

Uma Abordagem para Caracterização dos Aspectos Não-Funcionais e Classificação do Nível de Exigência em Editais de Licitação de Software no Setor Público Brasileiro

Bruno de F. Barros¹, Reinaldo C. Silva Filho², Allan R. dos Santos Araújo¹

¹Programa de pós-graduação em Engenharia de Software
Centro de Estudos e Sistemas Avançados do Recife (CESAR) – Recife, PE – Brazil

²Núcleo de Tecnologia da Informação
Universidade Federal de Alagoas (UFAL) – Maceió, AL – Brazil
{brunobarros, reinaldo}@nti.ufal.br, arsa@cesar.org.br

Abstract. *Efforts have led to software contracts measured by the amount of functionality delivered. However, this quantity is just one of many factors to derive the cost. An instrument was established to characterize the non-functional aspects in government software bidding documents. Four experimental studies were conducted to evaluate the effectiveness of the instrument. In order to do so, 11 professionals involved in biddings participated as evaluators and Kendall's Coefficient de Concordance (W) was used. The results indicate a substantial agreement of the requirements levels and it was possible to obtain evidence of effectiveness, allowing the instrument to serve as a common mechanism for comparing bidding documents.*

Resumo. *Esforços levaram a contratos de software medidos pela quantidade de funcionalidades entregues. Porém, essa quantidade é apenas um dos muitos fatores para derivar o custo. Foi estabelecido um instrumento para caracterizar os aspectos não-funcionais em editais de software para o governo. Quatro estudos experimentais foram conduzidos para avaliar a eficácia do instrumento. Para tanto, 11 profissionais que atuam em licitações participaram como avaliadores e o Coeficiente de Concordância (W) de Kendall foi utilizado. Os resultados indicam uma concordância substancial dos níveis de exigência e foi possível obter indícios de eficácia, permitindo que o instrumento sirva de mecanismo comum para comparar editais de licitação.*

1. Introdução

Organizações públicas e privadas buscam mais eficiência no campo da Tecnologia da Informação e Comunicação (TIC) por meio da terceirização, por meio de licitações [Meli 2012]. Segundo Meli (2012), há muitas razões para isso: manter os custos fixos a um baixo nível, aumentando o nível de eficiência nos gastos públicos; obter rápido acesso a competências e tecnologias revolucionárias, etc.

No domínio da TIC, esforços levaram a disseminação de contratos cuja as atividades de desenvolvimento e manutenção de software são medidas pela quantidade de funcionalidades entregues, utilizando o método de Análise de Pontos por Função (APF) [IFPUG 2010]. Porém, a quantidade de funcionalidades é apenas um dos muitos

fatores a serem considerados para derivar o custo do projeto de software [Meli 2012], entre eles: sistemas de software, que podem ser de diferentes naturezas (aplicações web, de tomada de decisões, *business intelligence*, por exemplo), aspectos de implementação (arquitetura, linguagem de programação, ferramentas, requisitos de qualidade, restrições técnicas, e muitos outros), atendimento a cláusulas contratuais de garantia de níveis de serviço. Esses requerem esforço de desenvolvimento que diferem muito, mesmo quando a quantidade de funcionalidade a ser entregue é igual [Meli 2012].

Análises mostram que não existe correlação próxima entre Pontos de Função Não Ajustados (UFP, do inglês *Unadjusted Function Points*) e o esforço requerido para entregar o produto [Green *et al.* 2012]. Em adição, as características gerais de sistemas e o fator de ajuste de valor são controversos e criticados tanto por razões práticas como teóricas [Lokan 2000]. Acórdãos do Tribunal de Contas da União (TCU) reforçam a determinação de não usar qualquer tipo de fator de ajuste na medição por Pontos de Função (PF) na contratação de serviços de desenvolvimento de software, para impossibilitar alterações na remuneração [Brasil, Ministério do Planejamento 2015].

Ainda assim o mecanismo de estimativa de formação do preço por PF sugere que todos os aspectos não-funcionais devem ser levados em consideração para que haja uma equidade no esforço e custo necessários para uma justa entrega dos produtos de software [Meli 2012]. Porém, em primeira análise, por meio da avaliação de editais de órgãos públicos brasileiros para aquisição de serviços de desenvolvimento e manutenção de software, não foi possível identificar uma sistemática que considerasse os aspectos não-funcionais na composição do custo do serviço a ser adquirido.

No Brasil, os valores de referência estabelecidos para o preço da contratação de serviços de software para o governo se baseiam no custo do PF. Em linhas gerais, a definição desse custo é fundamentada em pesquisa de mercado (contratações similares) e pesquisa junto a fornecedores [BRASIL 2014]. Essas formas de compor o preço não consideram, de forma sistemática e repetível, os aspectos não-funcionais especificados nos termos de referência dos editais. Ao mesmo tempo, a literatura de Engenharia de Software apresenta outras alternativas para quantificar os aspectos não-funcionais e apoiar a estimativa do custo do software, porém essas também apresentam limitações, especialmente no contexto de contratações de serviços de software para o governo.

Em análise de algumas licitações, foi percebido que as exigências para seleção do fornecedor aparentemente não influenciavam no preço de referência para a contratação. Poder-se-ia inferir que quanto mais exigências a serem atendidas por uma organização na prestação do serviço, maior os seus custos para prover o serviço e, conseqüentemente, maior deveria ser o valor de referência estabelecido para a contratação. No entanto não é isso que se observa na prática. Não há, aparentemente, uma relação entre a quantidade de exigências estabelecidas pela contratante e o custo de referência para a prestação do serviço.

