no projeto.

Os participantes foram divididos aleatoriamente em três grupos. O documento inspecionado era composto de dois casos de uso, requisitos funcionais e não-funcionais referentes aos casos de uso, além do glossário e descrição geral do sistema. Foram utilizados casos de uso reais do projeto referentes ao módulo MAP (módulo anterior ao MFI). No total, seis casos de uso foram selecionados e separados por complexidade (B – baixa, M – média e A – alta). Cada grupo utilizou inicialmente a técnica *ad-hoc* e, em seguida, a técnica *checklist*. A Tabela 1 sumariza o arranjo do estudo.

Tabela 1. Arranjo do estudo controlado

	Grupo 1	Grupo 2	Grupo 3
Treinamento	Sem treinamento – foi apresentada a descrição do domínio e definido o foco da inspeção nos atributos de qualidade		
Inspeção Ad-hoc	Casos de Uso • B1 e A1	Casos de Uso • B2 e M1	Casos de Uso • M2 e A2
Treinamento	Com treinamento – foram apresentadas as regras do <i>checklist</i>		

Inspeção checklist	Casos de Uso • B2 e M1	Casos de Uso • M2 e A2	Casos de Uso • B1 e A1
--------------------	---------------------------	---------------------------	---------------------------

O estudo controlado foi executado entre os dias 14/08/2008 e 04/09/2008 com doze alunos de mestrado e doutorado – cada grupo ficou com quatro alunos. Durante a disciplina, na qual o estudo foi executado, já havia sido ministrado aos alunos conceitos sobre inspeção e técnicas de inspeção (*checklist* e técnica de leitura). Os alunos também já haviam realizado uma inspeção *ad-hoc* no contexto da disciplina. A inspeção *ad-hoc* foi iniciada no dia 14/08 e teve prazo de uma semana para execução, enquanto que a *checklist* foi iniciada no dia 28/08 e também contou com uma semana para término. A atividade de discriminação de defeitos, onde as discrepâncias são classificadas em falso-positivo ou defeito, foi realizada pelo próprio pesquisador.

O método de análise de variância foi escolhido para a identificação de alguma diferença estatística significativa entre médias das medidas da inspeção *ad-hoc* e *checklist*. Assim, tínhamos como hipótese para o teste estatístico a não diferença entre as médias das medidas, e consideramos um *alpha* (nível de significância) menor que 0,05 para refutar esta hipótese. A Tabela 2 consolida os resultados do estudo conduzido. Conforme o arranjo descrito anteriormente, são apresentadas comparações entre as inspeções *ad-hoc* e *checklist*, agrupado pela complexidade dos casos de uso. Com este agrupamento, têm-se oito observações (medidas) em cada inspeção. Por exemplo, para as observações da inspeção *checklist* dos casos de uso de baixa complexidade foram agrupados os grupos 1 (caso de uso B2) e 3 (caso de uso B1) – conforme Tabela 1.

Tabela 2. Resultados da inspeção

Técnica	# Discrepâncias			# Defeitos		
	Média	Des. Padrão	Alpha	Média	Des. Padrão	Alpha
Complexidade Baixa						
<i>Ad-hoc</i>	2,875	1,727	0,0584	1,125	0,641	0,0001
<i>Checklist</i>	5,250	2,764		3,500	0,926	
Complexidade Média						
<i>Ad-hoc</i>	6,000	2,725	0,2294	3,250	1,752	0,0248
<i>Checklist</i>	9,375	7,090		6,875	3,681	
Complexidade Alta						
<i>Ad-hoc</i>	7,875	7,279	0,0583	3,625	2,264	0,0001
<i>Checklist</i>	14,750	5,994		9,000	1,852	

Analisando o resultado das inspeções nota-se uma diferença entre as médias, com um resultado superior à inspeção *checklist* em todos os casos. No entanto, apenas as diferenças relacionadas à defeitos mostraram-se estatisticamente significativas ($\alpha < 0,05$). Isto representa um resultado positivo na medida em que mostra que a técnica *checklist* detectou um menor número de falso-positivos no contexto desta inspeção. Outro dado interessante é que a inspeção *ad-hoc* surpreendentemente apresentou, em grande parte dos casos, um desvio padrão menor que *checklist* podendo ser um indício da influência de algum fator de confusão, como a experiência do inspetor. Por esta razão, uma análise foi feita buscando detectar alguma influência da experiência sobre o resultado. Para isto, os inspetores foram subdivididos em dois grupos, um com os mais experientes e outros com os menos. Não foi observada nenhuma mudança de comportamento nesta análise, ou seja, tanto para o grupo mais experiente quanto para o com menor experiência a técnica *checklist* foi superior no número de defeitos identificados.

