

Métricas de Tamanho de Software com Métodos Ágeis no Setor Público: um Survey

Alexandre F. Mendonça¹, Fernanda M. R. Alencar^{1,2}

¹Programa de Pós-Graduação Acadêmica em Engenharia de Computação
– Escola Politécnica de Pernambuco – Universidade de Pernambuco (UPE)
Rua Benfica, 455 – 50720-001 – Recife – PE – Brazil

²Universidade Federal de Pernambuco (UFPE) – Recife, PE – Brazil

{afm2, fernandaalenc}@ecom.poli.br, fernanda.ralencar@ufpe.br

Abstract. *Public administration is increasingly using agile methods, adapting to the reality of the market. While still required to measure your deliverables through some software size metric, choosing this metric properly can define the success level of public projects. This paper identifies software size metrics in conjunction with agile methods for use in the public sector. A survey is conducted to collect the metrics and method most likely to succeed in the public sector. As a result, the COSMIC metric and the SCRUM, XP and Kanban methods are the most promising for use in the public sector.*

Resumo. *A administração pública tem utilizado cada vez mais métodos ágeis, adequando-se à realidade do mercado. Muito embora continue obrigada a mensurar suas entregas através de alguma métrica de tamanho de software, a escolha adequada dessa métrica pode definir o nível de sucesso dos projetos públicos. Neste artigo, identificam-se as métricas de tamanho de software em conjunto com métodos ágeis a serem utilizados no setor público. Um survey é conduzido para coletar as métricas e método com maior probabilidade de sucesso no setor público. Como resultado, a métrica COSMIC e os métodos SCRUM, XP e Kanban se apresentam como os mais promissores para serem utilizados no setor público.*

1. Introdução

Mais e mais organizações estão adotando o desenvolvimento de software Ágil. Os métodos ágeis (MA) são particularmente úteis quando os requisitos estão mudando continuamente durante a vida do projeto e/ou estão fortemente incompletos na fase de inicialização do projeto.

Inúmeros métodos de dimensionamento de software foram criados, mas se depararam com limitações, tais como: não serem aplicados no início do processo de desenvolvimento de software; não serem aplicados uniformemente durante a vida útil do software; e, não serem entendidos de forma significativa pelos usuários do software que recebem estimativas baseadas nesses métodos de dimensionamento. Assim, os conceitos de métrica de tamanho de software (MTS) são projetados para superar essas limitações, deslocando o foco de medir como o software é implementado, para medir o tamanho em termos das funções requeridas pelo usuário.

A partir do trabalho de [Mendonça and Alencar 2019], se pode entender quais MTS e MA são mais aderentes ao contexto público como resultado de uma Revisão Sistemática da Literatura (RSL). Esse entendimento deve levar a escolhas que devem ser validadas com especialistas do setor público. Como resultado se espera responder quais MTS e MA são mais aceitas num processo de desenvolvimento de software para o setor público.

Segue a estrutura do trabalho: na seção 2, apresenta-se as MTS e os MA resultantes da RSL realizada no trabalho de [Mendonça and Alencar 2019]; na seção 3, apresenta-se todo o processo de elaboração e aplicação do *survey*; na seção 4, apresenta-se os principais resultados do *survey*; por fim, na seção 5, apresenta-se a conclusão e trabalhos futuros.

2. Background

A Revisão Sistemática da Literatura (RSL) resultante do trabalho de [Mendonça and Alencar 2019] indica as MTS e os MA mais relevantes para o contexto público. Nesse trabalho, foram realizadas três RSL parciais envolvendo MTS e MA, MTS e setor público, e MA e setor público. Após análise e consolidação das RSL parciais, se chegou ao resultado final da RSL. Com isso, é possível comparar resultados finais com o setor privado, sem considerar MTS e sem considerar MA. Entre as MTS, Ponto de Função da IFPUG [Fehlmann 2002] apresenta maior percentual de ocorrência no contexto público independente de se considerar MA. A outra MTS identificada é a COSMIC [Committee et al. 2015] que apresenta o mais alto percentual de ocorrência para o setor privado.

Para os MA, a RSL indica o Scrum [Tytkowska et al. 2015] como o MA com maior percentual de ocorrência tanto no contexto público quanto no contexto privado. O outro MA identificado é o eXtreme Programming (XP) [Beck 1999] que está presente em todos os resultados das RSL parciais e consolidada.

3. Survey

Após a conclusão da RSL indicar quais são as MTS e MA mais utilizados no setor público, o foco se volta para o entendimento de suas relevâncias na prática de trabalho do setor público. Capturando as características dessas MTS e MA, é possível apresentar para especialistas num formato de questionário para que seja possível identificar as melhores práticas e combinações de MTS e MA que potencializem o máximo de valor ao negócio da administração pública.

Para atingir essa meta, foi adotado a técnica do *survey*. Um processo para o planejamento e execução do *survey* foi adaptado dos estudos de Kitchenham e Pfleeger [Pfleeger and Kitchenham 2001, Kitchenham and Pfleeger 2002c, Kitchenham and Pfleeger 2002d, Kitchenham and Pfleeger 2002a, Kitchenham and Pfleeger 2002b] e pode ser observado na Figura 1. O primeiro passo está focado na identificação das características dessas MTS e MA. Na sequência, perguntas são formuladas para capturar a aderência da característica avaliada com a realidade de desenvolvimento de software no setor público. Finalmente, especialistas são acionados para colaborarem respondendo o questionário estruturado.