Eis que surgem algumas questões: os aspectos não-funcionais estão sendo considerados para definir esses valores de referência? Esse valor pode indicar qual edital apresenta mais aspectos não-funcionais a serem atendidos? Diante de alguns editais, qual deles vai requerer menor (ou maior) investimento para atender esses aspectos?

Nesse contexto, este trabalho apresenta um instrumento para caracterizar os aspectos não-funcionais explicitados em editais de licitação de software para o governo, no modelo fábrica de software. A caracterização dos aspectos não-funcionais poderá auxiliar na composição de um preço de referência mais justo tanto para o governo quanto para a empresa que fornecerá o serviço.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha o método de pesquisa para o desenvolvimento deste trabalho. A Seção 4 provê uma visão geral do instrumento baseado em heurísticas. A Seção 5 detalha o objetivo, contexto, o desenho e resultados do estudo baseado em abordagem experimental conduzido para avaliar o instrumento. Finalmente, a Seção 6 descreve as conclusões e direções futuras.

2. Trabalhos Relacionados

Há outros estudos relacionados a identificação e dimensionamento dos requisitos não-funcionais (RNF), porém com limitada aplicação em contratações de serviços de desenvolvimento e manutenção de software no contexto de licitações no setor público.

Buglione *et al.* (2008) investigaram uma predição, antecipada e ao nível do projeto, do tamanho do produto com a intenção de reduzir o efeito do “cone da incerteza”. Eles propuseram usar a técnica Project Size Unit (PSU) para prever o tamanho do produto (requisitos funcionais e RNF) medido em unidades de tamanho funcional [COSMIC 2015]. Porém, os autores concluem que sua aplicabilidade ainda carece de investigação em projetos reais e há preocupações relacionadas à validade.

Abordagem de como detectar, classificar e rastrear os RNF sugerem que esses tendem a estar implicitamente presentes em requisitos funcionais do sistema de software [Mahmoud and Williams 2016]. Tal conhecimento pode ser capturado, modelado, e utilizado para identificar RNF específicos do sistema. Porém, em editais de licitação do governo geralmente não são contempladas as especificações dos requisitos funcionais que devem ser entregues para os usuários.

Um dos esforços mais recentes na comunidade de Engenharia de Software para a avaliação de tamanho não-funcional, o Processo de Avaliação de Software Não-Funcional, ou SNAP [IFPUG 2015], limita-se a medir apenas o tamanho não-funcional do software, não considerando as tarefas relacionadas ao projeto. Esse próprio processo de avaliação afirma que, apesar de não afetarem o tamanho do software, tarefas relacionadas ao projeto afetam o esforço requerido para entregar o produto, e por consequência afetam o custo.

3. Método de Pesquisa

A abordagem foi desenvolvida em duas etapas: (i) concepção do instrumento para caracterizar o nível de exigência dos editais de licitação, e (ii) avaliação e evolução do instrumento, a qual foi realizada a partir de quatro ciclos de estudos baseados em abordagem experimental. A Figura 1 ilustra as duas etapas.

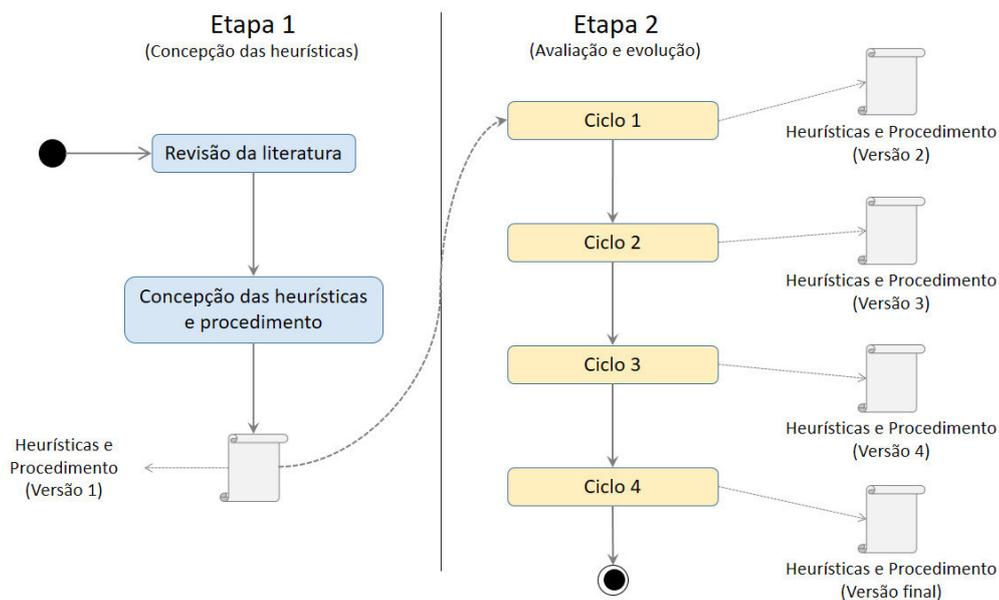


Figura 1. Etapas para o desenvolvimento da pesquisa

A pesquisa teve início na revisão da literatura para identificação de abordagens de identificação e dimensionamento de RNF. Em seguida, a partir do levantamento bibliográfico, foi elaborado o instrumento, baseado em heurísticas, para a caracterização do nível de exigência dos editais de licitação.