Além dos dados quantitativos apresentados, dados qualitativos foram coletados por meio de questionários pós-estudo. Dois tipos de questões foram feitas aos inspetores: questões relacionadas à condução e aplicação do estudo controlado e questões relacionadas à utilidade e qualidade da técnica *checklist*. A avaliação qualitativa deste questionário mostrou um retorno positivo por parte dos participantes tanto da utilidade da técnica *checklist* quanto sobre a condução do estudo controlado. Um dado que chamou atenção é que um número não desprezível de participantes (quatro) enfrentou algum tipo de dificuldade na aplicação do *checklist*. Em discussões informais com alguns dos participantes a percepção foi de que um número elevado, mas necessário, de questões torna o *checklist* um pouco cansativo e até mesmo repetitivo já que muitas exigem a verificação de todo o documento. Ainda assim, quando perguntados se o número de questões era suficiente ou insuficiente, muitos participantes responderam com sugestões de novos tópicos ou questões que o *checklist* deveria cobrir.

Dado os objetivos definidos no início desta seção, os resultados (quantitativos e qualitativos) obtidos neste estudo controlado mostraram um comportamento benéfico no que diz respeito à detecção de defeitos segundo o foco da compreensibilidade dos modelos de casos de uso. Foi importante observar a utilidade do *checklist* tanto para inspetores experientes quanto para os inexperientes. Além disto, ainda segundo o escopo dos objetivos traçados, a condução do estudo permitiu ao pesquisador, durante atividades como preparação do material de treinamento e discriminação de defeitos da inspeção, um maior entendimento da técnica. Não foi observada, no entanto, nenhuma necessidade de adaptação do *checklist* para a aplicação no projeto. Neste sentido, diante destes resultados, a aplicação da tecnologia no projeto real foi iniciada e avaliada por meio da metodologia da pesquisa-ação.

6. Planejamento do Estudo de Pesquisa-Ação no Projeto Real

O planejamento da pesquisa-ação passa por três etapas. A primeira consiste em uma revisão da literatura onde trabalhos relacionados ao tema de pesquisa são estudados – etapa descrita na seção 4. A partir disto, o enfoque da ação da pesquisa é fixado por meio da abordagem GQM [Basili *et al.* 1994] e os instrumentos utilizados definidos.

6.1. Enfoque da Ação

Objetivo

O *objetivo* é analisar a inspeção de modelos de caso de uso *com o propósito* de caracterizar *com respeito à* eficiência da identificação de defeitos que dificultam a compreensibilidade de modelos de casos de uso *do ponto de vista* do engenheiro de software *no contexto* de desenvolvimento de sistemas de informação Web.

Questões de Pesquisa

O objetivo será alcançado quando uma resposta for apresentada às seguintes questões:

- Q.1. O uso da técnica de inspeção *checklist* encontra mais defeitos que a técnica *ad-hoc*? Esta questão tem como objetivo verificar, no contexto do projeto, se o custo-eficiência utilizando a técnica *checklist* é melhor do que a técnica *ad-hoc*.
- Q.2 A técnica de inspeção *checklist* encontra defeitos diferentes do que a técnica *ad-hoc*? A idéia aqui é comparar o resultado das duas técnicas, o que servirá de base

para uma indicação sobre alguma diferença na compreensibilidade dos modelos de caso de uso.

Resultados Esperados (coleção de dados)

A questão Q.1 possui as seguintes questões práticas associadas:

- Q.1.1 Qual é o número de defeitos encontrados utilizando cada uma das técnicas?
- Q.1.2 Qual é o tempo médio aplicado à inspeção pelos inspetores nas duas técnicas?

A questão Q.2 possui a seguinte questão prática associada:

- Q.2.1 Qual a distribuição do tipo de defeitos identificados?

Para as questões práticas acima, serão analisados os dados informados pelos inspetores após a etapa de discriminação de defeitos.