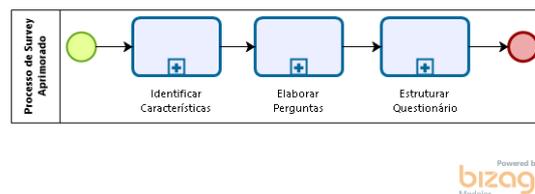


Figura 1. Processo de Survey

3.1. Identificar Características

O intuito desse processo é identificar características de cada MTS e MA relacionada para essa etapa da pesquisa a partir da RSL. As características procuradas devem delimitar o escopo da métrica ou método e diferenciar da outra métrica ou método da pesquisa. Os MA selecionados para esse trabalho são Scrum e XP por serem os métodos ágeis presentes no resultado da RSL consolidada, como visto em [Mendonça and Alencar 2019]. Já as MTS selecionadas são PF e . PF foi selecionado como a MTS mais presente na RSL consolidada. Já a COSMIC foi selecionada por aparecer na RSL consolidada e ser a mais presente no contexto do setor privado. Test Point não foi selecionado por estar presente em uma única referência e não ter também significância no setor privado. Para cada característica, é apresentada uma descrição que explica o escopo da característica associada. Com isso, se consegue facilitar a elaboração de perguntas que revelem o grau de relevância dessa característica para o especialista que está respondendo o questionário.

As características do Scrum elencadas para esse estudo estão focadas prioritariamente nas cerimônias, artefatos e papéis do *framework* do MA Scrum. Ao todo, 12 características do Scrum [Tomás 2009, Ali et al. 2015, Tridibesh 2016, Nunes 2017] foram identificadas, e estão listadas a seguir:

- C.01 Equipes pequenas
- C.02 Requisitos instáveis (incompletos ou desconhecidos)
- C.03 Sprints curtos
- C.04 Histórias dos usuários
- C.05 Priorização das histórias no Product Backlog
- C.06 Estimativas de histórias do usuário
- C.07 Reuniões diárias
- C.08 Entregas no final de cada Sprint
- C.09 Gráfico Burnup/burndown
- C.10 Papel - Product Owner
- C.11 Papel - Scrum Master
- C.12 Papel - Equipe Scrum

As características do XP elencadas para esse estudo estão focadas em suas práticas e valores. Ao todo, 16 características do XP [dos Santos Soares 2004, Nunes 2017] foram identificadas, e estão listadas a seguir:

- C.13 Jogo do Planejamento
- C.14 Entregas frequentes (*Release* curto)
- C.15 Metáfora

- C.16 Arquitetura simples
- C.17 TDD
- C.18 Refatoração
- C.19 Programação em pares
- C.20 Propriedade coletiva
- C.21 Integração contínua
- C.22 40 horas de trabalho semanal (ritmo sustentável)
- C.23 Cliente presente
- C.24 Código padronizado
- C.25 Comunicação
- C.26 Simplicidade
- C.27 *Feedback*
- C.28 Coragem

As características do PF elencadas para esse estudo estão focadas em elementos de adoção da MTS e técnicas de contagem. Ao todo, 9 características do PF [Karner 1993, Symons 2001, Bradley et al. 2010, BARBALHO 2014] foram identificadas, e estão listadas a seguir:

- C.29 Automatização da contagem
- C.30 Ponto de vista da medição
- C.31 Processo de contagem
- C.32 Forma de manutenção
- C.33 Abrangência
- C.34 *Benchmarking*
- C.35 Medição das transações
- C.36 Projetos de melhoria
- C.37 Aplicabilidade

As características da COSMIC elencadas para esse estudo estão focadas em elementos de adoção da MTS e técnicas de contagem. Ao todo, 12 características da COSMIC [Cuadrado-Gallego et al. 2010, Berardi and Santillo 2010, Desharnais et al. 2011, Berardi et al. 2011, BARBALHO 2014] foram identificadas, e estão listadas a seguir:

- C.38 Granularidade da medição
- C.39 Medição sem limitação de faixas de complexidade
- C.40 Considerar, dentro do possível, requisitos não funcionais
- C.41 Ponto de vista da medição
- C.42 Medição por camadas
- C.43 Medição sem modelo de dados
- C.44 Medição por movimentações de dados
- C.45 Projetos de melhorias
- C.46 Aplicabilidade
- C.47 Abrangência
- C.48 Checagem da qualidade dos requisitos
- C.49 Função pronta quando requisito pronto

3.2. Elaborar Perguntas

A partir das características identificadas com seu escopo definido, o próximo passo está focado na elaboração das perguntas. Elas devem atestar que o escopo das características sob avaliação está aderente a prática de trabalho no setor público por parte dos especialistas. Cada pergunta está associada a uma única característica. Uma característica está sendo coberta por uma ou mais perguntas.

A meta desse *survey* é avaliar o nível de concordância das características das MTS e MA elencadas pela RSL no âmbito profissional de especialistas que trabalham com MTS e MA no setor público. O contexto das perguntas do *survey* foram focadas no tempo presente. Com isso, trabalhamos na identificação da aderência de cada característica avaliada num contexto que absorve toda a bagagem de experiência atualizada do especialista.

Especialistas com as características adequadas se encontram pulverizados pelo Brasil. Portanto, o tipo de *survey* adotado foi não supervisionado. Ou seja, um questionário online foi disponibilizado para que os especialistas respondessem sem a oportunidade de explicar com maiores detalhes o que exatamente se espera de cada uma das perguntas. Para mitigar o risco de se ter respostas não direcionadas, um processo de revisão foi realizado com um dos especialistas.

Essa revisão teve como objetivo reduzir o viés das perguntas por parte dos autores, ajustar a sua complexidade ao perfil dos especialistas e direcionar cada pergunta para o seu objetivo particular. Com a conclusão dessa revisão, chegamos ao montante de 62% das perguntas aprimoradas. Essas perguntas já revisadas serão apresentadas a seguir.

Para cobrir o escopo das 12 características do Scrum foram elaboradas 16 perguntas, como visto na Tabela 1. Com exceção das características C_01, C_03, C_08 e C_10 que precisaram de duas perguntas cada, todas as demais estão integralmente cobertas com apenas uma pergunta.

