



Caracterização do Estado da Prática das Atividades de Teste em um Cenário de Desenvolvimento de Software Brasileiro

Arilo Claudio Dias Neto¹, Ana Cândida Cruz Natali², Ana Regina Rocha,
Guilherme Horta Travassos

Universidade Federal do Rio de Janeiro – COPPE/Sistemas
Caixa Postal 68.511 – CEP 21.941-972 – Rio de Janeiro – RJ – Brasil
{acdn, anatali, darocha, ght}@cos.ufrj.br

Abstract. *Software testing can be applied during the software development with the purpose of improving the final product quality. However, there is a gap between the state of the art and the state of the practice regarding software testing area. There are just a few evidences concerned with effectiveness of software testing practices suggested by the technical literature in the industry. Based on this context, this paper shows the characterization of the software testing actual state for a specific software development scenario, where it has been evaluated what practices are being applied by these organizations. The results have shown the lack, limitation, and interest of these organizations in the applying of testing for their software projects.*

Resumo. *As atividades de teste estão sempre presentes no desenvolvimento de um software com o objetivo de contribuir para a qualidade do produto final. No entanto, percebe-se a existência de uma grande diferença entre o estado da arte e o estado da prática de teste de software. Existem poucas evidências relatadas da efetividade das práticas de teste de software sugeridas pela literatura técnica na indústria. Baseado neste contexto, este artigo apresenta a caracterização do estado atual das práticas de teste em um cenário de desenvolvimento de software específico, onde foram avaliadas quais práticas de teste estão sendo adotadas pelas organizações. Os resultados indicam carência, limitação e o interesse dessas organizações em aplicar práticas de teste de software em seus projetos.*

1. Introdução

Teste de Software é uma das técnicas de Garantia da Qualidade de Software e consiste na análise dinâmica do produto, ou seja, na sua execução com o objetivo de provocar a falha nesse produto, contribuindo para uma futura detecção de defeitos através de um processo de depuração e, conseqüentemente, o aumento da confiança de que ele esteja correto (adaptado de [Rocha *et al.* 2001]).

As atividades de teste possuem um papel fundamental no processo de desenvolvimento de um software, pois correspondem a um último momento de corrigir eventuais problemas no produto antes da sua entrega ao usuário final [Pressman 2005]. Os testes de software devem ocorrer em diferentes níveis (teste de unidade, integração, sistema ou aceitação), de acordo com a atividade do processo de desenvolvimento.

No entanto, apesar de notável importância das atividades de teste de software, segundo Bertolino [Bertolino 2004], o cenário atual de teste de software apresenta a existência de uma grande distância entre o que é desenvolvido na comunidade acadêmica e o que é utilizado em organizações de software. Isso pôde ser observado também através da experiência do grupo de Engenharia de Software da COPPE/UFRJ durante suas atividades de consultoria em implantação de processos em organizações de software brasileiras [Rocha *et al.* 2005].

¹ Bolsista FAPEAM.

² Bolsista FAPERJ.



A literatura técnica da área de teste de software descreve um conjunto de práticas necessárias para a realização dos testes. Essas práticas refletem no desenvolvimento de diversas abordagens ou técnicas para apoiar tarefas relacionadas às atividades de teste que dificilmente chegam à indústria [Frankl *et al.* 1997; Kim *et al.* 2000]. Além disso, há poucos indícios na literatura da aplicação, efetividade e importância dessas práticas de teste em ambientes reais de desenvolvimento, o que dificulta a caracterização do estado da prática das atividades de teste de software.

Baseado neste contexto, este artigo apresenta uma avaliação do estado da prática das atividades de teste de software em organizações brasileiras localizadas em cenário de desenvolvimento de software específico. Esta avaliação foi obtida a partir dos resultados de um estudo (pesquisa de opinião ou *survey*), onde foi analisado um conjunto de práticas de teste de software extraídas da literatura em relação a sua aplicabilidade nessas organizações e à real importância dessas práticas durante o desenvolvimento de software, segundo a opinião dos profissionais que atuam diariamente nessas organizações. O objetivo do estudo é analisar se algumas práticas de teste de software consideradas “importantes” na literatura são também consideradas importantes pelos profissionais desta área.

Os resultados obtidos neste estudo possibilitaram uma avaliação de quais práticas de teste estão sendo adotadas por essas organizações e qual o nível de aplicação delas. Além disso, foi avaliada pelos profissionais a importância dessas práticas em seus projetos. Por fim, analisou-se a relação entre o nível de aplicabilidade e o grau de importância das práticas de teste nessas organizações. Isso revelou o nível de satisfação dos profissionais em relação ao que está sendo utilizado atualmente pelas organizações, além do interesse em aplicar novas práticas ainda não adotadas.

Sendo assim, este artigo está organizado da seguinte forma: na seção 2, são descritas informações sobre o planejamento do estudo aplicado nas organizações, seu objetivo, práticas de teste avaliadas, cenário onde o estudo foi realizado e o mecanismo para análise dos dados coletados. Na seção 3, são apresentados os dados obtidos através da execução do estudo. A avaliação da relação entre a aplicabilidade e o grau de importância das práticas de teste de software é descrita na seção 4, onde são identificadas quais práticas são aplicadas (ou não) e quais são importantes (ou não) nessas organizações. Por fim, na seção 5 são descritas as conclusões e trabalhos futuros que darão continuidade a este trabalho.

2. Definição e Planejamento do Estudo

O estudo foi aplicado nas organizações com o objetivo de analisar um conjunto de práticas de teste de software extraídas da literatura, com relação à sua **aplicabilidade e importância** nessas organizações sob o ponto de vista de gerentes e desenvolvedores de software.

A avaliação da **aplicabilidade** busca revelar quais práticas de teste de software estão sendo adotadas nas organizações, e o nível de aplicação dessas práticas nessas organizações, ou seja, se elas estão sendo aplicadas eventualmente, constantemente ou não estão sendo aplicadas.

A avaliação da **importância** busca revelar a opinião dos profissionais dessas organizações, de diferentes perfis, grau de formação e nível de experiência, em relação à real importância de cada prática de teste durante o desenvolvimento de um software.

Assim, pretende-se avaliar as seguintes questões gerais:

- Quais práticas de Teste de Software são aplicadas pelas organizações?
- Quais práticas de Teste de Software são julgadas importantes pelos profissionais que trabalham no desenvolvimento de software?



2.1. Descrição da Instrumentação

O ponto crucial no planejamento do estudo foi a definição do conjunto de práticas relativas a Teste de Software que seriam avaliadas nas organizações. Com isso, o questionário aplicado nas organizações foi elaborado a partir de informações descritas em três trabalhos disponíveis na literatura que descrevem e avaliam práticas de teste de software [Andersson e Runeson 2002; Craig e Jaskiel 2002; Gelperin e Hetzel 1998], e na experiência científica e prática do grupo de Engenharia de Software da COPPE/UFRJ na área de Qualidade de Software. As práticas avaliadas estão descritas na Tabela 1 e dividem-se nas seguintes categorias:

- **Genérico:** indicam simplesmente o nome de tipos de testes a serem realizados.
- **Organização:** práticas relacionadas à organização dos testes de software ou decisões tomadas pelos responsáveis por projetos na empresa.
- **Planejamento:** práticas relacionadas às atividades de planejamento dos testes.
- **Controle:** práticas relacionadas ao controle dos testes de software em um projeto ou em uma organização.
- **Medição e Análise:** práticas relacionadas às atividades de medição e análise dos testes, como mecanismo para obter uma maior interpretação dos dados coletados durante os testes, possibilitando uma melhor interpretação dos resultados para que sejam propostas melhorias constantes nas atividades de teste.
- **Ferramentas:** práticas relacionadas à utilização de ferramentas que apoiem a realização de alguma das categorias anteriores.

