



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

Métricas para Desenvolvimento Ágil em Larga Escala: Um Levantamento da Literatura e da Indústria de Software Brasileira

Renato Vieira de Menezes

Recife - PE, Agosto 2024

Renato Vieira de Menezes

Métricas para Desenvolvimento Ágil em Larga Escala: Um Levantamento da Literatura e da Indústria de Software Brasileira

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática Aplicada da Pró-Reitoria de Pós-Graduação da Universidade Federal Rural de Pernambuco, como parte dos requisitos necessários para obtenção do grau de Mestre.

Universidade Federal Rural de Pernambuco

Pró-Reitoria de Pós-Graduação

Programa de Pós-Graduação em Informática Aplicada

Orientador: Prof. Dr. Marcelo Luiz Monteiro Marinho

Coorientador: Prof. Dra. Suzana Candido de Barros Sampaio

Recife - PE

Agosto 2024

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

D543m Menezes, Renato Vieira de.

Métricas para desenvolvimento ágil em larga escala : um levantamento da literatura e da indústria de software brasileira / Renato Vieira de Menezes . – Recife, 2024.

116 f.: il.

Orientador(a): Marcelo Luiz Monteiro Marinho.

Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Informática Aplicada, Recife, BR-PE, 2024.

Inclui referências.

1. Métrica 2. Desenvolvimento ágil de software 3. Medição de software 4. Métodos ágeis
I. Marinho, Marcelo Luiz Monteiro, orient. II. Título

CDD 004

Renato Vieira de Menezes

Métricas para Desenvolvimento Ágil em Larga Escala: Um Levantamento da Literatura e da Indústria de Software Brasileira

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática Aplicada da Pró-Reitoria de Pós-Graduação da Universidade Federal Rural de Pernambuco, como parte dos requisitos necessários para obtenção do grau de Mestre.

Aprovada em Recife - PE, 02 de Agosto de 2024:

**Prof. Dr. Marcelo Luiz Monteiro
Marinho** (Orientador)
Universidade Federal Rural de Pernambuco

**Prof. Dr. Ivaldir Honório de Fárias
Júnior**
Universidade de Pernambuco

**Prof. Dr. Ricardo André Cavalcante de
Souza**
Universidade Federal Rural de Pernambuco

Recife - PE
Agosto 2024

Aos meus pais, Ricardo e Lenilda, pelo amor e apoio sem fim em minha vida inteira

Agradecimentos

Agradeço a Deus primeiramente, por ter me concedido saúde e resiliência para que eu fosse em busca dos meus objetivos.

Agradeço também a minha família, principalmente aos meus pais, Lenilda e Ricardo. Eles sempre fizeram o possível para me dar o incentivo e suporte necessários em todas as etapas da minha vida. Também sou grato a minha companheira Amanda pela compreensão e ajuda durante esse processo. Sem essa rede de apoio, teria sido muito mais difícil.

Também gostaria de agradecer a todos os professores do Programa de Pós-Graduação em Informática Aplicada da Universidade Federal Rural de Pernambuco, em especial aos professores Marcelo Marinho, meu orientador e Suzana Sampaio, minha coorientadora, por todos os ensinamentos e direcionamentos passados desde o início dessa jornada acadêmica.

Finalmente, estou grato a mim por ter conseguido, mais uma vez, vencer desafios e me esforçar continuamente para crescer como profissional e como pessoa.

“A felicidade não está na estrada que leva a algum lugar. A felicidade é a própria estrada.”
(Bob Dylan)