Tabela 1. Questionário - Scrum

ID	Pergunta ID	Pergunta
C_01	P_01	Você concorda que projetos de grande porte sejam executados por equipes pequenas (de até 9 pessoas)?
	P_02	Você concorda que projetos de grande porte sejam segmentados para que possam ser executados por equipes pequenas (de até 9 pessoas)?
C_02	P_03	Você concorda em iniciar a codificação de um projeto com um escopo de requisitos incompletos ou desconhecidos?
C_03	P_04	Você concorda em realizar entregas em períodos curtos?
	P_05	Você concorda em realizar entregas em períodos fixos de tempo?
C_04	P_06	Você concorda em trabalhar com requisitos escritos de forma não detalhada?
C_05	P_07	Você concorda que os requisitos sejam priorizados pelo cliente?

Continua na próxima página

Tabela 1 – *Continua da página anterior*

ID	Pergunta ID	Pergunta
C_06	P_08	Você concorda que os requisitos sejam estimados pela equipe de projeto?
C_07	P_09	Você concorda que encontros diários curtos ajudem a agregar valor ao projeto?
C_08	P_10	Você concorda que entregas constantes em curto espaço de tempo agregam valor ao andamento do projeto?
	P_11	Você concorda que entregas constantes em curto espaço de tempo agregam valor ao cliente?
C_09	P_12	Você concorda que acompanhamento em gráficos do projeto ajudam a manter o ritmo da equipe?
C_10	P_13	Você concorda que o cliente (ou alguém alocado por ele) deve ser o responsável por escrever os requisitos?
	P_14	Você concorda que o cliente (ou alguém alocado por ele) deve estar disponível diariamente para responder a equipe do projeto?
C_11	P_15	Você concorda que um integrante da equipe seja destacado das atividades técnicas para evitar que problemas de natureza não técnica interrompam atividades do projeto?
C_12	P_16	Você concorda que a equipe do projeto tenha autonomia para escolher qual tarefa executar durante o projeto?

Fonte: o autor

Para cobrir o escopo das 16 características do XP foram elaboradas 17 perguntas, como visto na Tabela 2. Com exceção da característica C_20 que precisou de duas perguntas, todas as demais estão integralmente cobertas com apenas uma pergunta. Outro destaque é a característica C_23 que compartilha da mesma pergunta P_14 já construída para cobrir uma característica do Scrum (C_10).

Tabela 2. Questionário - XP

ID	Pergunta ID	Pergunta
C_13	P_17	Você concorda em ter reuniões no início de cada <i>release</i> para priorizar e estimar o <i>release</i> junto com seu cliente?
C_14	P_18	Você concorda que seu projeto trabalhe com <i>release</i> curtos para que sejam disponibilizados o quanto antes em produção para uso do cliente?
C_15	P_19	Você concorda que os integrantes do projeto devem procurar utilizar uma linguagem própria de fácil entendimento entre si?
C_16	P_20	Você concorda em gerar uma arquitetura o mais simples possível para atender aos requisitos atuais do projeto?

Continua na próxima página

Tabela 2 – *Continua da página anterior*

ID	Pergunta ID	Pergunta
C_17	P_21	Você concorda em codificar um código de teste antes de codificar a própria funcionalidade a ser testada?
C_18	P_22	Você concorda em alterar o código de uma funcionalidade estável para tornar o software mais simples de ser mantido e de ser evoluído?
C_19	P_23	Você concorda que na sua equipe dois desenvolvedores trabalhem juntos na mesma máquina para implementar o mesmo requisito?
C_20	P_24 P_25	Você concorda que os desenvolvedores tenham acesso a todo o código do projeto? Você concorda que os desenvolvedores possam alterar no seu projeto o que julgarem importante sem solicitar autorização prévia?
C_21	P_26	Você concorda que toda a aplicação seja testada sempre que uma nova funcionalidade for implementada?
C_22	P_27	Você concorda que horas extras constantes no projeto podem afetar a produtividade da equipe?
C_23	P_14	Você concorda que o cliente (ou alguém alocado por ele) deve estar disponível diariamente para responder a equipe do projeto?
C_24	P_28	Você concorda que toda a equipe deve trabalhar regida por um padrão de codificação (seja qual for) definido para o projeto?
C_25	P_29	Você concorda que o cliente tenha uma comunicação direta com a equipe do projeto para tratar de assuntos técnicos do projeto?
C_26	P_30	Você concorda que a implementação deve ser a mais simples possível para atender aos requisitos atuais do projeto sem se preocupar com possíveis necessidades futuras do projeto?
C_27	P_31	Você concorda que o cliente pode estar reavaliando (excluindo, alterando ou adicionando) os requisitos do projeto a cada entrega realizada?
C_28	P_32	Você concorda que a equipe deve ter autonomia para diretamente tomar ações no projeto tais como, mas não restritas a, apontar um problema no projeto, solicitar ajuda, comunicar atrasos ao cliente, etc.?

Fonte: o autor

Para cobrir o escopo das 9 características do FP foram elaboradas 9 perguntas, como visto na Tabela 3. Todas as características do FP estão integralmente cobertas com apenas uma pergunta.

Tabela 3. Questionário - Ponto de Função (IFPUG)

ID	Pergunta ID	Pergunta
C_29	P_33	Você concorda que é significativo para uma métrica de tamanho de software ter sua contagem facilmente automatizada?
C_30	P_34	Você concorda que a medição do tamanho do software deve estar limitada ao ponto de vista do usuário?
C_31	P_35	Você concorda que a utilização de níveis de complexidade (baixa, média e alta) agilizam a medição de tamanho de software num nível de precisão aceitável?
C_32	P_36	Você concorda que é significativo para uma métrica de tamanho de software ser mantida por uma organização internacional?
C_33	P_37	Você concorda que é significativo para uma métrica de tamanho de software ser amplamente utilizada pelo mercado público e privado?
C_34	P_38	Você concorda que é significativo para uma métrica de tamanho de software ter um grande acervo que permite estudos e comparações?
C_35	P_39	Você concorda que as medições de tamanho de software podem ser definidas apenas pela quantidade de tipos de dados e arquivos referenciados?
C_36	P_40	Você concorda que para contagens de mudanças, o tamanho total da funcionalidade alterada seja considerado independente do nível de complexidade dessa mudança?
C_37	P_41	Você concorda que é significativo para uma métrica de tamanho de software seja aplicável apenas para software aplicativo de negócio?