Tabela 1. Lista de Práticas de Teste de Software avaliadas no estudo.

CATEGORIA	PRÁTICA DE TESTE DE SOFTWARE AVALIADA
GENÉRICO	Realização de Teste de Unidade.
	Realização de Teste de Integração.
	Realização de Teste de Sistema.
	Realização de Teste de Aceitação.
	Realização de Teste de Regressão.
ORGANIZAÇÃO	Existência de um responsável ou uma equipe alocada para as atividades de teste.
	Realização de treinamentos sobre testes.
	Separação das atividades de teste do desenvolvimento.
	Existência de testadores em tempo integral para realização de testes.
	Utilização de metodologia ou processo para sistematizar e organizar testes de software.
PLANEJAMENTO	Re-execução dos testes quando o software é modificado.
	Documentação do Plano de Teste descrevendo objetivos/abordagens.
	Documentação de Procedimentos e Casos de Teste e seus Resultados.
	Identificação e utilização de riscos para projetar, organizar e executar testes de software.
	Desenvolvimento de testes de software antes da codificação.
CONTROLE	Registro do tempo gasto em testes.
	Registro de falhas e defeitos detectados durante os testes.
	Conservação de registro (log) dos testes executados.
	Conservação dos dados dos testes para usos futuros.
	Monitoração da aderência ao processo de teste de software.
MEDIÇÃO E ANÁLISE	Avaliação dos documentos de teste quanto à sua qualidade.
	Medição dos testes e rastreamento dos custos com os testes.
	Medição da densidade de defeito (defeitos por mil linhas de código).
	Análise regular do padrão (tendência) de falhas e defeitos encontrados.
	Análise e medição da cobertura dos testes.
FERRAMENTA	Análise de defeitos identificados durante os testes para descobrir suas causas e como foram introduzidos (rastreadibilidade).
	Disponibilidade de um banco de dados de teste (pacote de testes que simulam o ambiente de teste) para reutilização.
	Utilização de ferramentas geradoras de procedimentos/casos de testes.
	Utilização de ferramentas de gerência de teste para rastrear e registrar os resultados da execução de teste.
	Utilização de ferramentas para estimar esforço de teste ou cronograma.
	Utilização de ferramentas de gerência de teste para acompanhar as atividades do processo de teste e os artefatos produzidos.
	Utilização de ferramentas de medida de cobertura nas fases de teste (unidade, integração, sistema)



Para avaliação do nível de aplicabilidade de uma prática de teste em uma organização, foram definidas escalas ordinais que variam de 0 a 4. O objetivo foi obter respostas mais precisas sobre a sua aplicabilidade nas organizações. As respostas expressam um aumento gradativo no nível de aplicação de uma prática nos projetos da organização:

- (0) Não Aplicável: Não faz parte do contexto da organização.
- (1) Não Utilizado pela organização.
- (2) Uso Não Frequente: Usado poucas vezes.
- (3) Uso Habitual: Usado em grande parte dos projetos de software da organização.
- (4) Uso Padrão: Usado em todos os projetos de software da organização.

Para avaliação do grau de importância de uma prática de teste, também se adotou uma escala ordinal composta por 5 opções pelos mesmos motivos da escala de aplicabilidade. As escalas expressam um aumento gradativo no grau de importância de uma prática na opinião de cada profissional da organização:

- (0) Não Importante: Não necessário.
- (1) Valor Baixo: Pouco importante.
- (2) Valor Limitado: Poderia ser adequado utilizá-lo.
- (3) Valor Significante: Prática recomendada.
- (4) Valor Essencial: Deve ser uma prática padrão para todos.

Durante a realização do estudo, cada profissional respondeu uma opção na escala de aplicabilidade e uma opção na escala de importância para cada prática de teste de software, de acordo com o questionário apresentado parcialmente na Figura 1.

APLICABILIDADE		GRAU DE IMPORTÂNCIA	
0	Não Aplicável	0	Não Importante: Não necessário
1	Não Utilizado	1	Valor Baixo: Pouco importante
2	Uso Não Frequente: Usado poucas vezes	2	Valor Limitado: Poderia ser adequado utilizá-lo
3	Uso Habitual: Usado em grande parte dos projetos de software	3	Valor Significativo: Prática recomendada
4	Uso Padrão: Usado em todos os projetos de software	4	Valor Essencial: Deveria ser prática padrão para todos os projetos

Caracterização das Atividades e Procedimentos de Validação (testes) de Software

PASSO 6: Em relação às atividades de validação, responda:

Descrição da atividade / procedimento	Indique o valor da escala adequada	
	Aplicabilidade	Importância
A1 Existência de um responsável ou uma equipe alocada para as atividades de teste.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A2 Realização de Teste de Unidade.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A3 Realização de Teste de Integração.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A4 Realização de Teste de Sistema.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A5 Realização de Teste de Aceitação.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A6 Realização de Teste de Regressão.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A7 Registro do tempo gasto em testes.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A8 Medição dos testes e rastreamento dos custos com os testes.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
A9 Registro de falhas e defeitos detectados durante os testes.	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4

Figura 1. Parte do questionário aplicado para avaliação das práticas de Teste de Software

2.2. Cenário de Aplicação do Estudo

O estudo foi conduzido inicialmente em organizações de software localizadas em um pólo de desenvolvimento de software brasileiro. Este cenário foi selecionado por conveniência, facilitado pelo incentivo oferecido pela Agência de Fomento a Pesquisa do Estado onde ocorreu a pesquisa (FAPEAM³), aliado ao interesse das organizações deste pólo em participar desta pesquisa.

³ Fundação de Amparo à Pesquisa do Estado do Amazonas – www.fapeam.am.gov.br.



2.3. Análise dos Dados

A seguir serão descritos os procedimentos de análise dos dados coletados para as duas características que foram avaliadas: *aplicabilidade* e *importância*.

2.3.1 Análise da Aplicabilidade das Práticas de Teste de Software

A avaliação da aplicabilidade indica como as práticas de teste estão sendo utilizadas nos projetos de software das organizações. Com isso, para cada prática avaliada, torna-se necessário obter um resultado único para cada organização visitada, pois uma organização pode ter mais que um participante no estudo. O objetivo é analisar a distribuição das respostas obtidas em cada organização entre as escalas possíveis (as escalas variam de 0 a 4, conforme descrito na seção 2.1). Para isso, foram seguidos alguns passos na análise da aplicabilidade das práticas de teste:

- (1) Para cada item do questionário, obteve-se a moda (o resultado mais frequente) entre os valores obtidos em uma organização. Este procedimento foi adotado pois a escala é ordinal, que sugere a utilização da moda como mecanismo de análise dos dados de uma amostra [Li e Conradi 2005].
- (2) Em seguida, os dados de todas as organizações visitadas foram agrupados e distribuídos entre as escalas para cada prática de teste avaliada. Com isso, é definida a análise quantitativa da aplicabilidade das práticas de teste.