A segunda etapa foi realizada em quatro ciclos de avaliação e evolução, com os seguintes objetivos: (i) avaliar a instrumentação; (ii) avaliar o esforço necessário dos participantes; (iii) obter elementos para aperfeiçoar o instrumento; (iv) obter *feedback* dos participantes, e; (v) avaliar a abordagem com relação à capacidade de caracterizar adequadamente os aspectos não-funcionais dos editais. Os ciclos de avaliação e evolução foram realizados com diferentes perfis de participantes. A população objeto de estudo foi a de profissionais que participaram de alguma fase de contratações de serviços de software através de licitação de terceirização. Em cada ciclo de avaliação, selecionou-se de forma aleatória editais de licitação de desenvolvimento e manutenção de software no modelo fábrica de software. Optou-se por utilizar editais disponibilizados pelo Núcleo de Contratações de Tecnologia da Informação (NCTI)¹. Cada ciclo de avaliação e evolução foi composto pelas atividades a seguir:

- Planejamento (seleção do perfil ideal dos participantes; seleção dos editais; e instrumentação)
- Execução
- Avaliação dos resultados e *feedback* dos participantes
- Evolução do instrumento

¹ <https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/sistema-de-administracao-dos-recursos-de-tecnologia-da-informacao-sisp/ncti-nucleo-de-contratacoes-de-tecnologia-da-informacao>

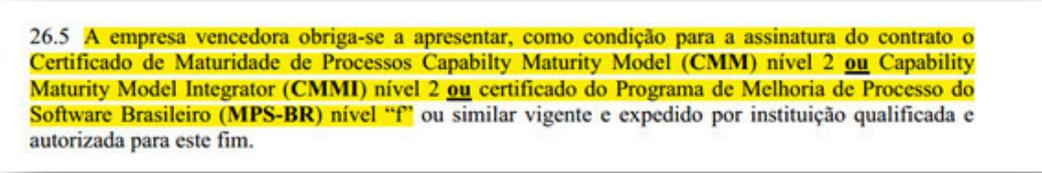
4. Heurísticas para Determinar o Nível de Exigência

As heurísticas foram concebidas para servir de mecanismo comum para comparar editais com vistas aos seus aspectos não-funcionais, e que fossem fáceis de serem aplicadas, inclusive por pessoas com pouco conhecimento técnico. Elas consistem basicamente em examinar os editais e julgar o seu nível de exigência requerido ao fornecedor a partir da própria opinião do avaliador. A seguir são apresentadas as regras heurísticas propostas.

I) Identificar exigências

Por exigências define-se obrigações relacionadas a capacidade técnica necessária para habilitar as empresas licitantes, ou seja, capacidades que as empresas já devem possuir antes do início da contratação. Essas geralmente estão expressas em forma de requisitos não-funcionais, incluindo como o software deve ser desenvolvido e mantido, e como deve se comportar quando em operação. Os requisitos não-funcionais dizem respeito a:

- A qualidade do sistema de software;
- O ambiente em que o sistema de software deve ser desenvolvido e o ambiente em que vai ser executado;
- Os processos e a tecnologia a serem usados para desenvolver e manter o sistema de software e a tecnologia a ser usada para sua execução.



26.5 A empresa vencedora obriga-se a apresentar, como condição para a assinatura do contrato o Certificado de Maturidade de Processos Capability Maturity Model (CMM) nível 2 ou Capability Maturity Model Integrator (CMMI) nível 2 ou certificado do Programa de Melhoria de Processo do Software Brasileiro (MPS-BR) nível "F" ou similar vigente e expedido por instituição qualificada e autorizada para este fim.

Figura 2. Exemplo de identificação de uma exigência

As frases das exigências identificadas *devem ser marcadas com um marca texto*. A Figura 2 ilustra como as exigências devem ser identificadas.

A fim de identificar exigências nos editais, o avaliador deve procurar declarações que indiquem *obrigatoriedades* a serem cumpridas pelo fornecedor para a contratação do serviço de software que está sendo considerado. Essas exigências também podem ser identificadas próximas a verbos, como por exemplo o verbo “dever” e suas variações (deve, deverá, devem etc.). Essas declarações são geralmente expressas em frases como “A contratada *deverá...*”, “A contratada *deve...*”. Os editais geralmente registram os critérios desejados para a contratação e contêm palavras ou frases relacionadas a esses critérios.

II) Descartar exigências de menor relevância

Na prática, algumas exigências são pouco expressivas sob o ponto de vista do investimento necessário para atendê-las, ou seja, são pequenas do ponto de vista global da contratação e há muita disponibilidade no mercado em questão para atendê-las. Por exemplo, a maioria das empresas do mercado atenderia a tais exigências sem a

necessidade de nenhum investimento. Essas exigências de menor relevância devem ser desconsideradas do nível de exigência dos editais, ou seja, não devem ser identificadas.

III) Avaliar o nível das exigências

Após a identificação de todas as exigências do edital, o avaliador deve atribuir um nível de exigência para cada exigência identificada no passo anterior.

O nível de exigência deve ser considerado “*muito comum*” quando as exigências forem fáceis de serem atendidas pela maioria das empresas atuantes no mercado nacional, independente do porte e sem considerar se elas já participaram de alguma licitação. Esse tipo de exigência não requer que a maioria das empresas realizem investimentos para atendê-las. Por exemplo, as qualificações exigidas para o pessoal são abundantes no mercado, ou a tecnologia a ser aplicada é amplamente difundida e não oferece desafios técnicos. Nesses casos, há grandes chances da maioria das empresas possuir pessoas com a referida qualificação em seu quadro e/ou já ter tido experiência com a tecnologia em questão.

No outro extremo, o nível de exigência deve ser considerado “*muito raro*” quando, por exemplo, as exigências envolverem recursos raros ou inexistentes no mercado nacional, ou seja, exigências muito difíceis de serem atendidas pelas empresas atuantes no mercado e que podem demandar altos investimentos. Algumas dessas exigências são atendidas por um grupo muito restrito de empresas. Uma empresa que desejasse atender esse nível de exigência teria que realizar grandes investimentos.