Fonte: o autor

Para cobrir o escopo das 12 características da COSMIC foram elaboradas 12 perguntas, como visto na Tabela 4. Todas as características da COSMIC estão integralmente cobertas com apenas uma pergunta.

Tabela 4. Questionário - COSMIC

ID	Pergunta ID	Pergunta
C_38	P_42	Você concorda que uma métrica de tamanho de software que permitisse uma medição numa granularidade mais fina seria mais benéfica para o projeto?
C_39	P_43	Você concorda que uma métrica de tamanho de software com medição focada em complexidade seria mais benéfica para o projeto?
C_40	P_44	Você concorda que uma métrica de tamanho de software que considera requisitos não funcionais em sua técnica de medição é benéfica para o projeto?

Continua na próxima página

Tabela 4 – *Continua da página anterior*

ID	Pergunta ID	Pergunta
C_41	P_45	Você concorda que a medição do tamanho do software poderia considerar, além do ponto de vista do usuário, também o ponto de vista da arquitetura do software?
C_42	P_46	Você concorda que uma métrica de tamanho de software que considera a quantidade de camadas de aplicação em seu processo de medição possa ser precisa?
C_43	P_47	Você concorda que uma métrica de tamanho de software que não considere o modelo de dados do projeto em seu processo de medição possa ser mais precisa?
C_44	P_48	Você concorda que as medições de tamanho de software podem ser definidas apenas pela quantidade de movimentações dos dados na transação?
C_45	P_49	Você concorda que para contagens de mudanças, apenas a quantidade de movimentações alteradas da funcionalidade seja considerado independente do nível de complexidade dessa mudança?
C_46	P_50	Você concorda que é significativo para uma métrica de tamanho de software seja aplicável para software de tempo-real?
C_47	P_51	Você concorda em utilizar uma métrica de tamanho de software reconhecida internacionalmente em seu projeto, mas que não tenha registro de uso no Brasil?
C_48	P_52	Você concorda em utilizar uma métrica de tamanho de software para ajudar na verificação da qualidade dos requisitos do seu projeto?
C_49	P_53	Você concorda que uma métrica de tamanho de software que separe funções transacionais de funções de dados tenha dificuldade para medir requisitos (histórias de usuários) concluídos?

Fonte: o autor

3.3. Estruturar Questionário

Após a elaboração das perguntas, elas foram agrupadas em cinco etapas: identificação, Scrum, XP, PF e COSMIC. A primeira etapa está focada na identificação das MST e MA utilizadas profissionalmente pelo especialista. Alguns dados de identificação como e-mail e empresa pública base para aplicação desse questionário também pode ser visualizados na Tabela 5.

Na sequência, estão agrupados os questionários de Scrum (Tabela 1), XP (Tabela 2), FP (Tabela 3) e COSMIC (Tabela 4), nessa ordem. Nessas etapas, todas as perguntas são objetivas e fechadas, exceto a pergunta final de cada grupo de etapas de MA e MTS na qual o especialista fica livre para mencionar algo relevante que não foi coberto pelas demais perguntas. Cada pergunta é formulada com quatro níveis de resposta para evitar respostas neutras. Uma quinta opção é disponibilizada para o caso do especialista não ter conhecimento da característica avaliada.

Tabela 5. Questionário – Dados de Identificação

Pergunta	Tipo de Questão
Endereço de e-mail	Resposta curta
Qual empresa pública de sua experiência que você está tomando como base para responder a esse questionário?	Resposta curta
Qual(is) método(s) ágil(eis) você vivenciou/vivencia nessa empresa pública?	Caixas de seleção com as opções: KANBAN, LEAN, Scrum, XP e Outro.
Qual(is) métrica(s) de tamanho de software você vivenciou/vivencia nessa empresa pública?	Caixas de seleção com as opções: Ponto de Caso de Uso, Ponto de Função (COSMIC), Ponto de Função (IFPUG) e Outro.

Fonte: o autor

O questionário ficou disponível para ser respondido do dia dois até vinte e três de abril de dois mil e dezenove. Ele foi enviado para 78 especialistas que trabalham ou trabalharam para o setor público, que utiliza ou utilizaram métodos ágeis em seus projetos que tiveram seus tamanhos funcionais de software contados através de alguma MTS. A partir das respostas desses especialistas, seguem os resultados obtidos na próxima seção.

4. Resultados e discussão

Durante os 22 dias que o *survey* ficou disponível na Internet, foram obtidas 42 respostas. Levando em consideração que foram enviadas 78 solicitações de resposta para especialistas, foi obtida uma taxa de resposta de 53,8%. A partir das respostas, algumas análises foram realizadas com intuito de apresentar credibilidade ao resultado final e a busca do entendimento da aderência de cada característica ao contexto dos projetos de software do setor público. Dentre as análises realizadas se pode destacar: MTS e MA presentes no setor público, e nível de concordância para cada MTS e MA.

4.1. MTS e MA Presentes no Setor Público

As respostas dos especialistas através do *survey* indicaram quais são as MTS e os MA presentes no setor público. Dentre as MTS, a mais citada pelos especialistas é IFPUG com 67% das ocorrências. Todas as outras MTS têm uma utilização bem mais baixa no setor público, conforme pode ser observado na Tabela 6. Algumas respostas não são precisamente uma MTS, mas estão listadas por terem sido respondidas pelos especialistas. Nesse contexto, se encontram UST e Story Point.