2.3.2 Análise do Grau de Importância dos Itens Avaliados

A avaliação do grau de importância indica a opinião exclusiva dos profissionais que responderam aos questionários. Com isso, para cada prática de teste de software avaliada, torna-se necessário obter o seu grau de importância considerando a formação e experiência de cada participante. Os passos para obtenção do grau de importância são:

- (1) Para diferenciar as respostas dos participantes, foi atribuído um peso a eles, considerando-se o seu nível de formação, tempo de atuação e grau de experiência nas atividades de revisão e teste de software. A fórmula usada para definição do peso de um participante foi baseada na proposta de [Farias 2002]:

$$PT(i) = \frac{TA(i)}{MedianaTA} + \frac{QP(i)}{MedianaQP} + f(i) + g(i) + h(i), \text{ onde:}$$

- $PT(i)$ é o peso total de um participante i ;
 - $TA(i)$ é o tempo de atuação na área de desenvolvimento de software de um participante i ;
 - $MedianaTA$ é a mediana do tempo de atuação, considerando o tempo de atuação de cada participante;
 - $QP(i)$ é a quantidade de projetos que um participante i já esteve envolvido;
 - $MedianaQP$ é a mediana da quantidade de projetos que um participante i já esteve envolvido, considerando o número de projetos de todos participantes;
 - $f(i)$ é a formação acadêmica na área de Ciência da Computação, sendo considerados os seguintes valores: (0) nível técnico, (1) nível superior, (2) especialização, (3) mestrado ou (4) doutorado;
 - $g(i)$ é o conhecimento do participante em relação à revisão⁴ de software, e $h(i)$ é o conhecimento em relação à teste de software, sendo considerados os seguintes valores: (0) nenhum, (1) baixo, (2) médio, (3) alto ou (4) excelente;
- (2) Em seguida, multiplicou-se a resposta de cada participante (as respostas variam entre 0 e 4, conforme descrito na seção 2.1) pelo seu peso e calculou-se o valor total para cada prática de teste avaliada.

⁴ O estudo avaliou originalmente práticas de revisão e teste de software, porém só os resultados de teste de software serão discutidos neste artigo.



- (3) Por fim, normalizou-se o valor obtido no passo 2 para cada prática de teste, ou seja, dividiu-se o valor obtido no passo anterior pelo valor máximo possível: $\text{VALOR OBTIDO} / (\text{PESO TOTAL} \times 4 [\text{escala máxima}])$, obtendo assim o grau de importância de cada prática de teste, que é um valor expresso em percentual.

A seguir, serão apresentados os dados coletados durante a aplicação do estudo.

3. Dados Coletados no Estudo

O estudo foi conduzido no período de 28/06/2005 a 13/07/2005 através de visitas às organizações de software contatadas e que se interessaram em participar. Os questionários foram preenchidos sempre na própria organização em que os profissionais atuavam e com o consentimento da alta direção. Ao final deste período, foram obtidos os seguintes dados:

- **Organizações de Software visitadas:** 13 organizações que realizam teste de software em seus projetos. As organizações que participaram do estudo desenvolvem diferentes tipos de software e possuem diferentes tamanhos (Figura 2). A definição das categorias de organização de acordo com seu tamanho (micros, pequenas, médias e grandes empresas) seguiu o critério estabelecido pela Secretaria de Projetos de Informática – SEPIN – do Ministério da Ciência e Tecnologia – MCT [SEPIN 2005].

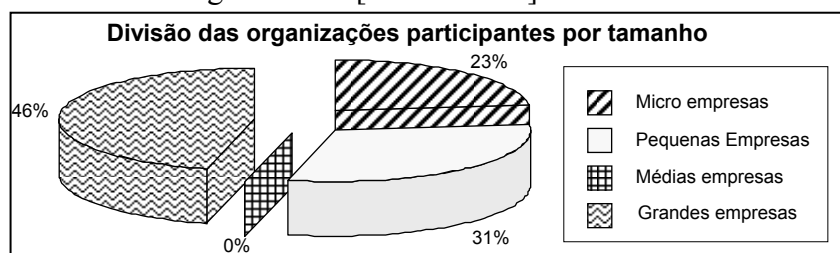


Figura 2. Divisão das organizações que participaram do estudo por tamanho

- **Questionários Respondidos:** 36 questionários (distribuídos entre as 13 organizações visitadas), entre gerentes de projeto e desenvolvedores. Além disso, os profissionais que participaram do estudo possuem diferentes níveis de formação e conhecimento nas atividades de Revisão e Teste de Software, conforme visualizado nos gráficos apresentados na Figura 3.

Agrupando-se os dados coletados ao longo do estudo, os resultados obtidos revelam a aplicabilidade das práticas de teste de software nas organizações visitadas e o grau de importância dessas práticas, segundo os seus profissionais.

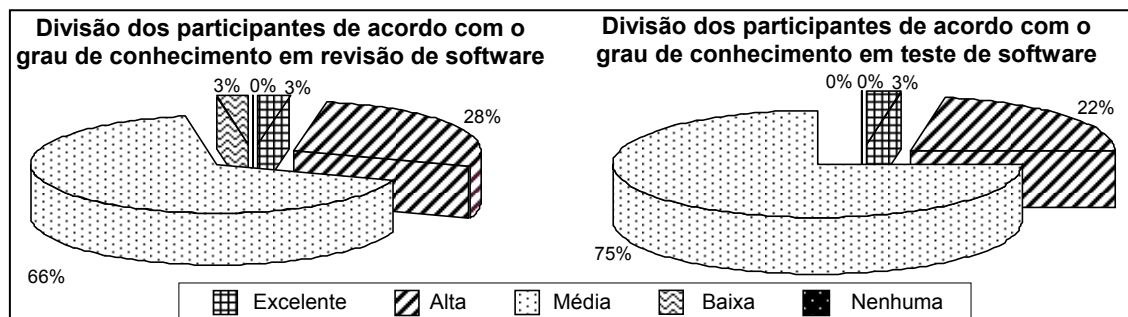


Figura 3. Divisão dos profissionais que participaram do estudo por experiência

3.1. Avaliação da Aplicabilidade dos Itens nas Organizações de Software

A partir da execução dos passos descritos na seção 2.3.1 com os dados coletados, inicialmente avaliaram-se os resultados de todas as organizações simultaneamente, ou seja, sem a divisão da amostra em grupos. Em seguida, avaliaram os dados através de



grupos de acordo com o tamanho das organizações. Com isso, a divisão da amostra ocorreu da seguinte forma:

- 7 organizações ficaram no grupo de micro e pequenas empresas;
- 6 organizações ficaram no grupo de grandes empresas;

A Tabela 2 apresenta a contabilização da quantidade de organizações em cada escala de aplicabilidade (0 a 4) para cada amostra do estudo (micros e pequenas empresas, grandes empresas e todas as organizações juntas).

Tabela 2. Resultados de Aplicabilidade obtidos no estudo.