6.2. Instrumentos

Como instrumento, utilizou-se a ferramenta ISPIS. A Figura 2 apresenta o formulário de inclusão de discrepâncias utilizado pelos inspetores. Cada discrepância pode ser classificada em cinco tipos [Kalinowski e Travassos 2004]:

- **Omissão:** (1) algum requisito importante relacionado à funcionalidade, ao desempenho, às restrições do projeto, ao atributo, ou à interface externa não foi incluído; (2) não está definida a resposta do software para todas as possíveis situações de entrada de dados; (3) faltam seções na especificação de requisitos.
- **Ambigüidade:** um requisito tem várias interpretações devido a diferentes termos utilizados para uma mesma característica.
- **Inconsistência:** dois ou mais requisitos são conflitantes.
- **Fato incorreto:** um requisito descreve um fato que não é verdadeiro, considerando as condições solicitadas para o sistema.
- **Informação estranha:** as informações fornecidas no requisito não são necessárias.
- **Outros:** outros defeitos como a inclusão de requisito em uma seção errada.

Nova Discrepância	
Localização	Pag 71 UC 20
Classificação	Informação Inconsistente
Severidade	Alta
Descrição	O passo 11 do fluxo principal diz: "11. O sistema verifica se o centro de custo de destino possui saldo no plano de contas associado ao protocolo"
Adicionar	
Concluído	

Figura 2. Formulário de discrepâncias de ISPIS

Alguns outros instrumentos de pesquisa foram utilizados para a caracterização dos inspetores e questionário pós-estudo.

7. Ações

Dez casos de uso referentes à segunda iteração do Módulo Financeiro estavam especificados e deveriam ser inspecionados, conforme o processo definido no projeto. Na medida em que atividades de inspeção não representam custos elevados ao projeto e para viabilizar resultados estatísticos significativos, definiu-se que quatro membros da equipe L (um projetista e três desenvolvedores) participariam da inspeção, e os seis membros da equipe R (todos desenvolvedores) também deveriam participar. Somando um total de dez inspetores para inspeção. Em inspeções anteriores, apenas os desenvolvedores mais experientes eram selecionados e, por este motivo, dos dez inspetores apenas três tinham experiência prévia com inspeção no contexto do projeto – um na equipe L e dois na equipe R. O arranjo deste estudo foi fortemente baseado no estudo controlado inicial descrito anteriormente e é apresentado na Tabela 3.

Tabela 3. Arranjo do estudo

	Equipe L	Equipe R
Treinamento	Sem treinamento – foi apresentada a descrição do domínio e definido o foco da inspeção nos atributos de qualidade	
Inspeção Ad-hoc	Casos de Uso <ul style="list-style-type: none"> • B1, B2, M1, M2 e A1 	Casos de Uso <ul style="list-style-type: none"> • B3, B4, M3, M4 e A2
Treinamento	Com treinamento – foram apresentadas as regras do <i>checklist</i>	
Inspeção checklist	Casos de Uso <ul style="list-style-type: none"> • B3, B4, M3, M4 e A2 	Casos de Uso <ul style="list-style-type: none"> • B1, B2, M1, M2 e A1

A equipe R foi a primeira a inspecionar os documentos entre os dias 17/02/2009 e 19/02/2009. O pesquisador deslocou-se até a cidade da equipe R para acompanhar as atividades. Os desenvolvedores foram instruídos sobre o foco da inspeção (mesmos atributos de qualidade do *checklist*) e sobre o uso da ferramenta ISPIS. Feito isto, a inspeção *ad-hoc*, com cinco casos de uso, foi iniciada em ISPIS. Como alguns inspetores ainda não haviam trabalhado com a ferramenta, eles foram auxiliados na sua utilização. A ferramenta é de uso bastante simples, especialmente na visão do inspetor, onde praticamente só tem acesso à facilidade de cadastro de discrepâncias. A inspeção *ad-hoc* pôde ser concluída em um dia. No dia seguinte, o treinamento previsto para a inspeção *checklist* foi aplicado e, logo em seguida, a inspeção *checklist*, com os outros cinco casos de uso, foi iniciada em ISPIS. A inspeção *checklist* foi concluída em dois dias, devido à necessidade do treinamento um pouco mais elaborado.