Do ponto de vista dos MA, Scrum é o mais utilizado com 55% das ocorrências. Em segundo lugar, com percentual significativo de 28% de ocorrência, se identificou o uso do Kanban. Mais detalhes estão presentes na Tabela 6.

4.2. Origem dos Especialistas

A origem dos especialistas se apresenta relevante para se tornar possível a identificação de algum tipo de concentração que possa direcionar o resultado da pesquisa. Um aspecto importante é o percentual de especialistas terceirizados que responderam o *survey*. Esse percentual em 57% com 24 especialistas.

Tabela 6. MTS e MA Presentes no Survey

(a) MTS			(b) MA		
MTS	Quantidade	Percentual	MA	Quantidade	Percentual
IFPUG	36	67%	Scrum	42	55%
PCU	5	9%	Kanban	21	28%
COSMIC	4	7%	Lean	9	12%
UST	4	7%	XP	4	5%
PEEX	2	4%			
Story Point	1	2%			
NESMA	1	2%			
SNAP	1	2%			

Fonte: o autor

A origem da empresa pública na qual os especialistas se basearam para responder as perguntas do questionário também foi capturada por essa pesquisa. Um total de nove empresas públicas distintas foram identificadas nas respostas dos especialistas. Aproximadamente um terço dos especialistas tiveram suas experiências no setor público federal e os demais dois terços no setor público estadual, como pode ser observado na Tabela 7.

Com relação ao estado federativo de origem das empresas públicas no qual os especialistas tiveram suas experiências, há uma predominância do estado de Pernambuco com 68% de participação. Ainda se tem participação do Distrito Federal (20%) e Rio de Janeiro (12%). Não foram coletadas respostas de especialistas que trabalham em empresas públicas nos demais estados federativos do Brasil. Dados podem ser observados na Tabela 7.

Tabela 7. Origem da Empresa Pública

(a) Tipo		(b) Estado	
Origem	%	Estado Brasileiro	%
Federal	36%	PE	68%
Estadual	64%	DF	20%
		RJ	12%

Fonte: o autor

4.3. Nível de Concordância

O nível de concordância das respostas dos especialistas ajuda a entender quais características ou elementos são mais aderentes a se trabalhar no contexto do setor público. Para identificar esse nível de concordância foi preciso quantificar a escala de respostas do *survey*, conforme pode ser observado na Tabela 8.

Resposta "Concordo totalmente" tem seu valor associado ao número 4, resposta "Concordo parcialmente" ao número 3, resposta "Discordo parcialmente" ao número 2 e resposta "Discordo totalmente" ao número 1. Ainda de acordo com essa escala, pergunta ou característica que obtenha média de respostas igual a 4, ou seja, todas as respostas "Concordo totalmente", alcançaria, por associação, 100% de nível de concordância por

Tabela 8. Quantificação da Escala de Respostas

Resposta	Pontuação	Percentual de Concordância
Concordo totalmente	4	100%
Concordo parcialmente	3	66%
Discordo parcialmente	2	33%
Discordo totalmente	1	0%

Fonte: o autor

parte dos especialistas; média de respostas igual a 3, 66% de nível de concordância; média de respostas igual a 2, 33% de nível de concordância; média de respostas igual a 1, ou seja, todas as respostas "Discordo totalmente", 0% de nível de concordância.

Fazendo essa associação, é possível mensurar o nível de concordância para cada elemento do *survey*, conforme pode ser visto na Tabela 9. Todas as respostas para todas as perguntas do Scrum obtiveram uma média de 3,45 pontos com nível de concordância calculado de 81,7%. As respostas do XP obtiveram 3,37 pontos com nível de concordância de 79%. As respostas do PF obtiveram 2,83 pontos com nível de concordância de 60,8%. As respostas da COSMIC obtiveram 2,87 pontos com nível de concordância de 62,4%. Os MA ficaram com níveis de concordância próximos. Uma pequena vantagem para o Scrum foi identificada com 2,7 pontos percentuais a mais de nível de concordância do que o XP. As MTS ficaram com níveis de concordância semelhantes entre si, porém mais baixos do que os MA. Uma pequena vantagem para a COSMIC com 1,6 ponto percentual a mais do que o PF de nível de concordância.

Tabela 9. Nível de Concordância

MA / MTS	Pontuação Média	% Concordância
Scrum	3,45	81,7%
XP	3,37	79,0%
PF	2,83	60,8%
COSMIC	2,87	62,4%

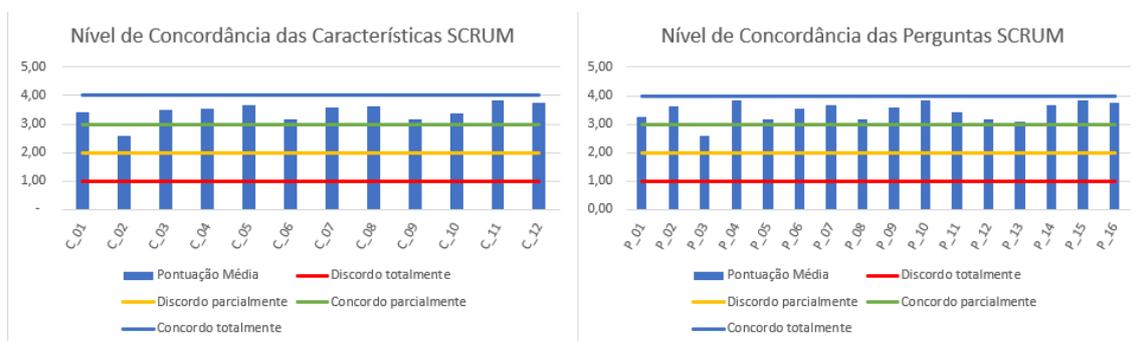
Fonte: o autor

4.3.1. Scrum

Continuando a identificação dos resultados de níveis de concordância, uma estratificação foi realizada para visualizar os resultados por características e perguntas. Inicialmente, é observado os resultados de níveis de concordância das características e perguntas do MA Scrum, como pode ser observado na Figura 2. Como destaque, a característica de maior nível de concordância é C_11 (Papel - Scrum Master), seguida da C_12 (Papel - Equipe Scrum). Com relação às perguntas, o maior nível de concordância se encontra nas perguntas P_04 e P_15. Com o menor nível de concordância, se tem a característica C_02 (Requisitos instáveis (incompletos ou desconhecidos)), seguida da C_06 (Estimativas de histórias do usuário). Por sinal, a característica C_02 foi a única que ficou entre as faixas de "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da escala "Concordo parcialmente". Com relação às perguntas, o menor nível de concordância se encontra nas perguntas P_03 e P_13. Somente a pergunta P_03 ficou entre

as faixas de "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da escala "Concordo parcialmente".