Resumo

A maioria das organizações de desenvolvimento ágil de software são baseadas em equipes auto gerenciadas responsáveis pelo desenvolvimento de um subconjunto de funcionalidades. Todavia, conforme os projetos de grande porte incorporam cada vez mais práticas de desenvolvimento ágil, as empresas de software têm se adaptado para escalar os métodos ágeis. Gerenciar grandes projetos ágeis para além do nível de apenas uma equipe de desenvolvimento de software em uma organização é desafiador em razão de um conjunto mais diverso de fatores de influência. Nesse ambiente mais flexível e dinâmico as realizações de medições adequadas para cada realidade e objetivo são indispensáveis para promover a melhoria contínua e monitorar o progresso. Neste trabalho, serão apresentados os resultados de uma revisão *multivocal*, na qual a análise envolve uma revisão abrangente da literatura cinza, proporcionando compreender o conjunto de fontes que abordam usabilidade, soluções e adaptações de métricas voltadas para o desenvolvimento ágil de software em larga escala. Também serão divulgados os resultados de um *survey* conduzido na indústria, cujo os dados foram coletados durante os meses de Abril e Maio/2024 com uma amostra de 145 profissionais do mercado de software brasileiro. A pesquisa usou como referencial o conhecimento extraído das fontes de literatura para delinear como os participantes percebem os ambientes ágeis de larga escala em que estão inseridos, analisando o estado da prática das estratégias de seleção de métricas, quem faz uso das medições, os objetivos e tendências para adoção, além dos impactos alcançados. Portanto, objetivo geral desta pesquisa é examinar o uso atual de métricas no desenvolvimento ágil em grande escala, destacando os propósitos e as situações que determinam a seleção dessas métricas. Este cenário converge para um conjunto diverso de descobertas sobre os tópicos em discussão que deve orientar profissionais do mercado de software e expandir possibilidades para pesquisas futuras na área. Além disso, um catálogo de métricas com alta probabilidade de eficiência nesse contexto é entregue, contendo exemplos como ‘Velocidade’, ‘Eficiência de fluxo’, ‘*Work in Progress*’ e ‘Dívida técnica’.

Palavras-chaves: Métrica, Desenvolvimento Ágil de Software, Escalonamento dos Métodos Ágeis, Métodos Ágeis.

Abstract

Most agile software development organizations are based on self-managing teams responsible for developing a subset of features. However, as large-scale projects increasingly incorporate agile development practices, software companies have adapted to scale agile methods. Managing large agile projects beyond the level of a single software development team within an organization is challenging due to a more diverse set of influencing factors. In this more flexible and dynamic environment, performing appropriate measurements for each reality and objective is indispensable to promote continuous improvement and monitor progress. In this work, the results of a multivocal review will be presented, where the analysis involves a comprehensive review of the gray literature, providing an understanding of the sources that address usability, solutions, and adaptations of metrics aimed at large-scale agile software development. Additionally, the results of a survey conducted in the industry will be disclosed, with data collected during April and May 2024 from a sample of 145 professionals in the Brazilian software market. The research used knowledge extracted from literature sources as a reference to outline how participants perceive the large-scale agile environments in which they are embedded, analyzing the current state of practice in metric selection strategies, who utilizes these measurements, the objectives and trends for adoption, and the impacts achieved. Therefore, this research aims to examine the current use of metrics in large-scale agile development, highlighting the purposes and situations that determine the selection of these metrics. This scenario converges into diverse findings on the topics under discussion, which should guide software market professionals and expand possibilities for future research in the field. Furthermore, a catalog of metrics with a high efficiency probability in this context is provided, containing examples such as 'Velocity', 'Flow Efficiency', 'Work in Progress' and 'Technical Debt'.

Keywords: Metric, Agile Software Development, Large-Scale Agile, Agile Methods.

Lista de ilustrações

Figura 1 – Resumo das etapas de pesquisa	12
Figura 2 – Visão geral do processo (RML)	18
Figura 3 – Relação entre idade e escolaridade concluída	44
Figura 4 – Distribuição de respondentes por cargo, anos de experiência e tamanho da empresa	45
Figura 5 – Categorização da escala de projetos ágeis	45
Figura 6 – Distribuição geográfica das equipes	46
Figura 7 – Utilização de métodos/frameworks ágeis escalonados	46
Figura 8 – Estratégias de seleção das métricas ágeis	47
Figura 9 – Público-alvo das métricas	47
Figura 10 – Uso de métricas relacionadas a responsividade aos clientes	48
Figura 11 – Uso de métricas relacionadas a entrega de valor aos clientes	49
Figura 12 – Uso de métricas relacionadas a produtividade	50
Figura 13 – Uso de métricas relacionadas a sustentabilidade de práticas ágeis	51
Figura 14 – Uso de métricas relacionadas a qualidade do produto	51
Figura 15 – Áreas de impacto da utilização de métricas	52