Para a equipe L a inspeção foi executada durante os dias 20/02/2009 e 09/03/2009. O treinamento conduzido foi o mesmo, mas, diferentemente da inspeção da equipe R, cada inspeção teve prazo de cerca de uma semana já que os desenvolvedores desta equipe trabalham em regime de tempo parcial e estavam em meio a outras atividades do projeto. Em nenhuma inspeção houve qualquer tipo de interrupção das atividades por conta de falha técnica de ISPIS ou não entendimento dos procedimentos apresentados aos inspetores.

8. Análise dos Resultados

Conforme definido anteriormente, a análise dos resultados segue o mesmo formato da apresentada no estudo inicial. Como uma lição aprendida na condução do estudo inicial, atentou-se para a necessidade do registro do tempo gasto por caso de uso já que foram divididos por complexidade. Esta seção está dividida em duas partes, inicialmente

concentra-se na análise quantitativa relacionada às questões de pesquisa Q.1 e Q.2 cujo principal objetivo é avaliar o desempenho da técnica *checklist* com relação à quantidade e qualidade (distribuição) dos defeitos encontrados. A segunda parte descreve sucintamente a avaliação qualitativa do estudo com base, principalmente, no questionário pós-estudo, mas trazendo também alguns comentários sobre a percepção da utilização da técnica no contexto do projeto.

Os resultados da inspeção são apresentados na Tabela 4. Considerando a classificação por complexidade dos casos de uso não houve, assim como no estudo inicial, nenhuma diferença expressiva entre as três classificações. Desta forma, apenas os resultados considerando todos os casos de uso (agrupando todas as complexidades) são apresentados. Além do registro do tempo mencionado anteriormente, o custo-eficiência da inspeção pôde ser computado. Também são apresentados resultados considerando a experiência do inspetor. A experiência foi avaliada com base no formulário de caracterização considerando o tempo como desenvolvedor, participações em outras inspeções, tempo de atuação no contexto do projeto, dentre outros. Cada subgrupo (experiente e inexperiente) ficou com quatro indivíduos já desconsiderando os casos atípicos (*outliers*) – somando um total de oito indivíduos das equipes L e R.