Figura 2. Nível de Concordância das Características Scrum

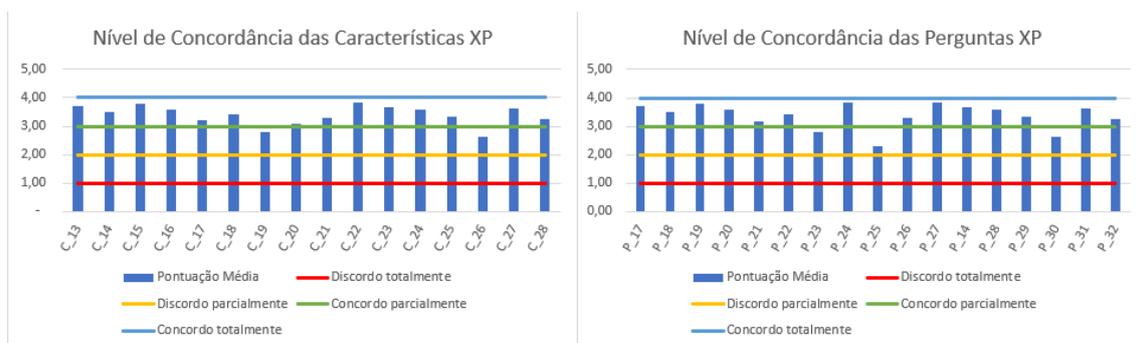


Fonte: o autor

4.3.2. XP

XP é outro MA observado a respeito aos resultados de níveis de concordância de suas características e perguntas, como pode ser observado na Figura 3. Como destaque, a característica de maior nível de concordância é C_22 (40 horas de trabalho semanal (ritmo sustentável)), seguida da C_15 (Metáfora). Com relação às perguntas, o maior nível de concordância se encontra nas perguntas P_24 e P_27. Com o menor nível de concordância, se tem a característica C_26 (Simplicidade), seguida da C_19 (Programação em pares). Por sinal, essas duas características são as únicas que ficaram entre as faixas de "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da escala "Concordo parcialmente". Com relação às perguntas, o menor nível de concordância se encontra nas perguntas P_25 e P_30. Além dessas perguntas, a P_23 fecha o conjunto de perguntas que ficaram entre as faixas de "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da escala "Concordo parcialmente".

Figura 3. Nível de Concordância das Características XP

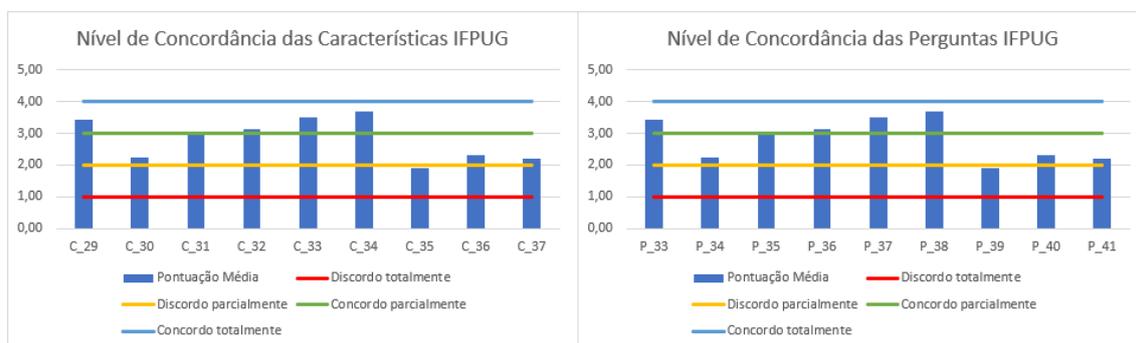


Fonte: o autor

4.3.3. IFPUG

Se voltando para as MTS, IFPUG também é observada com relação aos resultados de níveis de concordância de suas características e perguntas, como pode ser observado na Figura 4. Como destaque, a característica de maior nível de concordância é C_34 (Benchmarking), seguida da C_33 (Abrangência). Com relação às perguntas, o maior nível de concordância se encontra nas perguntas P_38 e P_36. Com o menor nível de concordância, temos a característica C_35 (Medição das transações), seguida da C_37 (Aplicabilidade). Por sinal, a característica C_35 é a única que ficou entre as faixas de escala "Discordo totalmente" e "Discordo parcialmente". Mais três características (C_30, C_36 e C_37) se encontram entre as faixas escala "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da faixa de escala "Concordo parcialmente". Com relação às perguntas, o menor nível de concordância se encontra nas perguntas P_39, seguida da pergunta P_41. De forma análoga às características, a pergunta P_39 é a única que ficou entre as faixas de escala "Discordo totalmente" e "Discordo parcialmente". Mais três perguntas (P_34, P_40 e P_41) se encontram entre as faixas escala "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da faixa de escala "Concordo parcialmente".