Lista de tabelas

Tabela 1 – Visão geral dos trabalhos relacionados	9
Tabela 2 – Critérios metodológicos da pesquisa	13
Tabela 3 – Critérios de inclusão e exclusão	15
Tabela 4 – Resumo dos resultados da busca e seleção para a literatura (RML)	17
Tabela 5 – Critérios de qualidade	17
Tabela 6 – Visão geral de estudos acadêmicos	25
Tabela 7 – Contexto organizacional dos estudos	34
Tabela 8 – Métricas no desenvolvimento ágil de software em larga escala (RML)	35
Tabela 9 – Métricas vinculadas ao Ciclo de Vida de Desenvolvimento de Software	36
Tabela 10 – Somos mais responsivos aos clientes na forma ágil de trabalhar?	38
Tabela 11 – Entregamos mais valor aos clientes com a forma ágil de trabalhar?	39
Tabela 12 – Somos mais produtivos trabalhando ágil?	40
Tabela 13 – Temos melhores práticas de sustentabilidade no ambiente organizacional?	41
Tabela 14 – Temos melhor qualidade de produto?	42
Tabela 15 – Métricas relatadas na pesquisa	53
Tabela 16 – Modelo utilizado para descrição de métricas ágeis	61

Lista de abreviaturas e siglas

UFRPE	Universidade Federal Rural de Pernambuco
MSL	Mapeamento Sistemático de Literatura
RSL	Revisão Sistemática de Literatura
RML	Revisão <i>Multivocal</i> da Literatura
LeSS	<i>Large Scale Scrum</i>
SAFe	<i>Scaled Agile Framework</i>
DA	<i>Disciplined Agile</i>
CVDS	Ciclo de Vida de Desenvolvimento de Software
GQM	<i>Goal Question Metric</i>
WIP	<i>Work In Progress</i>
NPS	<i>Net Promoter Score</i>
ARTs	<i>Agile Release Trains</i>
PI	<i>Program Increment</i>
eNPS	<i>Employee Net Promoter Score</i>
PMEs	Pequenas e Médias Empresas

Sumário

1	INTRODUÇÃO	1
1.1	Motivação	1
1.2	Justificativa	2
1.3	Objetivos	3
1.4	Organização do trabalho	3
2	REFERENCIAL TEÓRICO	4
2.1	Métricas de Software	4
2.2	Desenvolvimento Ágil de Software em Larga Escala	5
2.3	Trabalhos Relacionados	6
2.4	Considerações Finais do Capítulo	9
3	METODOLOGIA	11
3.1	Etapas da Pesquisa	11
3.2	Critérios Metodológicos	12
3.3	Mapeamento Sistemático de Literatura	13
3.4	Revisão <i>Multivocal</i> de Literatura	14
3.4.1	Mecanismos de Busca	15
3.4.2	String de Busca	15
3.4.3	Critérios de Inclusão e Exclusão	15
3.4.4	Seleção	16
3.4.5	Avaliação da Qualidade	16
3.4.6	Extração de Dados	18
3.5	<i>Survey</i>	18
3.5.1	Público-Alvo e População	19
3.5.2	Amostragem	20
3.5.3	Desenho do Estudo	20
3.5.4	Instrumento e Avaliação	20
3.5.5	Análise dos Dados	21
3.5.5.1	Validação e limpeza de dados	21
3.5.5.2	Análise quantitativa	22
3.5.5.3	Análise qualitativa	22
3.6	Considerações Finais do Capítulo	22
4	RESULTADOS	24
4.1	Resultados do Mapeamento Sistemático de Literatura	24