Divisão da quantidade de empresas por escala para cada amostra do estudo.		MIC E PEQ					GRANDE					GERAL				
As escalas são: (E0) Não aplicável, (E1) Não Utilizado, (E2) Uso Não-frequente, (E3) Uso Habitual e (E4) Uso Padrão.		Qtde de Empresas por Escala					Qtde de Empresas por Escala					Qtde de Empresas por Escala				
PRÁTICAS RELACIONADAS Á TESTE DE SOFTWARE		E0	E1	E2	E3	E4	E0	E1	E2	E3	E4	E0	E1	E2	E3	E4
GENÉRICO	Realização de Teste de Unidade.	0	0	1	2	4	0	2	1	1	2	0	2	2	3	6
	Realização de Teste de Integração.	0	0	1	2	4	0	0	1	2	3	0	0	2	4	7
	Realização de Teste de Sistema.	0	0	1	2	4	0	0	0	2	4	0	0	1	4	8
	Realização de Teste de Aceitação.	0	1	2	3	1	1	0	2	0	3	1	1	4	3	4
	Realização de Teste de Regressão.	1	2	2	1	1	1	0	2	1	2	2	2	4	2	3
ORGANIZAÇÃO	Existência de um responsável ou uma equipe alocada para as atividades de teste.	0	3	0	2	2	0	0	0	2	4	0	3	0	4	6
	Realização de treinamentos sobre testes.	0	5	1	1	0	0	0	4	0	2	0	5	5	1	2
	Separação das atividades de teste do desenvolvimento.	0	3	1	2	1	0	0	1	3	2	0	3	2	5	3
	Existência de testadores em tempo integral para realização de testes.	1	4	0	0	2	0	0	2	1	3	1	4	2	1	5
	Utilização de metodologia ou processo para sistematizar e organizar testes de software.	0	2	1	2	2	0	0	1	2	3	0	2	2	4	5
PLANEJAMENTO	Re-execução dos testes quando o software é modificado.	0	1	3	0	3	0	0	2	1	3	0	1	5	1	6
	Documentação do Plano de Teste descrevendo objetivos/abordagens.	0	1	1	2	3	0	0	0	2	4	0	1	1	4	7
	Documentação de Procedimentos/Casos de Teste e os resultados	0	1	1	2	3	0	0	1	1	4	0	1	2	3	7
	Identificação e utilização de riscos para projetar, organizar e executar testes de software.	0	2	4	1	0	1	2	2	1	0	1	4	6	2	0
	Desenvolvimento de testes de software antes da codificação.	0	3	3	1	0	0	3	2	1	0	0	6	5	2	0
CONTROLE	Registro do tempo gasto em testes.	1	3	1	2	0	1	0	3	0	2	2	3	4	2	2
	Conservação de registro (log) dos testes executados.	0	3	0	3	1	0	0	1	1	4	0	3	1	4	5
	Conservação dos dados dos testes para usos futuros.	0	3	1	2	1	0	0	1	2	3	0	3	2	4	4
	Registro de falhas e defeitos detectados durante os testes.	0	1	0	3	3	0	0	0	2	4	0	1	0	5	7
	Monitoração da aderência ao processo de teste de software.	0	4	2	0	1	0	2	1	2	1	0	6	3	2	2
MÉDICAÇÃO e ANÁLISE	Avaliação dos documentos de teste quanto à sua qualidade.	1	2	2	2	0	0	1	0	3	2	1	3	2	5	2
	Medição dos testes e rastreamento dos custos com os testes.	1	3	3	0	0	1	2	2	1	0	2	5	5	1	0
	Medição da densidade de defeito (defeitos por mil LOC).	0	6	0	1	0	1	2	2	0	1	1	8	2	1	1
	Análise regular do padrão (tendência) de falhas e defeitos encontrados.	1	3	2	1	0	0	1	2	2	1	1	4	4	3	1
	Análise e medição da cobertura dos testes.	0	5	1	0	1	0	1	1	2	2	0	6	2	2	3
FERRAMENTA	Análise de defeitos identificados durante testes para descobrir suas causas e como foram introduzidos (rastreadibilidade).	0	3	1	2	1	0	1	2	1	2	0	4	3	3	3
	Disponibilidade de um banco de dados de teste (pacote de testes que simulam o ambiente de teste) para reutilização.	0	4	2	0	1	0	1	3	0	2	0	5	5	0	3
	Utilização de ferramentas geradoras de procedimentos/casos de testes.	0	4	3	0	0	1	2	2	0	1	1	6	5	0	1
	Utilização de ferramentas de gerência de teste para rastrear e registrar os resultados da execução de teste.	0	5	1	1	0	0	2	2	0	2	0	7	3	1	2
	Utilização de ferramentas para estimar esforço de teste ou cronograma.	0	6	1	0	0	0	2	2	1	1	0	8	3	1	1
FERRAMENTA	Utilização de ferramentas de gerência de teste para acompanhar as atividades do processo de teste e os artefatos produzidos.	0	5	2	0	0	0	2	1	1	2	0	7	3	1	2
	Utilização de ferramentas de medida de cobertura nas fases de teste (unidade, integração, sistema).	0	4	2	0	1	0	2	2	0	2	0	6	4	0	3

Esta divisão ocorreu pelo seguinte motivo: com a avaliação global dos dados obtidos, isto é, sem separá-los em grupos, percebeu-se que os resultados de determinados itens eram bastante dispersos, ou seja, existem organizações que sempre aplicam um item e outras que nunca aplicam o mesmo item.



Com isso, optou-se pela divisão da amostra em grupos de acordo com o tamanho de cada organização. O objetivo desta divisão é reduzir o viés causado por grandes empresas, que aparentemente tendem a aplicar mais práticas de Engenharia de Software em seus projetos, pois eles são normalmente mais complexos que os projetos de micros e pequenas empresas e estas possuem mais recursos disponíveis.

3.2. Avaliação do Grau de Importância pelos profissionais

Para avaliação do grau de importância, foram seguidos os passos descritos na seção 2.3.2. Ao final, chegou-se aos dados apresentados na Tabela 3, onde são descritos os resultados obtidos analisando todas as organizações em um único grupo e de acordo com os grupos definidos na seção 3.1.

Tabela 3. Resultados de Grau de Importância obtidos no estudo.