Figura 4. Nível de Concordância das Características PF



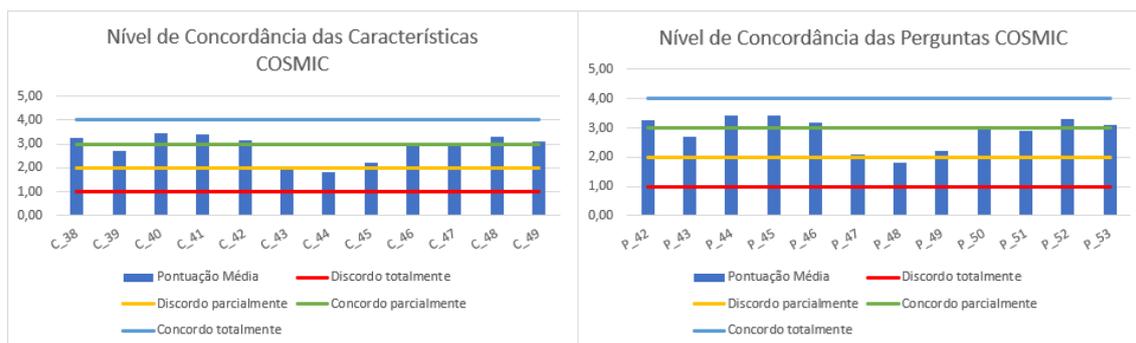
Fonte: o autor

4.3.4. COSMIC

Fechando com a MTS COSMIC, são observados os resultados de níveis de concordância de suas características e perguntas, como pode ser observado na Figura 5. Como destaque, a característica de maior nível de concordância é C_40 (Considerar, dentro do possível, requisitos não funcionais), seguida da C_41 (Ponto de vista da medição). Com relação às perguntas, o maior nível de concordância se encontra nas perguntas P_44 e P_45. Com o menor nível de concordância, temos a característica C_44 (Medição por movimentações de dados), seguida da C_43 (Medição sem modelo de dados). Por sinal, a característica C_44 é a única que ficou entre as faixas de escala "Discordo totalmente" e "Discordo parcialmente". Mais quatro características (C_39, C_43, C_45 e C_47) se encontram entre as faixas escala "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da faixa de escala "Concordo parcialmente". Com relação às perguntas, o menor nível de concordância se encontra nas perguntas P_48, seguida da pergunta P_47.

De forma análoga às características, a pergunta P_48 é a única que ficou entre as faixas de escala "Discordo totalmente" e "Discordo parcialmente". Mais quatro perguntas (P_43, P_47, P_49 e P_51) se encontram entre as faixas escala "Discordo parcialmente" e "Concordo parcialmente". Todas as demais estão acima da faixa de escala "Concordo parcialmente".

Figura 5. Nível de Concordância das Características COSMIC



Fonte: o autor

4.4. Discussões

Como observado nas seções anteriores, a MTS IFPUG é largamente utilizada no setor público. Por ser tão aplicada, os resultados com baixo nível de concordância se demonstram com maior credibilidade. A utilização da IFPUG faz com que os especialistas conheçam suas características e possam ter maior embasamento para poder identificar o nível de concordância de cada uma. Já a MTS COSMIC não é tão aplicada, de acordo com essa pesquisa. As respostas sobre suas características são respondidas mais por associação às características da IFPUG do que por conhecimento prático de uso da COSMIC.

Observando as características da IFPUG com nível de concordância abaixo de 50%, e comparando com características semelhantes da COSMIC, é possível identificar que essas características da COSMIC, apresentaram níveis de concordância maiores, exceto a característica de projetos de melhoria. A projeção de utilização da COSMIC se apresenta com resultados mais satisfatórios do que a prática de uso da IFPUG.

Numa outra análise, confrontando as características da IFPUG e da COSMIC com mais de 75% de concordância, é possível entender seus pontos fortes. Para IFPUG, sua vasta utilização do setor público faz com que gere uma massa de dados entre os mais diversos órgãos públicos que facilita o processo de *benchmarking*. Outra consequência de seu uso, é a existência de um grande acervo de informações sobre IFPUG em diversas organizações. Por último, a característica de automatização da contagem, apesar da literatura indicar dificuldades devido a elementos valorados por seres humanos, os especialistas entenderam que planilhas configuradas que facilitam o processo de contagem já indicam um nível de amadurecimento no processo de automatização de contagens.

Para COSMIC, a capacidade de medir pequenos pedaços de funcionalidades facilita o processo de contagem de projetos em andamento ou inacabados. A possibilidade de considerar no processo de contagem aspectos não funcionais apresenta uma maior aderência a complexidade e esforço de desenvolvimento do projeto alvo de contagem.

Um outro ponto significativo é considerar no processo de contagem outros pontos de vista, como o ponto de vista dos desenvolvedores, ao invés de utilizar unicamente o ponto de vista dos usuários. Por último, o processo de contagem COSMIC se torna mais eficiente caso a documentação esteja parametrizada e completa. Portanto, a utilização da COSMIC provê uma verificação de qualidade dos artefatos de requisitos.

Se voltando para os MA, nenhuma característica do Scrum ou XP apresenta percentual menor do que 50% de nível de concordância. Com relação às características com percentual maior do que 75% de concordância, 75% das características Scrum analisadas nessa pesquisa e o mesmo percentual das características XP estão nesse nível de concordância. Isso demonstra que os especialistas concordam que o nível de maturidade dos MA está adequado em sua utilização no setor público.

5. Conclusões e Trabalhos Futuros

O *survey* gerou como resultado final a identificação que as características do Scrum e XP são bem aceitas pelos especialistas do setor público, com níveis superiores a 75% de concordância. Além disso, evidenciou o uso do Kanban de forma significativa por parte dos especialistas. Apesar do nível elevado de aceitação das características dos MA, quatro delas merecem uma atenção especial por terem um nível menor de concordância (C_02: Requisitos instáveis (incompletos ou desconhecidos), C_19: Programação em pares, C_20: Propriedade coletiva e C_26: Simplicidade).