4.1.1	Visão Geral dos Estudos Acadêmicos	24
4.1.2	Quais são os estudos publicados a respeito da utilização de métricas no desenvolvimento ágil em ambiente escalável?	24
4.1.3	Quais são as métricas adotadas no desenvolvimento ágil em ambiente escalável?	26
4.1.4	Discussão dos Resultados do Mapeamento	31
4.2	Resultados da Revisão <i>Multivocal</i> de Literatura	33
4.2.1	Contexto Organizacional	33
4.2.2	Quais métricas são usadas no desenvolvimento ágil de software em larga escala?	34
4.2.3	Quais são as razões e os resultados do uso de métricas no desenvolvimento ágil em larga escala?	37
4.2.4	Discussão dos Resultados da <i>Multivocal</i>	41
4.3	Resultados do <i>Survey</i>	43
4.3.1	Dados Demográficos	43
4.3.2	Seleção das Métricas Ágeis	46
4.3.3	Público-Alvo das Métricas	46
4.3.4	Objetivos e Impactos da Utilização de Métricas	47
4.3.5	Métricas Relatadas na Pesquisa	52
4.4	Discussão dos Resultados do <i>Survey</i>	52
4.4.1	Como e para quem as métricas são selecionadas?	53
4.4.2	Quais são os objetivos e impactos da utilização das métricas?	55
4.4.3	Quais métricas utilizadas no desenvolvimento ágil em larga escala não foram identificadas na literatura?	56
4.5	Análise Comparativa dos Resultados da Revisão <i>Multivocal</i> e do <i>Survey</i> sobre Métricas Ágeis Escalonadas	56
4.6	Considerações Finais do Capítulo	58
5	CATÁLOGO DE MÉTRICAS PARA DESENVOLVIMENTO ÁGIL EM LARGA ESCALA	60
5.1	Visão Geral das Métricas	60
5.2	Métricas	60
5.2.1	Responsividade aos Clientes	61
5.2.1.1	<i>Lead Time</i>	61
5.2.1.2	Tempo gasto desde o início do projeto até a entrega do primeiro incremento (Em inglês: <i>Time to delivery first increment</i>)	62
5.2.1.3	Tempo gasto desde o início do projeto até o encerramento do projeto (Em inglês: <i>Time to project closure</i>)	62
5.2.1.4	Tempo gasto (todas as tarefas)	63
5.2.1.5	Tempo gasto (tarefas principais usando a conclusão de subtarefas)	64
5.2.1.6	Tempo de resposta da solicitação de atendimento ao cliente	64

5.2.1.7	Distribuição de fluxo	65
5.2.1.8	Tempo corrido	65
5.2.1.9	Tempo disponível	66
5.2.1.10	Tempo restante	66
5.2.1.11	Tempo de espera pelas alterações	67
5.2.1.12	Tempo de ciclo por recurso	67
5.2.2	Entrega de Valor	68
5.2.2.1	Eficiência de fluxo	68
5.2.3	Produtividade	68
5.2.3.1	Velocidade	68
5.2.3.2	Impedimentos	69
5.2.3.3	Tarefas bloqueadas	69
5.2.3.4	Eficiência	70
5.2.3.5	Gráfico <i>burndown</i>	70
5.2.3.6	Gráfico de <i>burnup</i>	71
5.2.4	Sustentabilidade de Práticas Ágeis	71
5.2.4.1	<i>Work In Progress</i>	71
5.2.5	Qualidade do Produto	72
5.2.5.1	Cobertura de testes unitários por histórias	72
5.2.5.2	Dívida técnica	72
5.2.5.3	Número de problemas externos	73
5.2.5.4	Dias abertos com problemas externos	73
5.3	Considerações Finais do Capítulo	74
6	DISCUSSÃO	75
6.1	Limitações e Ameaças à Validade	76
6.2	Trabalhos Futuros	77
6.3	Considerações Finais do Capítulo	78
7	CONCLUSÕES	79
7.1	Contribuições	80
	REFERÊNCIAS	82
A	APÊNDICE - FORMULÁRIO DE PESQUISA.	88
B	APÊNDICE - GUIA DE MÉTRICAS.	98