Para exigências “*muito comuns*” deve ser atribuído o número 1 (um) e, no outro extremo, para exigências “*muito raras*” deve ser atribuído o número 5 (cinco). A Figura 3 estabelece os níveis da escala de nível de exigência.

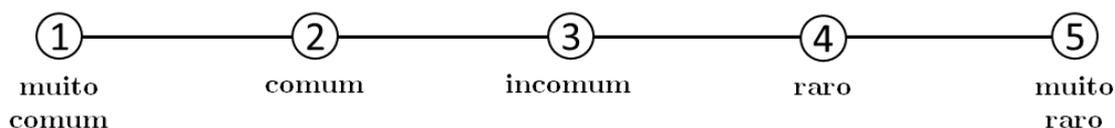


Figura 3. Escala com os níveis de exigência

A ilustração da Figura 4 contém um exemplo de como as exigências identificadas devem ser marcadas com o nível de exigência avaliado.

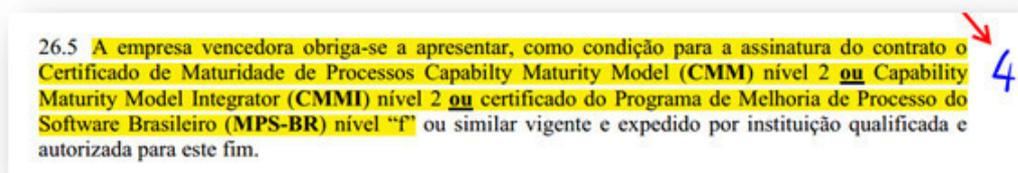


Figura 4. Exemplo de avaliação de um nível de exigência

5. Avaliação do Instrumento

Em linha com o entendimento de Shull *et al.* (2004), que estudos experimentais devem ser realizados e repetidos para melhorar a credibilidade da pesquisa, quatro avaliações baseada em abordagem experimental [Wohlin *et al.* 2000] foram executadas para guiar a evolução do instrumento.

As seções a seguir apresentam o detalhamento do quarto e último estudo baseado em abordagem experimental, realizado com o propósito de avaliar o uso do instrumento.

5.1. Definição

Utilizando o paradigma GQM [Basili *et al.* 1994], o objetivo do quarto ciclo foi declarado como:

Analisar	O instrumento
Com propósito de	Avaliar
Com respeito a	Eficácia da caracterização do nível de exigência de editais de licitação
Do ponto de vista	Profissionais que atuam em alguma etapa de licitações públicas
No contexto de	Contratação de serviços de software para o governo, na modalidade fábrica de software

Com base nesse objetivo foi formulada a seguinte questão:

Questão 1: O instrumento é eficaz na caracterização dos níveis de exigência dos editais?

O instrumento será considerado eficaz se vários participantes aplicarem o instrumento para avaliar editais de licitação, escolhidos ao acaso, e for constatada concordância com relação ao nível de exigência do edital. Essa questão é respondida de acordo com as métricas associadas, descritas a seguir:

Métrica 1: Nível de exigência, nível associado a cada exigência identificada no edital. Deve-se obter o valor associado a cada exigência identificada no edital. Esse valor segue uma escala ordinal de 1 a 5.

Métrica 2: Distribuição dos níveis de exigência do edital, quantifica a frequência de ocorrências de cada possível resultado (escala de 1 a 5) dos níveis de exigência encontrados no edital.

Métrica 3: Classificação de abundâncias dos níveis de exigência do edital, através da transformação da distribuição de frequência dos níveis de exigência em ordem de classificação dos mais abundantes, de forma a torná-los adequados para análise estatística. Essa classificação foi baseada no estudo de Legendre (2005) que propôs o uso do Coeficiente de Concordância (W) de Kendall para identificar grupos de espécies significativamente associados.

Métrica 4: Coeficiente de Concordância (W) de Kendall [Kendall and Smith 1939], é uma medida de concordância entre várias variáveis (m) quantitativas ou semiquantitativas que avaliam um conjunto de *n* objetos de interesse.

5.2. Contexto

O instrumento foi aplicado com profissionais que já participaram de alguma etapa de contratações desses serviços através de licitações, e foi avaliado a partir de editais que já

foram executados, ou seja, o resultado da concorrência pública já foi divulgado. Dessa forma, a experiência é caracterizada como *on-line* (projetos industriais), baseado em problemas do mundo real, e conta com participação de profissionais.

5.3. Variáveis

Neste estudo, as variáveis independentes são a experiência dos participantes, as exigências dos editais e o instrumento proposto. A variável dependente é a eficácia.

5.4. Formulação das Hipóteses

Dois tipos de hipóteses foram formuladas para avaliar este estudo: uma *Hipótese Nula*, que o estudo pretende rejeitar com a maior significância possível; e uma *Hipótese Alternativa*.

- *Hipótese Nula*: Não há concordância entre os níveis de exigência indicados pelos participantes a partir da aplicação do instrumento.

$$H_0 : W = 0$$

- *Hipótese Alternativa*: Há concordância entre os níveis de exigência indicados pelos participantes a partir da aplicação do instrumento.

$$H_1 : W \neq 0$$

Onde o valor de W é referente ao coeficiente de concordância de Kendall [Kendall and Smith 1939] e pode variar de 0 a 1. Quanto maior o valor de W , mais forte é a concordância. Um coeficiente de concordância de Kendall elevado ou significativo denota que os avaliadores estão aplicando essencialmente o mesmo padrão ao avaliar as amostras [Siegel and Jr. 2006].