ITENS RELACIONADOS Á VALIDAÇÃO DE SOFTWARE		GRAU DE IMPORTÂNCIA		
		MIC E PEQ	GRANDE	GERAL
GENÉRICO	Realização de Teste de Unidade.	90,80%	92,82%	92,02%
	Realização de Teste de Integração.	90,80%	92,02%	91,54%
	Realização de Teste de Sistema.	89,31%	92,08%	90,98%
	Realização de Teste de Aceitação.	92,34%	97,63%	95,53%
	Realização de Teste de Regressão.	85,70%	83,43%	84,33%
ORGANIZAÇÃO	Existência de um responsável ou uma equipe alocada para as atividades de teste.	89,91%	100,00%	95,98%
	Realização de treinamentos sobre testes.	76,84%	91,54%	85,69%
	Separação das atividades de teste do desenvolvimento.	84,71%	92,35%	89,31%
	Existência de testadores em tempo integral para realização de testes.	74,07%	92,19%	84,98%
	Utilização de metodologia ou processo para sistematizar e organizar testes de software.	76,45%	82,43%	80,05%
	Re-execução dos testes quando o software é modificado.	81,32%	86,38%	84,36%
PLANEJAMENTO	Documentação do Plano de Teste descrevendo objetivos/abordagens.	80,53%	84,13%	82,69%
	Documentação de Procedimentos e Casos de Teste e seus Resultados.	83,43%	69,33%	74,95%
	Identificação e utilização de riscos para projetar, organizar e executar testes de software.	89,65%	98,83%	95,17%
	Desenvolvimento de testes de software antes da codificação.	82,71%	94,28%	89,67%
	Registro do tempo gasto em testes.	73,18%	89,41%	82,95%
	Registro de falhas e defeitos detectados durante os testes.	87,01%	93,53%	90,94%
	Conservação de registro (<i>log</i>) dos testes executados.	70,59%	84,41%	78,91%
CONTROLE	Conservação dos dados dos testes para usos futuros.	74,09%	92,00%	84,87%
	Monitoração da aderência ao processo de teste de software.	86,33%	98,83%	93,85%
	Avaliação dos documentos de teste (<i>testware</i>) quanto à sua qualidade.	87,45%	92,28%	90,36%
	Medição dos testes e rastreamento dos custos com os testes.	68,43%	80,40%	75,63%
MEDIÇÃO E ANÁLISE	Medição da densidade de defeito (defeitos por mil linhas de código).	61,65%	61,67%	61,66%
	Análise regular do padrão (tendência) de falhas e defeitos encontrados	63,49%	82,20%	74,75%
	Análise e medição da cobertura dos testes.	81,63%	78,38%	79,67%
	Análise de defeitos identificados durante os testes para descobrir suas causas e como foram introduzidos (rastreadibilidade).	78,74%	78,16%	78,39%
	Disponibilidade de um banco de dados de teste (pacote de testes que simulam o ambiente de teste) para reutilização.	74,67%	88,38%	82,92%
FERRAMENTA	Utilização de ferramentas geradoras de procedimentos/casos de testes.	82,22%	74,60%	77,63%
	Utilização de ferramentas de gerência de teste para rastrear e registrar os resultados da execução de teste.	77,29%	72,63%	74,48%
	Utilização de ferramentas para estimar esforço de teste ou cronograma.	72,12%	84,48%	79,56%
	Utilização de ferramentas de gerência de teste para acompanhar as atividades do processo de teste e os artefatos produzidos.	90,56%	95,11%	93,30%
	Utilização de ferramentas de medida de cobertura nas fases de teste (unidade, integração, sistema).	80,84%	86,82%	84,44%

4. Avaliação da Aplicabilidade x Grau de Importância

A partir dos dados descritos nas Tabelas 2 e 3, foi avaliada a relação entre a aplicabilidade das práticas de teste e o seu grau de importância. O objetivo foi analisar



qualitativamente os resultados obtidos e desta forma obter indicações sobre o cenário atual das práticas de teste de software nessas organizações. Abordagem semelhante foi utilizada em [Xexéo 2001] para entender a relação entre o nível de importância e de percepção de fatores críticos de qualidade em software médico. Com isso, surgem quatro grupos possíveis: (1) *aplicadas e importantes*; (2) *não-aplicadas e importantes*; (3) *aplicadas e não-importantes*; e (4) *não-aplicadas e não-importantes*. Para a definição das práticas que compõem cada grupo, utilizou-se o seguinte critério:

- Considera-se uma prática de teste *aplicada* aquela que a soma da quantidade de organizações que obtiveram as escalas 3 (Uso Habitual) ou 4 (Uso Padrão) for maior que a soma das organizações que obtiveram as demais escalas de aplicabilidade. Caso contrário ela será considerado *não-aplicada*.
- Para definir se uma prática de teste é considerada *importante*, obtiveram-se os quartis associados aos resultados do grau de importância (Tabela 3), conforme expresso na Tabela 4. A partir disso, somente as práticas cujo grau de importância está acima do segundo quartil, ou seja, está acima da mediana, são consideradas *importantes* neste estudo. Caso contrário, a prática é considerada *não-importante*.

Tabela 4. Divisão dos resultados obtidos no contexto geral em quartis.

Quartis da distribuição	Valor	Classificação
Quartil 0 (valor mínimo da amostra)	até 61,66%	Não Importante
1º Quartil	até 79,56%	
2º. Quartil (mediana)	até 84,43%	
3º. Quartil	até 90,94%	Importante
4º. Quartil (valor máximo da amostra)	até 95,98%	

A divisão desses grupos permitirá uma maior interpretação dos resultados obtidos. Ela pode revelar a satisfação das organizações em relação ao que está sendo aplicado e as reais necessidades e carências dessas organizações em relação às práticas de teste de software, seguindo este raciocínio:

- **APLICADA E IMPORTANTE:** isso indica que essa prática de teste vem apresentando resultados positivos quando aplicada nas organizações. Desta forma, os profissionais a julguem importante e deverão continuar a utilizá-la em seus projetos. As pesquisas na área de teste de software devem prover meios para que ela continue a ser aplicada nas organizações.
- **APLICADA E NÃO-IMPORTANTE:** isso indica que essa prática de teste não vem apresentando resultados positivos quando aplicada nas organizações. Desta forma, as pessoas não a consideram importante em um projeto, apesar de usá-la. A tendência é que essas práticas sejam abandonadas com o tempo, por não apresentarem resultados para a organização. As pesquisas devem buscar entendimento sobre a real necessidade destas práticas para as atividades de teste.
- **NÃO-APLICADA E NÃO-IMPORTANTE:** isso indica que essa prática de teste não faz parte dos planos das organizações para as atividades de teste, por não serem consideradas importante. Os motivos relacionados a isso podem ser vários, como por exemplo, falta de conhecimento sobre a prática ou por julgar que ela não faz parte do escopo dos projetos da organização. As pesquisas na área de teste de software devem buscar mecanismos para aplicação destas práticas em ambientes reais de desenvolvimento, para avaliar a sua importância a partir dos resultados de sua implantação nesses ambientes.
- **NÃO-APLICADA E IMPORTANTE:** isso indica que essa prática, apesar de ser considerada importante, não está sendo aplicada nas organizações. Os motivos podem ser diversos, carência de recursos humanos, falta de apoio da



gerência, falta de apoio ferramental, dentre outros. Essas práticas indicam necessidade imediata das organizações em relação às atividades de teste de software. As pesquisas devem buscar mecanismos para apoiar a sua aplicação em ambientes reais de desenvolvimento.

4.1. Avaliação da Aplicabilidade x Grau de Importância no Contexto Geral

A partir da definição dos critérios para agrupamento das práticas de teste, inicialmente avaliaram-se as práticas de teste em um contexto geral, ou seja, todas as organizações juntas. Com isso, obteve-se a divisão apresentada na Tabela 5 e representada graficamente na Figura 4.

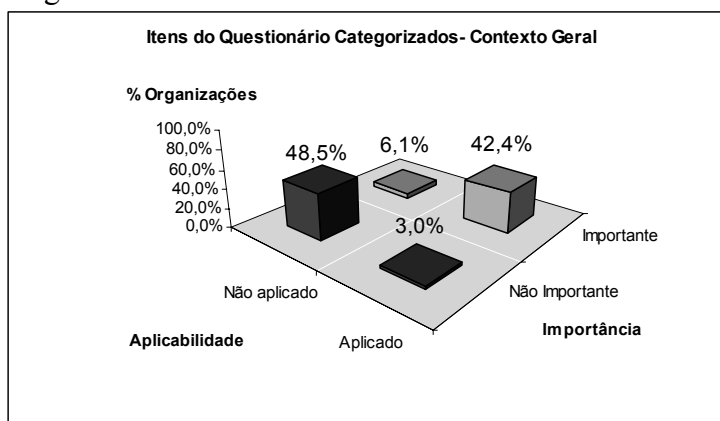


Figura 4. Avaliação de Aplicabilidade x Importância no Contexto Geral

Nesta tabela, as siglas (G),(O),(P),(C),(M) e (F) antes do título de cada prática representam as suas categorias (Genérico, Organização, Planejamento, Controle, Medição e Análise e Ferramentas, respectivamente).