1 Introdução

Ao longo dos anos a adoção de variadas metodologias ágeis segue crescendo, grandes organizações estão adotando cada vez mais práticas ágeis para ganhar uma vantagem competitiva, priorizando a satisfação do cliente como principal objetivo (TRIPATHI *et al.*, 2015; MONTANO *et al.*, 2022). Inseridos nessa nova realidade, os mais diversos setores da indústria passaram a depender estrategicamente do mercado de tecnologia, demandando ampla atenção para o desenvolvimento ágil de software em contextos de larga escala (TRIPATHI *et al.*, 2015; GRIMALDI *et al.*, 2016).

Coordenar múltiplas equipes trabalhando em um único projeto ou em diferentes projetos no contexto de um grande programa e portfólio é um desafio complexo na abordagem ágil, pela necessidade de alinhar comunicação, dependências, integração e *deploy* (MARINHO; CAMARA; SAMPAIO, 2021; CAMARA; MARINHO; MOURA, 2021). Para aumentar as chances de sucesso de projetos de software, o gerenciamento em métodos ágeis tem sido apoiado pelo uso de métricas, como velocidade, tempo de ciclo, dentre muitas outras (LEAL *et al.*, 2022). Métricas são medições que quantificam propriedades de entidades que são mensuráveis (FENTON; BIEMAN, 2014). À medida que a indústria de software cresce, a dependência de métricas torna-se cada vez mais evidente (ERTABAN; SARIKAYA; BAGRIYANIK, 2018). Dingsøy e Brede (DINGSØYR; MOE, 2014) estabelecem como alta a prioridade de identificar métricas apropriadas para que as partes interessadas monitorem o progresso promovendo a transparência. Greening (GREENING, 2015) discute como determinadas métricas podem fornecer a base para relatórios usados por treinadores que buscam melhorar o desempenho das equipes. Kettunen (KETTUNEN, 2014) defende que organizações podem ser melhoradas e transformadas instruindo as equipes a medir seu desempenho de acordo com os estados desejados.

Logo, há uma preocupação dos pesquisadores sobre as métricas em ambientes escaláveis, que desempenham um papel fundamental por duas razões principais: avaliar o desempenho das equipes de desenvolvimento e promover a busca constante pela melhoria (ERTABAN; SARIKAYA; BAGRIYANIK, 2018). Além disso, a medição oferece uma valiosa capacidade de compreensão, permitindo identificar problemas e ineficiências no fluxo de valor, trazendo segurança para sua eliminação (ERTABAN; SARIKAYA; BAGRIYANIK, 2018).

1.1 Motivação

Usar métricas ágeis de forma eficaz é desafiador para grandes organizações de software devido a uma grande quantidade de fatores de influência, demandas de habilidades e

infraestrutura e falta de treinamento, tornando complicada a medição eficiente (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021). Desse modo, para monitorar processos e projetos é necessário depender do *feedback* qualitativo e comunicar-se regularmente com cada equipe ágil para orientar a tomada de decisões da gestão (PHILIPP et al., 2023; CAMARA; MARINHO; MOURA, 2021). Nessas circunstâncias, as métricas adequadas cumprem a função de aprimorar ou substituir o *feedback* qualitativo (ERTABAN; SARIKAYA; BAGRIYANIK, 2018; PHILIPP et al., 2023).

No entanto, apesar do papel fundamental das métricas neste contexto e dos apelos repetidos dos pesquisadores por mais conhecimento documentado, ainda há escassez na literatura acadêmica de estudos dedicados para preencher essa lacuna (DINGSØYR; MOE, 2014; EDISON; WANG; CONBOY, 2021; KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021; PHILIPP et al., 2023). Em vista disso, torna-se importante pesquisar com profundidade as evidências empíricas de aplicabilidades das métricas utilizadas pelas empresas que praticam agilidade em larga escala.