Tabela 4 - Resultados da inspeção

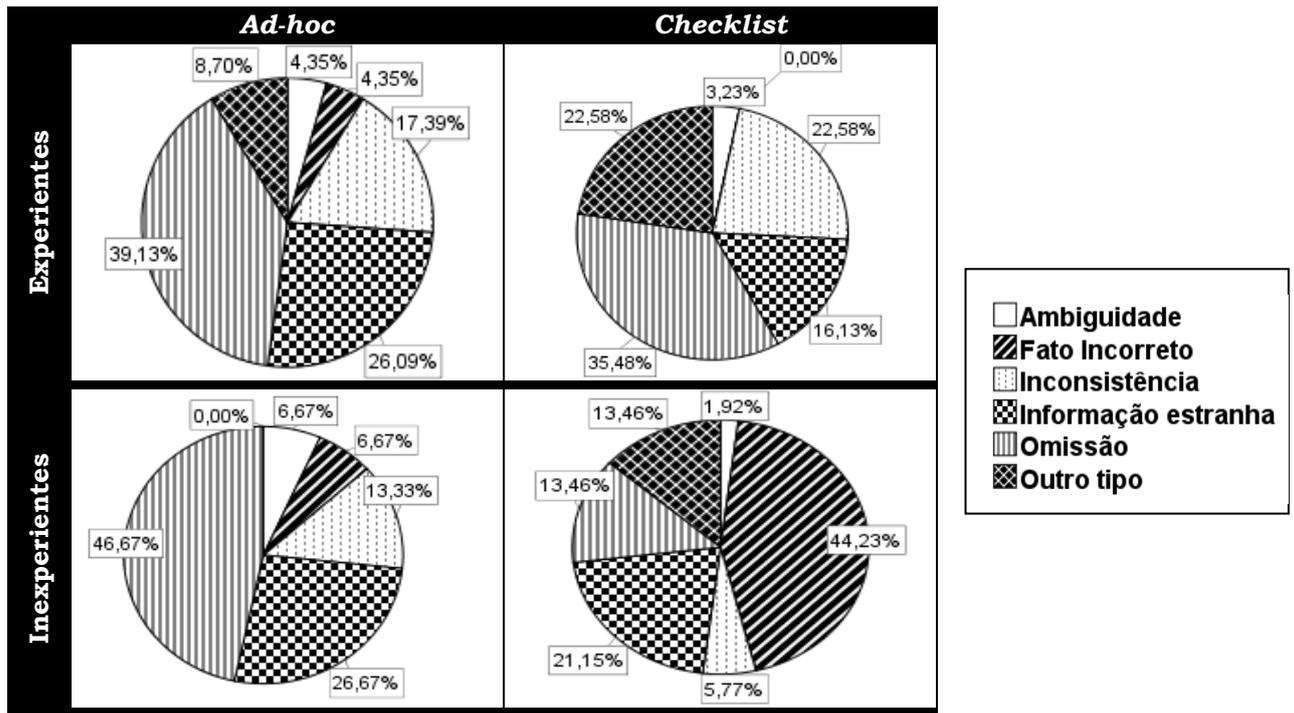
Técnica	Tempo (min.)			Discrepância			Defeito			Custo-Eficiência		
	Média	Des. Padrão	Alpha	Média	Des. Padrão	Alpha	Média	Des. Padrão	Alpha	Média	Des. Padrão	Alpha
Inexperientes												
<i>Ad-hoc</i>	221,00	122,70	0,517	17,75	7,13	0,479	4,00	3,16	0,011*	0,97	0,29	0,038*
<i>Checklist</i>	166,75	99,54		21,75	7,84		13,00	3,83		5,73	3,59	
Experientes												
<i>Ad-hoc</i>	283,50	167,91	0,606	14,75	5,56	0,470	5,50	3,00	0,311	1,90	1,67	0,557
<i>Checklist</i>	227,50	119,05		11,50	6,35		7,75	2,75		2,58	1,41	
Média (agrupando experientes e inexperientes)												
<i>Ad-hoc</i>	252,25	140,18	0,391	16,25	6,13	0,9214	4,75	2,96	0,007*	1,43	1,21	0,033*
<i>Checklist</i>	197,12	106,65		16,65	8,58		10,37	4,17		4,16	3,04	

Os resultados apresentados acima mostram, assim como no estudo inicial conduzido, uma superioridade da técnica *checklist* na identificação de defeitos em relação à *ad-hoc* ainda que aponte, em média, o mesmo número de discrepâncias. Não foi possível perceber nenhuma diferença significativa para o tempo gasto entre as técnicas, o que é um indício de que a técnica *checklist* não afeta intensamente a forma de leitura dos modelos de casos de uso ainda que exista um foco maior nas questões tratadas no *checklist*. Em função do registro do tempo pôde ser calculado o custo-eficiência, onde também se observou um bom desempenho da técnica *checklist* influenciado em grande parte pelo um maior número de defeitos identificados. *Em resumo, estes resultados mostram que checklist identifica um maior número de defeitos, resultante de um menor número de falso-positivos apontados, em um mesmo intervalo de tempo em comparação com a técnica ad-hoc.*

É necessário analisar também, o resultado mais característico em relação ao estudo inicial que é a existência de uma diferença significativa de comportamento entre as técnicas quando a análise é feita considerando a experiência dos inspetores. Nesta perspectiva é possível perceber que a técnica *checklist* é mais útil aos inspetores menos experientes do que aqueles com mais experiência. Isto parece ser um indício de que os

inspetores mais experientes consideraram todas ou grande parte das preocupações presentes nas questões do *checklist* durante a realização da inspeção *ad-hoc*. No sentido contrário, outro fator que pode explicar este resultado é que a experiência do inspetor pode tê-lo feito apontar um número idêntico de defeitos em ambas as técnicas, mas defeitos diferentes. Tenta-se avaliar esta questão, adiante, por meio da análise das distribuições dos defeitos.