Tabela 5. Divisão das práticas: Aplicabilidade x Importância no Contexto Geral.

	Não-Aplicada	Aplicada
Não-Importante	<ul style="list-style-type: none"> •(G) Realização de testes de regressão •(P) Identificação dos riscos de teste •(P) Desenvolvimento de teste antes da codificação •(C) Registro de tempo dos testes •(C) Monitoração da aderência ao processo de testes •(F) Todos os 6 itens de ferramenta (ver Tabela 1) •(M) Todos os 5 itens de medição e análise (ver Tabela 1) <p>TOTAL: 16 práticas</p>	<ul style="list-style-type: none"> • (C) Conservação dos registros (<i>log</i>) dos testes executados; <p>TOTAL: 1 prática</p>
Importante	<ul style="list-style-type: none"> • (O) Realização de treinamento sobre testes; • (O) Existência de testadores em tempo integral para realização de testes; <p>TOTAL: 2 práticas</p>	<ul style="list-style-type: none"> •(G) Realização de testes de unidade, integração, sistema e aceitação •(O) Existência de um responsável ou equipe alocada para os testes •(O) Separação entre testes e desenvolvimento •(O) Utilização de metodologia para sistematizar os testes •(O) Re-execução dos testes quando o software é modificado •(P) Documentação do Plano de Testes •(P) Documentação de procedimentos/casos de Teste e seus resultados •(C) Registro das falhas e defeitos detectados •(C) Conservação dos dados dos testes para uso futuro •(C) Avaliação dos documentos de teste <p>TOTAL: 13 práticas</p>

Avaliando-se os resultados obtidos, observa-se que a maioria das práticas de teste é considerada *não-aplicada* e *não-importante* (48,5%) ou *aplicada* e *importante* (42,4%). O alto número de práticas de teste consideradas *aplicadas* e *importantes* reflete na satisfação dos profissionais em relação ao que já está sendo utilizado em seus projetos. Essas práticas estão divididas entre as categorias *Genérico*, *Organização*, *Planejamento* ou *Controle*. No entanto, não há nenhuma categoria que possui todas as suas práticas consideradas *aplicadas*, indicando uma limitação das organizações em relação à aplicação das práticas de teste que compõem essas categorias.



Além disso, todas as práticas das categorias *Medição e Análise* e *Ferramenta* são consideradas *não-importantes* e *não-aplicadas* pelas organizações que participaram do estudo (além de outras práticas de outras categorias). Isso expressa a carência das organizações nas atividades de Medição e Análise, e na adoção de ferramentas de apoio aos testes. A grande quantidade de práticas nesse grupo pode ser ocasionada por diversos fatores, como: pouco conhecimento dos profissionais em relação às práticas de medição e análise dos testes, limitação de recursos humanos a serem alocados para essas tarefas, falta de interesse da gerência, dentre outros.

Entre os resultados, somente uma prática de teste é *aplicada* pelas organizações, mas é considerada *não-importante*: *Conservação dos registros dos testes executados*. A sua avaliação indica que as organizações contatadas não estão utilizando com sucesso os dados coletados nos testes. Isso ocorre provavelmente porque a conservação dos dados dos testes só tem sentido caso eles sejam utilizados na prevenção de futuros defeitos em novos projetos, o que ocorre através de adoção de práticas de medição e análise (que não são *aplicadas* nessas organizações, de acordo com os resultados descritos).

As práticas de teste consideradas *importantes* e *não-aplicadas* fazem parte da categoria *Organização*. Isso reflete que essas organizações já estão preocupadas com a organização dos testes, mas possuem o desejo de implantar novas práticas que auxiliam nessa organização, como a realização de treinamentos com seus profissionais e a definição de testadores em tempo integral.

4.2. Aplicabilidade x Grau de Importância em Micros e Pequenas Empresas

De forma semelhante ao que foi descrito na seção 4.1, foi avaliado o relacionamento dos resultados de aplicabilidade e de grau de importância somente com os dados das 7 micros e pequenas empresas que participaram do estudo. A divisão das práticas de teste está apresentada na Tabela 6 e representada graficamente na Figura 5.

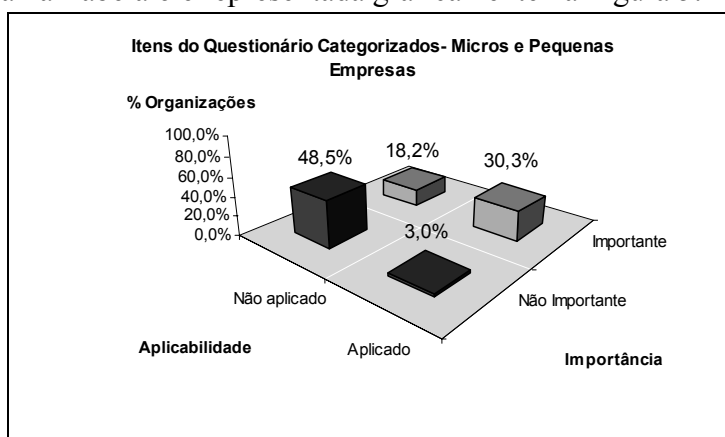


Figura 5. Aplicabilidade x Importância em Micros e Pequenas Empresas

Estes resultados apresentam um cenário similar ao apresentado no contexto geral (seção 4.1). Foi obtido o mesmo número de práticas consideradas *não-importantes* e *não-aplicadas* (16 práticas), e são basicamente as mesmas, com pequenas alterações. De forma semelhante, todas as práticas de medição e análise são consideradas *não-importantes* e *não-aplicadas*. Além disso, a mesma prática (*Conservação de registro (log) dos testes executados*) foi isoladamente considerada *não-importante* e *aplicada* nas organizações desta amostra. Porém, algumas variações puderam ser observadas.