1.2 Justificativa

A crescente adoção de metodologias ágeis ao longo dos anos reflete não apenas uma mudança nas práticas de desenvolvimento de software, mas também uma transformação fundamental na maneira como as organizações abordam a entrega de valor aos clientes (TRIPATHI et al., 2015). Grandes organizações estão cada vez mais adotando práticas ágeis como meio de garantir uma vantagem competitiva, priorizando a satisfação do cliente como objetivo central (MONTANO et al., 2022). Esse movimento é especialmente evidente nos setores da indústria que dependem estrategicamente do mercado de tecnologia, onde a agilidade no desenvolvimento de software se tornou uma necessidade imperativa para permanecer relevante e competitivo (GRIMALDI et al., 2016).

Nesse cenário, o uso de métricas tem emergido como um suporte fundamental para o gerenciamento eficaz. Contudo, há uma lacuna significativa na literatura acadêmica que destaca a necessidade de uma pesquisa aprofundada para compreender melhor as práticas, desafios e oportunidades relacionadas ao uso de métricas em ambientes ágeis de grande porte (DINGSØYR; MOE, 2014; EDISON; WANG; CONBOY, 2021; KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021; PHILIPP et al., 2023).

Portanto, a presente pesquisa visa responder como as métricas são utilizadas no desenvolvimento ágil de software em larga escala, identificando o panorama contemporâneo da sua aplicação, enfatizando os objetivos e contextos que orientam a escolha dessas métricas. Por meio da investigação dessas questões, pretende-se contribuir para o avanço do conhecimento na área, fornecendo *insights* valiosos para profissionais do setor e preenchendo uma lacuna importante na literatura acadêmica. Os resultados desta pesquisa têm o

potencial de informar práticas e políticas relacionadas ao desenvolvimento de software em larga escala, além de beneficiar as organizações que buscam melhorar sua eficácia e eficiência no contexto ágil, proporcionando o desenvolvimento de estratégias de medição que possibilitem o alcance de resultados mais precisos.

1.3 Objetivos

Diante da justificativa descrita, esta pesquisa visa como objetivo geral: apresentar o cenário atual das utilizações de métricas no desenvolvimento ágil em larga escala, destacando as finalidades e circunstâncias pelas quais são selecionadas.

Como objetivos específicos, pretende-se:

- identificar as métricas utilizadas por organizações de software que implementam metodologias ágeis escaláveis;
- investigar como e para quem as métricas são selecionadas;
- investigar os objetivos específicos de cada métrica, tais como melhoria de processos, controle de qualidade, produtividade e satisfação do cliente;
- mapear áreas em que as métricas tem impacto nas organizações;
- mapear métricas emergentes e práticas inovadoras que não foram amplamente documentadas na literatura acadêmica;
- elaborar um catálogo de métricas aplicáveis a projetos de desenvolvimento ágil em larga escala.

1.4 Organização do trabalho

Este trabalho está organizado em 7 capítulos. O Capítulo 2 apresenta o embasamento teórico. Explicações sobre Métricas de Software e Desenvolvimento Ágil de Software em Larga Escala são encontradas neste capítulo. Além disso, também apresenta os trabalhos relacionados ao tema desta pesquisa, descrevendo suas metodologias e resultados. O Capítulo 3 relata as etapas de pesquisa, os critérios metodológicos e o detalhamento dos procedimentos seguidos neste estudo. O Capítulo 4 descreve os resultados alcançados acerca das questões de pesquisa deste estudo por meio da aplicação de cada metodologia. Uma discussão focada por etapa de pesquisa também é efetuada. O Capítulo 5 compartilha um catálogo resultante de métricas validadas e eficazes para o desenvolvimento ágil em larga escala. O Capítulo 6 apresenta uma discussão integrada fundamentada na literatura sobre os resultados da pesquisa, expõe suas limitações e sugere direcionamentos para trabalhos futuros. Por fim, o Capítulo 7 refere-se as conclusões e contribuições da pesquisa.