Com relação a MTS, ficou evidente que a COSMIC, apesar de ser pouco conhecido e utilizado pelo setor público, tem um nível de aceitação maior de suas características em comparação com a IFPUG. Enquanto a COSMIC apresenta um percentual de 58% de suas características com nível de concordância, a IFPUG apresenta 56% de suas características com nível de concordância. Em contrapartida, COSMIC apresenta um percentual de 8% de suas características com nível de discordância enquanto a IFPUG apresenta um percentual de 11% de suas características com nível de discordância. Apoiado, também, por resultados da RSL, COSMIC é a MTS escolhida por essa pesquisa.

Fechando essa análise através do *survey*, é importante ressaltar que algumas limitações foram identificadas para essa pesquisa. As tomadas de decisões desse trabalho foram tomadas com base nas respostas de um *survey* que continha respostas de apenas três estados brasileiros. Outras realidades não foram evidenciadas que poderiam eventualmente auxiliar com direcionamentos mais assertivos. Ainda sobre o *survey*, o entendimento do contexto foi elaborado com base num número razoável de respostas. Ao todo foram recebidas 42 respostas. Esse número apresenta uma amostragem relativamente pequena que eventualmente pode pertencer a um nicho específico de especialistas.

Como trabalhos futuros é recomendada a estruturação de processo de desenvolvimento do setor público que utilize COSMIC como MTS e Scrum, XP e Kanban como MA. Após essa estruturação, é importante validar a solução em projetos públicos reais. Como as MA são de alta aceitação pelos especialistas do setor público, o foco da validação deve se concentrar na utilização da COSMIC tanto no processo de contagem quanto como ferramenta para aprimorar o planejamento de projetos públicos.

Referências

- Ali, M., Shaikh, Z., and Ali, E. (2015). Estimation of project size using user stories. In *International Conference on Recent Advances in Computer Systems*. Atlantis Press.
- BARBALHO, F. d. O. (2014). Uma abordagem comparativa entre duas técnicas de medição de software. Master's thesis, Universidade Federal de Pernambuco.
- Beck, K. (1999). Embracing change with extreme programming. *Computer*, (10):70–77.
- Berardi, E., Buglione, L., Cuadrado-Collego, J., Desharnais, J.-M., Gencel, C., Lesterhuis, A., Santillo, L., Symons, C., and Trudel, S. (2011). Guideline for the use of cosmic fsm to manage agile projects. *Common Software Measurement International Consortium*.
- Berardi, E. and Santillo, L. (2010). Cosmic-based project management in agile software development and mapping onto related cmmi-dev process areas. In *International Workshop on Software Measurement-IWSM 2010*. Citeseer.
- Bradley, M. et al. (2010). Function point counting practices manual, release 4.3.1. *International Function Point Users Group (IFPUG)*.
- Committee, C. M. P. et al. (2015). The cosmic functional size measurement method-version 4.0. 1-measurement manual. *The COSMIC Consortium, Montréal*.
- Cuadrado-Gallego, J. J., Buglione, L., Domínguez-Alda, M. J., de Sevilla, M. F., de Mesa, J. A. G., and Demirors, O. (2010). An experimental study on the conversion between ifpug and cosmic functional size measurement units. *Information and Software Technology*, 52(3):347–357.
- Desharnais, J., Buglione, L., and Kocaturk, B. (2011). Improving agile software projects planning using the cosmic method. In *workshop on Managing Client Value Creation Process in Agile Projects (Torre Cane, Italy)*.
- dos Santos Soares, M. (2004). Metodologias ágeis extreme programming e scrum para o desenvolvimento de software. *Revista Eletrônica de Sistemas de Informação*, 3(1).
- Fehlmann, T. (2002). Combinatory metrics for software development. In *8th International Symposium in QFD*.
- Karner, G. (1993). Resource estimation for objectory projects. *Objective Systems SF AB*, 17:1–9.
- Kitchenham, B. and Pfleeger, S. L. (2002a). Principles of survey research part 4: questionnaire evaluation. *ACM SIGSOFT Software Engineering Notes*, 27(3):20–23.
- Kitchenham, B. and Pfleeger, S. L. (2002b). Principles of survey research: part 5: populations and samples. *ACM SIGSOFT Software Engineering Notes*, 27(5):17–20.
- Kitchenham, B. A. and Pfleeger, S. L. (2002c). Principles of survey research part 2: designing a survey. *ACM SIGSOFT Software Engineering Notes*, 27(1):18–20.
- Kitchenham, B. A. and Pfleeger, S. L. (2002d). Principles of survey research: part 3: constructing a survey instrument. *ACM SIGSOFT Software Engineering Notes*, 27(2):20–24.

- Mendonça, A. and Alencar, F. (2019). Métricas de tamanho de software com métodos ágeis no setor público: uma revisão sistemática. *22nd Workshop on Requirements Engineering*.
- Nunes, R. D. (2017). A implantação das metodologias ágeis de desenvolvimento de software scrum e extreme programming (xp): uma alternativa para pequenas empresas do setor de tecnologia da informação. *ForScience*, 4(2).
- Pfleeger, S. L. and Kitchenham, B. A. (2001). Principles of survey research: part 1: turning lemons into lemonade. *ACM SIGSOFT Software Engineering Notes*, 26(6):16–18.
- Symons, C. (2001). Come back function point analysis (modernized)—all is forgiven!. In *Proc. of the 4th European Conference on Software Measurement and ICT Control, FESMA-DASMA*, pages 413–426. sn.
- Tomás, M. R. (2009). Métodos ágeis: características, pontos fortes e fracos e possibilidades de aplicação.
- Tridibesh, S. (2016). A guide to the scrum body of knowledge (sboktmguide). *Scrumstudy*. Phoenix, Arizona.
- Tytkowska, M., Werner, A., and Bach, M. (2015). Project management in the scrum methodology. In *International Conference: Beyond Databases, Architectures and Structures*, pages 483–492. Springer.