5.5. Participantes

Os participantes foram escolhidos com base na técnica não-probabilística de amostragem por conveniência² e *snowball*³, e é composta de profissionais que participaram de alguma fase de contratações de serviços de software através de licitação.

5.6. Arranjo Experimental

A definição, hipóteses e medidas para este experimento indicam que o tipo de projeto é: “*um fator com um tratamento*”, já que o objetivo deste estudo não é avaliar o instrumento contra algum outro método existente.

A variável dependente é medida em escala ordinal, e portanto um teste não-paramétrico é indicado [Wohlin *et al.* 2000]. Neste caso em particular, o teste de Coeficiente de Concordância (W) de Kendall [Kendall and Smith 1939] é utilizado.

² Se baseia na coleta de dados de membros da população que estão convenientemente disponíveis para participar do estudo, ou seja, os membros mais acessíveis [Marshall 1996].

³ Permite que cada indivíduo da amostra seja perguntado para indicar o nome de outros indivíduos da mesma população de estudo [Goodman 1961].

5.7. Instrumentação

Os instrumentos utilizados neste estudo foram:

- Questionário para ser preenchido pelos participantes no início do experimento - Esses dados provêm a entrada para caracterização do participante.
- Material de apoio ao uso das heurísticas – consiste de documento para treinamento contendo a descrição completa das heurísticas e exemplos.
- Editais de licitação – principais seções dos editais de desenvolvimento e manutenção de software, selecionados aleatoriamente.
- Questionário para a fase final de execução do experimento. Proverá dados quantitativos e qualitativos sobre experiência dos participantes no experimento.
- Softwares de apoio a medição – os dados serão organizados em planilhas eletrônicas e pacotes estatísticos para a análise dos resultados e hipóteses.

A Tabela 1 a seguir apresenta os editais selecionados.

Tabela 1. Editais selecionados para o estudo

Nº	Ano	Órgão	UF
26	2014	ANVISA – Agência Nacional de Vigilância Sanitária	DF
11	2015	Secretaria de Portos da Presidência da República	DF

5.8. Operação

Todo o material necessário para a execução deste estudo foi preparado previamente e submetido a um avaliador externo para correção de possíveis mau-entendimentos. Em seguida iniciou-se a busca por indivíduos pertencentes a população desta pesquisa. Esses eram então convidados a participar da pesquisa com uma breve explicação do objetivo, do procedimento e do tempo necessário para execução.

A execução deste estudo foi conduzida no período de março a abril de 2017. Os dados foram coletados manualmente pelos participantes que preenchiam os questionários e seguiam as orientações do instrumento para caracterizar os aspectos não-funcionais dos editais.

Foram abordadas pessoalmente 18 pessoas para participar da pesquisa presencial, sendo que 12 aceitaram participar da avaliação e as outras seis informaram não estar disponíveis. As abordagens um-a-um foram particularmente importantes para garantir uma melhor taxa de resposta. Os participantes realizaram a avaliação sob as mesmas circunstâncias (com a presença do pesquisador durante a avaliação). As avaliações eram iniciadas com uma breve apresentação da pesquisa, em seguida os participantes leram e assinaram o Termo de Consentimento Livre e Esclarecido. Então, os participantes preencheram o questionário de caracterização e deram início a leitura das orientações contendo o instrumento para avaliar as principais seções de dois editais de licitação selecionados ao acaso. Por fim, os participantes preencheram o questionário pós-execução como o objetivo de obter opiniões e considerações quanto a utilização do instrumento para avaliar os editais.

Convites também foram enviados para participação *online*. Desses, 11 convites foram enviados diretamente para os indivíduos, e outros convites foram enviados a grupo de discussão na internet sobre o tema de contratações públicas de TI, e grupo de trabalho sobre métricas de software. A pesquisa ficou disponível por 12 dias, mas, apesar de somados esses grupos possuírem cerca de 400 membros, *apenas dois indivíduos* aceitaram participar da pesquisa.

Os dados foram coletados de 14 pessoas que concordaram em participar da pesquisa. Após a validação dos dados, admitiu-se 11 participantes para as análise e interpretação dos resultados.

5.9. Análise e Interpretação

Como primeiro passo na análise dos dados, utilizou-se estatísticas descritivas para visualizar os dados coletados. Para obter um melhor entendimento dos dados relacionados a experiência dos participantes, em função da sua dispersão e inclinação, um *boxplot* (diagrama de caixa) e gráficos de barras são apresentados na Figura 5.