Tabela 5 - Distribuição de defeitos das inspeções *ad-hoc* e *checklist* classificado pela experiência dos inspetores



Dois fatos interessantes podem ser observados através da distribuição de defeitos apresentada na Tabela 5. O primeiro é que, olhando apenas para a coluna *ad-hoc*, percebe-se que tanto os mais experientes quanto os menos têm distribuições semelhantes, sugerindo que apesar da inspeção ter sido *ad-hoc* o foco foi bastante homogêneo entre todos os inspetores. O segundo, agora comparando *ad-hoc* e *checklist* (as duas colunas da tabela), é que para os mais experientes a distribuição manteve-se razoavelmente similar enquanto que para os menos experientes muda sensivelmente. No caso dos menos experientes, os tipos “fato incorreto” e “outro tipo” ampliaram expressivamente em detrimento, principalmente, dos defeitos do tipo “omissão”. Em uma análise qualitativa superficial dos defeitos percebeu-se que os defeitos do tipo “omissão” estavam mais associados ao domínio do negócio e do sistema, enquanto que “fato incorreto” e “outro tipo” foram os tipos mais usados para classificar questões apontadas diretamente pelo *checklist*. Este mesmo padrão de diferença entre as distribuições *ad-hoc* e *checklist* pode ser percebido para os mais experientes, porém de forma bastante tênue, mas neste caso os mais experientes usaram o tipo “inconsistência” para classificar questões apontadas pelo *checklist* em grande parte dos casos em vez de “fato incorreto”. Parece que os menos experientes entenderam que um não atendimento ao *checklist* representava um fato incorreto enquanto que os mais experientes classificaram corretamente.

Em suma, diante da análise dos resultados, é possível argumentar que a técnica *checklist* auxilia na identificação de defeitos singulares (diferentes da abordagem *ad-hoc*) em modelos de casos de uso, tornando-os em princípio mais compreensíveis.

Cabe agora, por fim, analisar a avaliação feita pelos participantes em relação à técnica empregada. O mesmo questionário do estudo controlado foi respondido pelos membros do projeto e os resultados foram idênticos aos identificados no estudo inicial. Todos informaram ter seguido completamente as instruções do *checklist* o que torna mais sólido e consistente os resultados quantitativos analisados. Um fato, entretanto, que tem um significado importante, foi a visão, por um dos inspetores, de que o *checklist* atuou como obstáculo à identificação de defeitos. Este inspetor está entre os mais experientes. Com relação a isto, o pesquisador pôde presenciar durante reuniões de projeto comentários entre os inspetores demonstrando alguma insatisfação no uso da técnica *checklist*, quando ao segui-la completamente abdicaram da possibilidade de apontar alguns possíveis defeitos não capturados no *checklist*. Talvez isto denote uma incompreensão destes inspetores de que de fato a idéia do *checklist* é justamente focar a inspeção em alguns aspectos tratados pelo próprio *checklist*. Mas, por outro lado, reafirma que a experiência é um fator preponderante ao desempenho do inspetor mesmo com uso de técnicas mais sistemáticas. De fato, este é um comportamento de certa forma já observado em estudos anteriores [Shull *et al.* 2000] onde inspetores mais experientes tendem a aplicar heurísticas adquiridas anteriormente combinando-as com possíveis novos aspectos da técnica utilizada. Para equipes experientes, o mais apropriado parece ser utilizar técnicas de inspeção como mecanismo de aprendizado para aspectos específicos de interesse do projeto e, a partir disto, retornar a utilização da técnica *ad-hoc*.

9. Conclusões

As questões de pesquisa deste estudo estiveram associadas à investigação do desempenho da técnica de inspeção *checklist* proposta por Gregolin (2007) frente à inspeção *ad-hoc*. As questões Q.1 e Q.2 puderam ser avaliadas por meio da análise quantitativa, onde foi possível perceber que a técnica *checklist* foi capaz não apenas de identificar um maior número de defeitos, mas, sobretudo, diferentes daqueles identificados em inspeções *ad-hoc* realizadas até então no contexto do projeto. Esta era a principal expectativa em relação à utilização da técnica no projeto. Ainda assim, apesar de em princípio este objetivo ter sido alcançado, a avaliação qualitativa do questionário pós-estudo mostrou que uma parcela dos inspetores, principalmente os experientes, sentiu-se limitada pela aplicação do *checklist*. Na verdade, a percepção do pesquisador em relação a esta idéia de limitação por parte dos inspetores, é que ela esteve associada ao fato de que o *checklist* foca mais questões sintáticas do que semânticas. No entanto, um modelo de casos de uso semanticamente aderente, mas com deficiências sintáticas como, por exemplo, imprecisões e inconsistências, pode provocar tantos defeitos em etapas seguintes do processo de desenvolvimento quanto um modelo de casos de uso com características opostas.