**Tabela 6. Aplicabilidade x Importância em Micros e Pequenas Empresas.**

	Não- Aplicada	Aplicada
Não-Importante	<ul style="list-style-type: none"> • (O) Realização de treinamento sobre testes • (O) Existência de testadores em tempo integral para realização de testes • (P) Desenvolvimento de testes antes da codificação • (C) Registro do tempo gasto em testes • (C) Conservação dos dados dos testes para uso futuro • (C) Monitoração da aderência ao processo de testes de sw • (C) Avaliação dos documentos de teste quanto à sua qualidade • (M) Todos os 5 itens de medição e análise (ver Tabela 1) • (F) Disponibilidade de um banco de dados de teste (pacote de testes que simulam o ambiente de teste) para reutilização • (F) Utilização de ferramentas de gerência de teste para rastrear e registrar os resultados da execução dos testes • (F) Utilização de ferramentas de gerência de teste para acompanhar as atividades do processo de teste e os artefatos produzidos • (F) Utilização de ferramentas de medida de cobertura nas fases de teste <p>TOTAL: 16 práticas</p>	<ul style="list-style-type: none"> • (C) Conservação de registro (<i>log</i>) dos testes executados <p>TOTAL: 1 prática</p>
Importante	<ul style="list-style-type: none"> • (G) Realização de testes de regressão • (O) Separação dos testes do desenvolvimento • (O) Re-execução dos testes quando o software é modificado • (P) Identificação e utilização de riscos para projetar, organizar e executar testes de software • (F) Utilização de ferramentas geradoras de procedimentos/casos de teste • (F) Utilização de ferramentas para estimar esforço de teste ou cronograma <p>TOTAL: 6 práticas</p>	<ul style="list-style-type: none"> • (G) Realização de testes de unidade, integração, sistema e aceitação • (O) Existência de um responsável ou uma equipe alocada para as atividades de teste • (O) Utilização de metodologia ou processo para sistematizar e organizar testes de software • (P) Documentação do Plano de Testes • (P) Documentação de Procedimentos e Casos de Teste e seus resultados • (C) Registro das falhas e defeitos detectados <p>TOTAL: 9 práticas</p>

A quantidade de práticas *aplicadas* e julgadas como *importante* reduziu bastante em relação à quantidade obtida no contexto geral (de 42,4% para 30,3%). As práticas que deixaram esse grupo são referentes à organização (separação entre os testes e o desenvolvimento e re-execução dos testes após modificações no produto) e ao controle dos testes (conservação dos dados dos testes e avaliação da qualidade dos documentos de teste). Isso permite observar a limitação de micros e pequenas empresas em relação à organização e controle dos testes. Essa limitação é ocasionada possivelmente devido à carência de recursos humanos nessas organizações, o que impossibilita a definição de uma equipe específica responsável pelas atividades de testes (inclusive o seu controle).

Aliado a isso, a quantidade de práticas consideradas *importantes* e *não-aplicadas* aumentou em relação ao contexto geral (de 6,1% para 18,2%). As práticas que ingressaram neste grupo são referentes ao planejamento dos testes e ferramentas de apoio à automação de tarefas de teste (geração de casos e procedimentos de teste e geração de estimativas de esforço e cronograma). Isso demonstra o interesse destas organizações em relação ao planejamento dos testes, que é pouco aplicado nos seus projetos, e na adoção de ferramentas de automação, como mecanismo para suprir a carência de recursos humanos responsáveis pelo planejamento dos testes (que é a fase onde é realizada a geração dos casos e procedimentos de teste e a estimativa de esforço e cronograma).

4.3. Aplicabilidade x Grau de Importância em Grandes Empresas

A avaliação do relacionamento dos resultados de aplicabilidade e de grau de importância foi realizada também para os dados das 6 grandes empresas que participaram do estudo. A divisão das práticas de teste está descrita na Tabela 7 e representada graficamente na Figura 6.

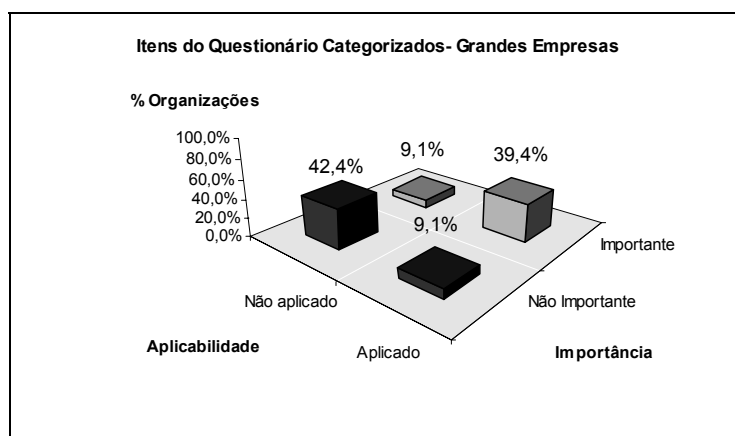


Figura 6. Aplicabilidade x Importância em Grandes Empresas

Estes resultados apresentam um cenário diferenciado em relação ao contexto geral e em relação aos resultados de micros e pequenas empresas.

Tabela 7. Aplicabilidade x Importância em Grandes Empresas.

	Não-Aplicada	Aplicada
Não-Importante	<ul style="list-style-type: none"> (G) Realização de testes de regressão (P) Identificação dos riscos de teste (P) Desenvolvimento de teste antes da codificação (C) Monitoração da aderência ao processo de testes (M) Medição dos testes e rastreamento dos custos com os testes (M) Medição da densidade de defeitos (M) Análise regular do padrão (tendência) de falhas e defeitos encontrados (M) Análise de defeitos identificados durante os testes para descobrir suas causas e como foram introduzidos (rastreadibilidade) (F) Todos os 6 itens de ferramenta (ver Tabela 1) <p>TOTAL: 14 práticas</p>	<ul style="list-style-type: none"> (C) Conservação de registro (<i>log</i>) dos testes executados (C) Avaliação dos documentos de teste quanto à sua qualidade (M) Análise e medição da cobertura dos testes <p>TOTAL: 3 práticas</p>
Importante	<ul style="list-style-type: none"> (G) Realização de testes de unidade (O) Realização de treinamento sobre testes (C) Registro do tempo gasto em testes <p>TOTAL: 3 práticas</p>	<ul style="list-style-type: none"> (G) Realização de testes de integração, sistema e aceitação (O) Existência de um responsável ou uma equipe alocada para as atividades de teste (O) Separação das atividades de teste do desenvolvimento (O) Existência de testadores em tempo integral para realização de testes (O) Utilização de metodologia ou processo para sistematizar e organizar testes de software (O) Re-execução dos testes quando o software é modificado (P) Documentação do Plano de Testes (P) Documentação de Procedimentos e Casos de Teste e seus resultados (C) Registro das falhas e defeitos detectados (C) Conservação dos dados dos testes para uso futuro <p>TOTAL: 12 práticas</p>

Nesta amostra, a quantidade de práticas *aplicadas* é bem maior que em micros e pequenas empresas (48,5% contra 33,3%). Isso permite observar que grandes empresas aplicam mais práticas de teste de software em seus projetos, possivelmente por desenvolverem softwares mais complexos e terem mais recursos disponíveis.

Praticamente todas as práticas da categoria *Organização* foram consideradas *aplicadas* e *importantes* (5 práticas entre as 6). Isso indica que grandes empresas possuem uma maior preocupação em organizar formalmente os testes que micros e pequenas por trabalharem com equipes maiores, de forma que é necessária a divisão de responsabilidades para facilitar a comunicação entre as partes envolvidas.

Nesta amostra, 3 práticas são consideradas *não-importantes*, mas são *aplicadas* nessas organizações: *Conservação de registro (log) dos testes executados*, *Avaliação dos documentos de teste quanto à sua qualidade* e *Análise e medição da cobertura dos*



testes. Essas práticas estão relacionadas ao armazenamento, medição e análise dos dados dos testes. Elas indicam uma deficiência das organizações na aplicação dessas práticas, pois o que está sendo aplicado atualmente não está apresentando resultados para a organização.