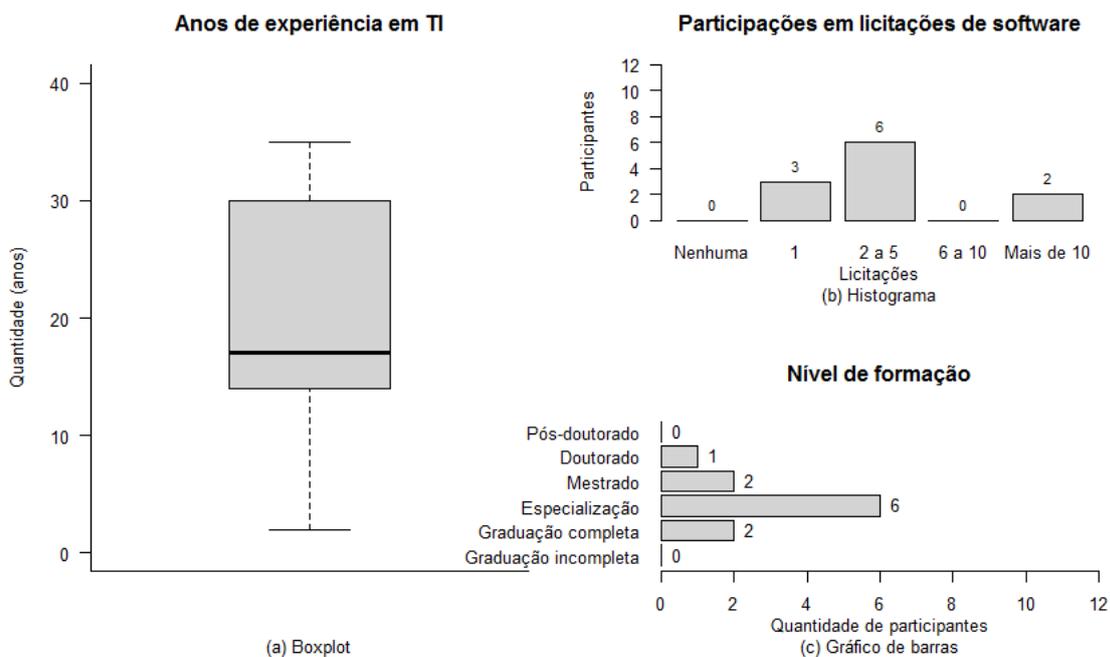


Figura 5. (a) *Boxplot* para os anos de experiência em TI dos participantes. (b) Histograma para a quantidade de participantes por participações em licitações de software. (c) Gráfico de barras com os níveis de formação dos participantes.

Conforme a Figura 5, observa-se que a maioria dos participantes possui entre 15 e 30 anos de experiência em TI, a maioria dos participantes informou já ter participado de 2 a 5 licitações e, quanto ao nível de formação, a maioria informou possuir especialização, seguido por graduação completa e mestrado.

5.9.1. Testes de Hipóteses

O Coeficiente de Concordância W de Kendall [Kendall and Smith 1939] foi utilizado para analisar os resultados e determinar se há concordância entre os níveis de exigência

identificados e avaliados pelos participantes a partir da aplicação do instrumento. O Coeficiente de Concordância W de Kendall quantifica o grau de concordância entre três avaliadores ou mais em relação à sua classificação do mesmo grupo de sujeitos. Seu valor varia de 0 a 1, onde 0 representa ausência total de concordância e 1 representa concordância perfeita. A Tabela 2 sugere interpretação para o coeficiente de concordância W de Kendall.

Tabela 2. Interpretação do Coeficiente de Concordância (W) de Kendall. Adaptado de [Landis and Koch 1977].

W de Kendall	Interpretação
0	Não há concordância
0,01 – 0,20	Fraca concordância
0,21 – 0,40	Ligeiramente concordantes
0,41 – 0,60	Moderadamente concordantes
0,61 – 0,80	Concordância substancial
0,81 – 0,99	Concordância quase perfeita
1	Concordância perfeita

A Tabela 3 lista os dados da distribuição de frequência e classificação de abundâncias dos níveis de exigências avaliados pelos participantes para o primeiro edital selecionado (Edital ANVISA 26/2014).

Tabela 3. Tabela superior: distribuição de frequência dos níveis de exigência avaliados para o “Edital ANVISA”. Tabela inferior: dados transformados em classificação de abundâncias dos níveis de exigência.

	Distribuição de frequência dos níveis de exigência				
	1 (muito comum)	2 (comum)	3 (incomum)	4 (raro)	5 (muito raro)
Participante 1	15	14	7	0	0
Participante 2	9	29	6	1	0
Participante 3	5	13	3	3	2
Participante 4	23	14	13	12	9
Participante 5	12	5	9	4	2
Participante 6	5	17	14	9	1
Participante 7	57	53	7	1	0
Participante 8	9	31	6	3	1
Participante 9	22	35	52	8	1
Participante 10	9	27	3	9	1
Participante 11	24	12	24	8	4

	Classificação das abundâncias dos níveis de exigência				
	1 (muito comum)	2 (comum)	3 (incomum)	4 (raro)	5 (muito raro)
Participante 1	5	4	3	1,5	1,5
Participante 2	4	5	3	2	1
Participante 3	4	5	2,5	2,5	1
Participante 4	5	4	3	2	1
Participante 5	5	3	4	2	1
Participante 6	2	5	4	3	1
Participante 7	5	4	3	2	1
Participante 8	4	5	3	2	1
Participante 9	3	4	5	2	1
Participante 10	3,5	5	2	3,5	1
Participante 11	4,5	3	4,5	2	1

Os dados da tabela superior (ver Tabela 3) foram transformados em uma classificação da ordem de abundâncias (tabela inferior), ou seja, os dados foram ordenados pela quantidade de exigências identificadas em cada nível de exigência, onde “1” indica o nível com a menor quantidade de exigências, e “5” o nível com a maior quantidade de exigências. Observa-se também que é atribuída a média dos postos quando ocorrem observações empatadas.

O gráfico da Figura 6, plotado com os dados das classificações de abundâncias (Tabela 3, tabela inferior), ilustra que pode haver convergência nos resultados das avaliações realizadas no edital pelos participantes. Porém, para verificar essa convergência, surge a necessidade de aplicar teste estatístico apropriado.