O fator experiência revelou-se neste estudo como elemento principal da análise dos resultados. Os menos experientes puderam ser mais eficientes com o uso da técnica *checklist* do que os mais experientes. De fato, este resultado é coerente com hipóteses sugeridas por Shull *et al.* (2000) e Anda e Sjöberg (2002), que apontam o fator experiência para explicar a não diferença entre técnicas *ad-hoc* e *checklist* em estudos

anteriores [Cheng e Jeffery 1996, Lanubile e Visaggio 1996, Miller *et al.* 1998]. Da mesma forma, esta hipótese também deveria ser avaliada quando há diferença a favor de *checklist* como outros estudos mostram [Wohlin *et al.* 2002, Cox *et al.* 2004]. No entanto, em nenhum destes dois trabalhos [Shull *et al.* 2000, Anda e Sjøberg 2002] foi feita uma avaliação experimental desta questão. No trabalho de Anda e Sjøberg (2002) esta hipótese surgiu devido à observação de que no estudo executado por eles, onde nenhuma diferença entre *ad-hoc* e *checklist* foi identificada, todos os participantes eram experientes.

Todavia, ao contrário de Anda e Sjøberg (2002), o grupo de inspetores deste estudo tinha experiência bastante heterogênea, o que permitiu a avaliação experimental que culminou na confirmação da hipótese. Não foi possível identificar na literatura técnica nenhum trabalho que tenha investigado esta questão como foi neste estudo. Em Maldonado *et al.* (2006) o fator experiência foi avaliado experimentalmente, mas considerando técnicas de leitura, mas não *checklist*. Nele, não é identificada diferença de desempenho para a técnica de leitura PBR (*Perspective-based reading*) (Basili *et al.*, 1996) entre inspetores de diferentes níveis de experiência.

Ainda assim, uma avaliação mais detalhada melhoria da compreensibilidade dos modelos de casos precisa ser conduzida em trabalhos futuros e representa uma limitação deste estudo. Isto porque, apesar de identificarmos que a técnica *checklist* utilizada identifica mais defeitos (além de diferentes) que a técnica *ad-hoc*, não podemos afirmar ainda que de fato houve uma melhora na compreensibilidade dos casos de uso. Para esta avaliação, um estudo com ênfase em dados qualitativos, buscando capturar se a leitura dos modelos de casos de uso é facilitada em função da utilização do *checklist*, possivelmente trará maior clareza se realmente os atributos de qualidade tratados pelo *checklist* afetam a compreensibilidade dos modelos.

De qualquer maneira, baseado na nossa experiência no projeto e nos resultados apresentados, indicamos a aplicação do *checklist* como ferramenta para identificação de um maior número de defeitos relacionados à compreensibilidade de modelos de casos de uso. A técnica foi incorporada ao processo do projeto e internalizada pelos desenvolvedores.

Agradecimentos

Os autores gostariam de agradecer à Rosângela Gregolin e ao Prof. José Eduardo Deboni pelos materiais fornecidos e informações adicionais sobre a técnica *checklist*.

Referências

- Anda, B. and Sjøberg, D.I.K. (2002). "Towards an Inspection Technique for Use Case Models". In: Fourteenth International Conference on Software Engineering and Knowledge Engineering (SEKE), Ischia, Italy, July 15-19.
- Basili, V.R., Caldiera, G., Rombach, H.D. (1994) "The Goal Question Metric approach. In the Encyclopedia of Software Engineering", vol. 2, pp. 528-532, John Wiley & Sons, Inc.
- Basili, V., Green, S., Laitenberger, O., Shull, F., Sorumgaard, S., Zelkowitz, M. (1996) "The empirical investigation of perspective based reading". In: *Empir Softw Eng-Int J.* 1:133-164.
- Baskerville, R. L. (1999) "Investigating information systems with action research". In: *Communications of the Association for Information Systems*, volume 2.
- Checkland, P., Holwell, S. (1998) "Action Research: Its Nature and Validity". In: *Systemic Practice and Action Research* 11(1): 9-21.