2 Referencial Teórico

Este capítulo apresentará as fundamentações teóricas a respeito dos principais temas de interesse desta dissertação. Além disso, serão apresentados os principais trabalhos que se relacionam de alguma forma com a temática deste trabalho, a fim de proporcionar o conhecimento sobre as iniciativas, metodologias, resultados e uma breve discussão comparativa.

2.1 Métricas de Software

Métricas são medições, variáveis que quantificam propriedades de entidades que são mensuráveis (FENTON; BIEMAN, 2014; LEAL et al., 2022). No desenvolvimento de software tradicional, as métricas são utilizadas para auxiliar a orientação de projetos e tomada de decisão (CHLOROS; GEROGIANNIS; KAKARONTZAS, 2022). O processo de medição tradicional, no entanto, pode ser trabalhoso e burocrático para organizações que adotam métodos ágeis (LOMIO et al., 2022).

Com a popularização dos valores ágeis surgiram diferentes processos de desenvolvimento de software, enquanto as métricas aplicadas no desenvolvimento de software tradicional medem o cumprimento de um plano pré-definido, as métricas ágeis precisaram passar a suportar mudanças em todas as fases do desenvolvimento (HOSSAIN; AHMED; ARAFAT, 2021; LEAL et al., 2022). Métricas ágeis devem enfatizar o valor entregue, afirmar e reforçar os princípios ágeis, são fáceis de coletar e podem fornecer *feedback* de forma frequente e regular (LEAL et al., 2022). Devido às suas vantagens, uma infinidade de métricas ágeis têm sido propostas ou adaptadas para ajudar a medir vários aspectos dos processos e produtos de desenvolvimento de software (LEAL et al., 2022). As equipes ágeis têm usado métricas com foco em obter benefícios nas seguintes áreas: planejamento da *sprint*, acompanhamento do progresso, medição da qualidade do software, correção de problemas no processo de software e motivação das pessoas (KUPIAINEN; MÄNTYLÄ; ITKONEN, 2015).

Inúmeras são as justificativas para que as equipes ágeis de desenvolvimento de software utilizem as medições (BARCELLOS; ROCHA, 2008; FENTON; BIEMAN, 2014; KUPIAINEN; MÄNTYLÄ; ITKONEN, 2015; ERTABAN; SARIKAYA; BAGRIYANIK, 2018; KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021). Avaliação de desempenho de equipes de desenvolvimento e busca da melhoria contínua (BARCELLOS; ROCHA, 2008; ERTABAN; SARIKAYA; BAGRIYANIK, 2018); dar visibilidade do que está sendo desenvolvido, analisar a qualidade do desenvolvimento e fornecer *insights* para organização (KUPIAINEN; MÄNTYLÄ; ITKONEN, 2015); proporcionar a compreensão dos problemas

e perdas no fluxo de valor para que possam ser eliminados (ERTABAN; SARIKAYA; BAGRIYANIK, 2018); mensurar ou prever atributos de processos, produtos ou recursos (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021); planejamento da *sprint*, acompanhamento do progresso, medição da qualidade do software, motivação das pessoas (KUPIAINEN; MÄNTYLÄ; ITKONEN, 2015); entre outras. Ainda que as equipes ágeis utilizem muitas métricas sugeridas na literatura, elas também usufruem de muitas métricas personalizadas (KUPIAINEN; MÄNTYLÄ; ITKONEN, 2015; HOSSAIN; AHMED; ARAFAT, 2021).

Embora a literatura científica tenha discutido o papel que as métricas desempenham no contexto ágil, ainda há evidências empíricas insuficientes sobre esse papel no contexto industrial, especialmente na visão de grandes empresas (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021).

2.2 Desenvolvimento Ágil de Software em Larga Escala

Grande parte das organizações de desenvolvimento ágil de software são compostas por equipes que se auto gerenciam e são responsáveis pelo desenvolvimento de um subconjunto de recursos (GRIMALDI et al., 2016; CAMARA et al., 2020; LEAL et al., 2022). Escalar agilidade é uma forma de gerenciar processos para além do desenvolvimento de software a nível de equipe utilizando ferramentas ágeis definidas (LAANTI, 2014).