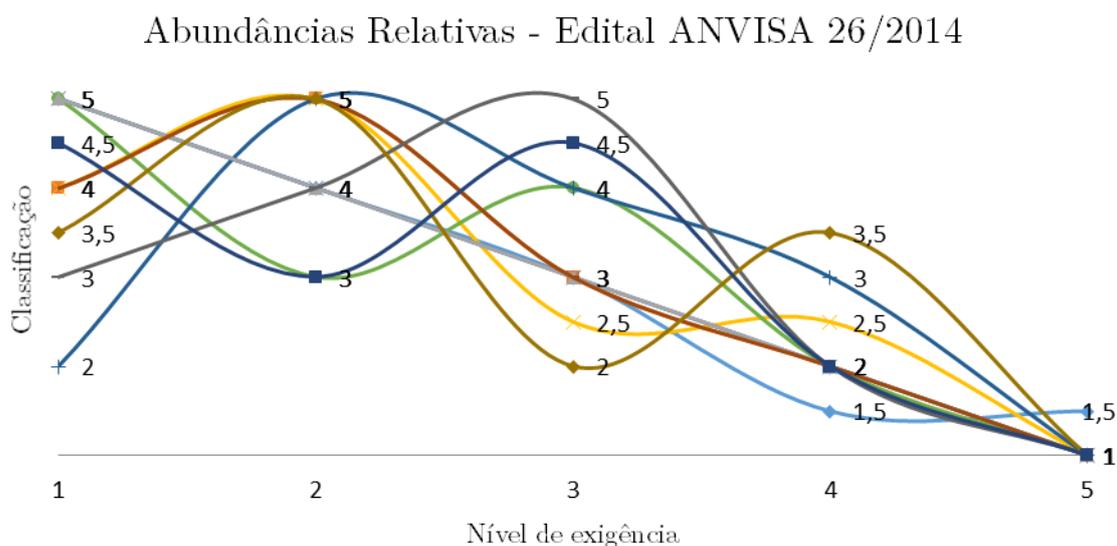


Figura 6. Gráfico de curvas das abundâncias relativas da classificação dos níveis de exigência identificados e avaliados pelos participantes.

A análise estatística foi realizada utilizando o pacote estatístico Minitab 17, no nível de significância $\alpha = 0,05$. A seguir são apresentados os resultados do teste estatístico W.

Tabela 4. Resultados do teste estatístico W

Teste da estatística W – edital 1					
Objetos =	5				
Avaliadores =	11				
W de Kendall =	0,75				
Qui-quadrado =	32,98	Graus de liberdade (GL) =	4	Valor p =	0,00 Rejeita H_0

Com os resultados apresentados na Tabela 4, o grau de concordância entre os 11 avaliadores é expresso por $W = 0,75$. Portanto, o W de Kendall para este estudo apresenta uma *concordância substancial* entre os participantes na avaliação do nível de exigência do edital.

Para determinar a significância estatística dessa concordância, a quantidade é aproximadamente distribuída como um qui-quadrado com $N - 1$ graus de liberdade

[Siegel and Jr. 2006]. Conforme a Tabela 4, obteve-se que com um qui-quadrado = 32,98, com grau de liberdade = 4, tem probabilidade de ocorrência sob H_0 de $p < 0,05$. Logo, pode-se concluir que a concordância entre os 11 participantes é mais alta do que seria se fosse devido ao acaso, isto é, se suas avaliações tivessem sido aleatórias ou independentes. A probabilidade muito baixa sob H_0 associada com o valor observado de W , *permite rejeitar a hipótese nula* de que não há concordância entre os níveis de exigência indicados pelos participantes a partir da aplicação do instrumento, e obter indícios de que há uma *concordância substancial* entre os participantes no que se refere ao nível de exigência do primeiro edital avaliado.

5.10. Resultados

Após análise e interpretação dos resultados, pode-se responder à questão: *O instrumento é eficaz na caracterização dos níveis de exigência dos editais?*

Diante dos resultados, sumarizados na Tabela 5, obteve-se indícios de que há uma *concordância substancial* entre os participantes no que se refere aos níveis de exigência avaliados pelos participantes nos dois editais. Portanto, respondendo à Questão 1, o instrumento pode ser considerado eficaz quanto a capacidade de caracterização dos níveis de exigência dos editais.

Tabela 5. Resumo dos testes de hipóteses

Edital	Hipótese nula	Resultado (W de Kendall)	Valor p	Rejeita H_0 ?
1	$W = 0$	0,75	0,00	Sim
2	$W = 0$	0,70	0,00	Sim

A Figura 7 a seguir ilustra a utilização do instrumento, permitindo que ele sirva de mecanismo comum para comparação de editais de licitação de software.

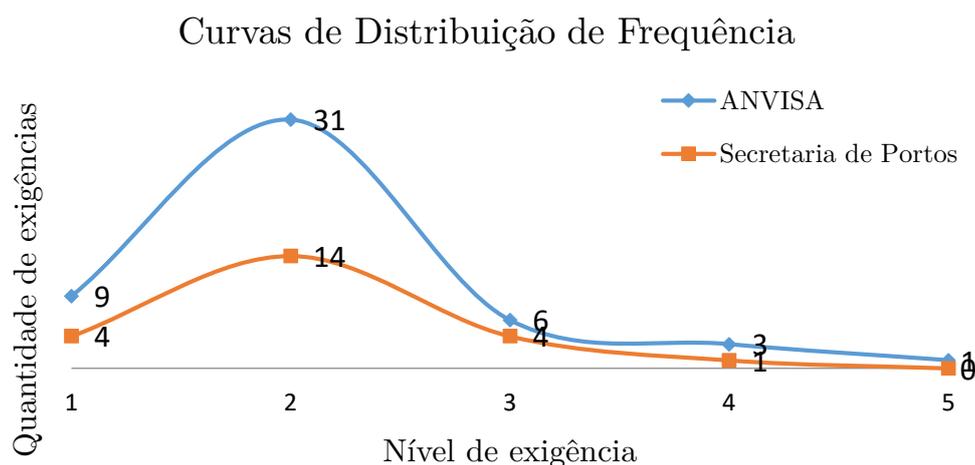


Figura 7. Uso do instrumento para comparar editais de licitação

Conforme a Figura 7, observa-se a comparação dos níveis de exigência de dois editais avaliados por um participante utilizando o instrumento. A partir da caracterização dos aspectos não-funcionais, pode-se classificar o “Edital ANVISA” como sendo mais exigente do que o “Edital Secretaria de Portos”.