- Cheng, B., Jeffery, R. (1996) "Comparing Inspection Strategies for Software Requirement Specifications". In: Proc. Australian Software Engineering Conference, Los Alamitos, USA.
- Cox, K., Aurum, A., Jeffery, R. (2004) "An experiment in inspecting the quality of use case descriptions". In: Journal of Research and Practice in Information Technol. 36(4):211-229.
- Davis, A, Overmyer, S, Jordan, K. (1993). "Identifying and Measuring Quality in a Software Requirements Specification". In: Proc. of the 1st Int. Soft. Metrics Symposium, Baltimore.
- Deboni, J. E. Z., Gregolin, R. (2008). "Inspeção de Qualidade em Descrições de Casos de Uso: Uma Proposta de Modelo e Artefatos". In: VII SBQS, Florianópolis – SC, Brasil.
- Dias Neto, A. C., Sinola, R. O., Bott, A., Travassos, G. H. (2007). "Estratégia de Teste de Software no Desenvolvimento Incremental de um Sistema de Informação". In: Workshop on Systematic and Automated Software Testing, 2007, João Pessoa – PB, Brasil.
- Fabbrini, F., Fusani, M., Gnesi, S., Lami, G. (2001). "An Automatic Quality Evaluation for Natural Language Requirements". In: 7th Int. Workshop on Requirements Engineering.
- Fagan, M. (2001) "A History of Software Inspections", Sd&m Conference, Disponível em: <http://www.mfagan.com/resources.html>. Acessado em 23 de janeiro de 2010.
- Gregolin, R. (2007). "Uma proposta de inspeção em modelos de caso de uso". Dissertação de Mestrado em Engenharia da Computação, Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, 108 f.
- IEEE Std 830-1998. (1998). "IEEE Recommended Practice for Software Requirements Specifications". Software Engineering Standards Committee of the IEEE Computer Society.
- Kalinowski, M., Spínola, R.O., Dias Neto, A.C., Bott, A., Travassos, G. H. (2007) "Inspeções de Requisitos de Software em Desenvolvimento Incremental: Uma Experiência Prática", In: VI SBQS, Porto de Galinhas – PE, Brasil.
- Kalinowski, M., Travassos, G.H. (2004) "A Computational Framework for Supporting Software Inspections", 19th IEEE Int. Conf. on Automated Soft. Engineering, Linz, Austria.
- Lanubile, F., Visaggio, G. (1996) "Assessing defect detection methods for software requirements inspections through external replication". In: ISERN-96-01, January.
- Maldonado, J., Carver, J., Shull, F., Fabbri, S., Doria, E., Martimiano, L., Mendonça, M., and Basili, V. (2006) "Perspective-Based Reading: A Replicated Experiment Focused on Individual Reviewer Effectiveness". In: Emp. Sof. Eng.: An Int. Journal. v. 11, n. 1.
- Miller, J., Wood, M., Roper, M. and Brooks, A. (1998) "Further Experiences with Scenarios and Checklists". Empirical Software Engineering, Vol. 3(1), pp. 37-64, January.
- Porter A. and Votta L. (1998) "Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment Using Professional Subjects". In: Empirical Software Eng. J., vol. 3, no. 4, pp. 355-379.
- Santos, P. S. M. e Travassos, G. H. (2008) "Colaboração entre Academia e Indústria: Oportunidades para Utilização da Pesquisa-Ação em Engenharia de Software". In: 5th Experimental Software Engineering Latin American Workshop, v. 1, p. 1-10.
- Santos, P.S.M. e Travassos, G. H. (2009) "Action Research Use in Software Engineering: an Initial Survey". In: 3rd Int. Symp. on Empirical SE and Measurement, Orlando, USA.
- Shull, F., Rus, I., and Basili, V. (2000). "How Perspective-Based Reading Can Improve Requirements Inspections". IEEE Software, July 73-79.
- SOFTEX: MPS.BR (2007) "Melhoria de Processo do Software Brasileiro". Guia Geral Versão 1.2, Campinas, SP.
- Susman, G.L., Evered, R.D. (1978) "An assessment of the scientific merits of action research". In: Administrative Sciences Quarterly, 23, pp. 582-603.
- Wohlin, C., Aurum, A., Petersson, H., Shull, F., Ciolkowski, M. (2002) "Software inspection benchmarking - a qualitative and quantitative comparative opportunity". In: Proceedings of 8th International Software Metrics Symposium, 118-130.