Apesar disso, alguns resultados foram similares aos obtidos no contexto geral. As práticas *não-aplicadas* e *não-importantes* estão relacionadas principalmente à medição e análise e utilização de ferramentas de apoio aos testes. Já as práticas consideradas *importantes* e *não-aplicadas* estão relacionadas à realização de testes de unidade, realização de treinamentos sobre testes e registro do tempo gasto em testes.

5. Conclusões e Trabalhos Futuros

Neste artigo foram apresentados os objetivos, a instrumentação utilizada, os dados coletados e os resultados analisados em um estudo (pesquisa de opinião) que avaliou como algumas práticas de teste de software estão sendo utilizadas em um cenário de desenvolvimento de software específico. Para isso, avaliaram-se dois aspectos: a *aplicabilidade* dessas práticas nas organizações de software, e o *grau de importância* dessas práticas na visão dos profissionais dessas organizações.

Os resultados do estudo não podem ser generalizados para outros cenários de desenvolvimento, mas permite observar que no cenário avaliado, as organizações estão aplicando práticas de teste em seus projetos, porém com limitações. Uma maior preocupação é dada à organização dos testes, enquanto que as atividades de planejamento, controle, medição e análise são pouco aplicadas na indústria de software. Observou-se também que ferramentas de apoio aos testes (relacionadas à gerência ou automação dos testes) não estão sendo aplicadas nessas organizações.

Avaliou-se também o cenário dos testes de software de acordo com o tamanho dessas organizações: micros, pequenas e grandes empresas. Com isso, observou-se que grandes empresas aplicam mais práticas de teste em seus projetos, principalmente relacionadas à organização dos testes. Por outro lado, micros e pequenas empresas possuem um maior interesse em implantar ferramentas que apóiam a automação de tarefas relacionadas ao planejamento dos testes (que é pouco aplicado nesse cenário), como forma de suprir a sua carência em recursos humanos.

Além disso, foi observada uma grande quantidade de práticas consideradas *não-importantes* e *não-aplicadas*, independentemente do tamanho das organizações. As razões para isso podem ser diversas, tais como: falta de interesse da organização em relação a essas práticas de teste, falta de conhecimento sobre as práticas, falta de recursos humanos disponíveis para a sua implantação, falta de apoio da alta gerência para sua implantação, carência em abordagens, técnicas de teste ou ferramentas que apóiem essas práticas e que atendam às reais necessidades das organizações. Esses dados expressam de certa forma uma inconsistência entre as respostas dos profissionais, pois, por exemplo no contexto geral, ao mesmo tempo que testes de regressão são considerados *não-aplicados* e *não-importantes*, a conservação dos dados dos testes para uso futuro é *aplicada* e considerada *importante*. Isso torna necessária uma avaliação mais detalhada das respostas obtidas e suas causas, a fim de entender os dados obtidos de forma qualitativa, pois a análise somente dos dados quantitativamente obtidos pode distorcer o cenário real das atividades de teste na indústria.

Com isso, neste artigo foram apresentadas indicações sobre o estado da prática das atividades de teste de software em um cenário de desenvolvimento de software. Os resultados revelam as limitações e necessidades dessas organizações, e motivam pesquisas na área de teste de software, principalmente relacionadas à definição de estratégias de apoio ao planejamento, controle, medição e análise dos testes, além da construção de infra-estruturas computacionais que visam apoiar a estas atividades.



Como próximo passo, está sendo realizada a análise dos dados referentes à caracterização de revisão de software. Os resultados iniciais indicam uma limitação nas organizações em relação à utilização de técnicas de revisão (como checklist ou técnicas de leitura), sistematização e ferramentas de apoio à revisão. Além disso, estão previstas execuções do estudo em novos cenários de desenvolvimento de software, para que seja observado como as práticas de teste de software são avaliadas em diferentes cenários. Atualmente, este estudo já foi realizado com 5 organizações do sul do país (além das 13 organizações relatadas neste artigo), porém os dados coletados ainda estão sendo analisados. Além disso, está prevista para o primeiro semestre de 2006 a sua aplicação em organizações do sudeste do país, principalmente no Estado do Rio de Janeiro.

Agradecimentos

Agradecemos a todas as pessoas que participaram do estudo e às diretorias das organizações localizadas no pólo desenvolvimento de software do Estado do Amazonas que aceitaram o convite feito. Agradecemos à FAPEAM e CNPq pelo apoio fornecido ao longo deste trabalho. Este trabalho é realizado no contexto do Projeto eSEE – CNPq (4732135/2005-0) e do Projeto “6.34: Infra-estrutura Computacional para apoiar o Planejamento e Controle de Testes de Software”, submetido ao Programa Brasileiro de Qualidade e Produtividade em Software – Ciclo 2005.

Referências Bibliográficas

- Anderson, C., Runeson, P. (2002) “Verification and Validation in Industry - A Qualitative Survey on the State of Practice”, In: International Symposium on Empirical Software Engineering (ISESE'02), p.37, October 03-04.
- Bertolino, A. (2004) “The (Im)maturity level of software testing. SIGSOFT Softw”. Eng. Notes 29, 5, 1-4, Setembro.
- Craig, R.D., Jaskiel, S. P., “Systematic Software Testing”, Artech House Publishers, Boston, 2002.
- Farias, L.D. (2002) “Planejamento de Riscos em Ambientes de Desenvolvimento de Software Orientados à Organização”, Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, Agosto.
- Frankl, P. G., Weiss, S. N., Hu, C. (1997) “All-Uses vs Mutation Testing: An Experimental Comparison of Effectiveness”, Journal of Systems and Software. Vol. 38, Pág: 235-253, Setembro.
- Gelperin, D., Hetzel, B. (1988) “The Growth of Software Testing”, Communications of the ACM, Vol. 31(6), pág: 687-695.
- Kim, J. M., Porter, A., Rothermel, G. (2000) “An Empirical Study of Regression Test Application Frequency”. In Proceedings of the 22nd International Conference of Software Engineering, Maio, Pág. 126-135. Ireland. IEEE Computer Society Press.
- Li, J., Conradi, R., *et al.* (2005) “An Empirical Study on Off-the-Shelf Component Usage in Industrial Projects”. Proc. 6th International Conference on Product Focused Software Process Improvement (PROFES'2005), Junho, Finland, Springer Verlag LNCS Vol. 3547, pp.54–68.
- Pressman, R. S., “Software Engineering: A Practitioner’s Approach”, McGraw-Hill, 6th ed, Nova York, NY, 2005.
- Rocha, A. R., Maldonado, J. C., Weber, K. C., *et al.*, “Qualidade de software – Teoria e prática”, Prentice Hall, São Paulo, 2001.
- Rocha, A.R. *et al.*, (2005) “Estação TABA: Uma Infra-estrutura para Implantação do Modelo de Referência para Melhoria de Processo de Software”, In: 3º Simpósio Brasileiro de Qualidade de Software (SBQS), Porto Alegre, RS.
- SEPIN (Secretaria de Política de Informática) – MCT (Ministério da Ciência e Tecnologia), “Qualidade e Produtividade no Setor de Software – Resultados da Pesquisa de 2005”. <http://www.mct.gov.br/sepin/Dsi/Quali2005/Public2005.htm>, acessado em Março/2006.
- Xexéo, J.A.M. (2001) “Sistemas de Informação como Instrumento de Programas de Qualidade”, Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, dezembro.