5.11. Ameaças à Validade

A presença do pesquisador pode ser uma ameaça à validade. Dúvidas pontuais esclarecidas aos participantes relacionadas às orientações para avaliação dos editais podem ter influência sobre os resultados.

A avaliação a partir de apenas algumas seções dos editais também pode ser uma ameaça à validade. Assumiu-se que o tamanho elevado dos editais (quantidade de páginas) poderia demandar muito esforço para os participantes no momento da avaliação. Logo, resultados podem ser diferentes caso o instrumento seja aplicado em todo o termo de referência de um edital.

O tamanho pequeno da amostra, e a distribuição geográfica dos participantes, não são ideais do ponto de vista estatístico. Embora os resultados deste estudo não possam ser generalizados, eles apoiam a indicação de que o instrumento é eficaz na caracterização do nível de exigência dos editais.

6. Conclusões e Trabalhos Futuros

As formas de compor o preço de referência de contratações de software não consideram, de forma sistemática e repetível, os aspectos não-funcionais especificados nos termos de referência dos editais. Ao mesmo tempo, a literatura de Engenharia de Software apresenta outras alternativas para quantificar os aspectos não-funcionais e apoiar a estimativa do custo do software, porém essas também apresentam limitações, especialmente no contexto de contratações de serviços de software para o governo.

Dada a extensão do problema, foi descrita uma proposta baseada em heurísticas e um procedimento para caracterizar os aspectos não-funcionais de projetos de software no contexto de editais de licitação no setor público brasileiro. As heurísticas propostas foram montadas com a premissa de que devem servir de mecanismo para comparar editais com vistas aos seus aspectos não-funcionais, e que fossem fáceis de serem aplicadas, inclusive por pessoas com pouco conhecimento técnico. Com o propósito de avaliar a eficácia do instrumento estudos baseados em abordagem experimental foram conduzidos.

Considerando as limitações, os estudos possibilitaram observar a evolução do instrumento e obter indícios de que há uma concordância substancial entre os participantes no que se refere aos níveis de exigência avaliados utilizando o instrumento desenvolvido.

Como trabalhos futuros, há a intenção de: analisar se há correlação entre os preços estimados das contratações e os níveis de exigência (caracterizados através deste instrumento) de editais de licitação de software para o setor público; replicar esta experiência com participantes do mesmo público-alvo desta pesquisa, e de outros estados brasileiros, permitindo ampliar o entendimento sobre a eficácia do instrumento.

Referências

Basili, V., Caldiera, G. and Rombach, H. D. (1994). Goal Question Metric Paradigm. *Encyclopedia of Software Engineering*, v. 2, p. 528–534.

Brasil, Ministério do Planejamento, O. e G. S. de L. e T. da I. (2015). *Roteiro de*

Métricas de Software do SISP: Versão 2.1. Brasília: .

BRASIL, M. S. (2014). *Guia de Boas Práticas em Contratação de TI*. v 2 ed. Brasília: .

Buglione, L., Ormandjieva, O. and Daneva, M. (2008). Using PSU for early prediction of COSMIC size of functional and non-functional requirements. *Lecture Notes in Computer Science*, v. 5338 LNCS, p. 352–361.

COSMIC (2015). The COSMIC Functional Size Measurement Method Version 4.0.1 - Measurement Manual. n. April, p. 1–98.

Goodman, L. A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, v. 32, n. 1, p. 148–170.

Green, C., Bradley, D., Ben-Cnaan, T., et al. (2012). Software Non-Functional Assessment Process. *The IFPUG Guide to IT and Software Measurement*. Auerbach Publications. p. 495–523.

IFPUG (2010). *Counting Practices Manual*. 4.3 ed.

IFPUG (2015). *Software Non-functional Assessment Process (SNAP)*. *Assessment Practices Manual*. International Function Point Users Group.

Kendall, M. G. and Smith, B. B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*, v. 10, n. 3, p. 275–287.

Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, v. 33, n. 1, p. 159–174.

Legendre, P. (2005). Species Associations: the Kendall Coefficient of Concordance Revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, v. 10, n. 2, p. 226–245.

Lokan, C. J. (2000). An empirical analysis of function point adjustment factors. *Information and Software Technology*, v. 42, n. 9, p. 649–660.

Mahmoud, A. and Williams, G. (2016). Detecting , classifying , and tracing non-functional software requirements. *Requirements Engineering*, v. 21, n. 3, p. 357–381.

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, v. 13, n. 6, p. 522–525.

Meli, R. (2012). Software Measurement in Procurement Contracts. *The IFPUG Guide to IT and Software Measurement*. Auerbach Publications. p. 561–583.

Shull, F., Mendonça, M. G., Basili, V., et al. (2004). Knowledge-Sharing Issues in Experimental Software Engineering. *Empirical Software Engineering*, v. 9, n. 1, p. 111–137.

Siegel, S. and Jr., N. J. C. (2006). *Estatística não-paramétrica para as ciências do comportamento*. 2ª ed. Porto Alegre: Artmed.

Wohlin, C., Runeson, P., Höst, M., et al. (2000). *Experimentation in Software Engineering: An Introduction*. Boston, MA: Springer US. v. 6