De acordo com Ambler e Lines (AMBLER; LINES, 2013), o escalonamento ágil pode ser interpretado de três maneiras. Primeiro, escalar o ágil em toda a organização significa escalar o ágil entre as equipes, isso se manifesta quando equipes de software ou multidisciplinares colaboram de forma integrada para alcançar objetivos comuns. Em segundo lugar, escalar o ágil para outros projetos significa escalar o tamanho do projeto. Terceiro, escalar o ágil para todo o fluxo de valor de uma organização significa dimensionar e maximizar o desenvolvimento do valor do negócio sem focar apenas nas equipes ágeis, mas também na coordenação de todo empreendimento. Por isso, diversos são os fatores que podem ser considerados ao escalar o ágil, como tamanho e distribuição geográfica das equipes, complexidade do sistema, cultura e distribuição organizacional, governança e metas de negócio (RAZZAK et al., 2017; SHAMEEM et al., 2017).

Além disso, como o número de equipes é uma condição determinante na escala do projeto, este estudo segue a definição de agilidade em larga escala de Dingsoyr *et al.* (DINGSØYR; FÆGRI; ITKONEN, 2014), que categoriza a escala de projetos ágeis de desenvolvimento de software por número de equipes. Ainda segundo os autores, projetos de pequena escala consistem em uma única equipe, coordenada por meio de práticas ágeis padrão. Projetos de grande porte envolvem de 2 a 9 equipes, exigindo metodologias escalonadas para coordenação. Projetos com 10 ou mais equipes são considerados de grande escala, necessitando de um *framework* de escalonamento.

Os principais *frameworks* de escalonamento ágeis conhecidos atualmente são: *Scrum-at-Scale*, *Large Scale Scrum* (LeSS), *Scaled Agile Framework* (SAFe), *Disciplined Agile* (DA), Spotify e *Nexus* (SHAMEEM et al., 2017; EDISON; WANG; CONBOY, 2021). Nas linhas seguintes, serão fornecidas uma breve descrição dos seis *frameworks* ágeis de escalonamento mais comuns:

- SAFe - *Scaled Agile Framework* foca em grandes empresas e cobre a abordagem em larga escala organizacional para adoção de agilidade (RAZZAK et al., 2017). Além disso, o SAFe dispõe de um rico sistema de métricas nativas incorporadas (SAFE, 2022);
- *Scrum-at-Scale* - *Scrum-at-Scale* é uma estrutura organizacional na qual equipes operando de forma consistente com o *Scrum Guide* podem resolver problemas adaptativos complexos, ao mesmo tempo em que entrega produtos de maior valor possível (SCRUM@SCALE, 2022);
- LeSS - *Large Scale Scrum* é um *framework* que possui duas abordagens: (i) o *framework* LeSS é projetado para médias empresas, e (ii) LeSS *huge* projetado para grandes empresas (LARMAN; VODDE, 2016);
- Spotify - o modelo Spotify foca na organização das equipes em *squads*, sendo que cada *squad* utiliza práticas ágeis preferíveis (KNIBERG; IVARSSON, 2012);
- *Nexus* - o *Nexus* fornece uma estrutura que consiste em um conjunto de funções, práticas e eventos que servem para grandes equipes de desenvolvimento distribuído de software (SCHWABER, 2015);
- DA - *Disciplined Agile* é uma estrutura que amplia as práticas Scrum e as combina com outros métodos ágeis (AMBLER; LINES, 2012).

Em comparação ao SAFe, os demais *frameworks* apresentam menos métricas nativas, artefatos, funções e eventos além do Scrum (RAZZAK et al., 2017).

2.3 Trabalhos Relacionados

O primeiro trabalho relacionado a ser destacado neste capítulo é (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021), cuja ideia principal consiste em suprir uma lacuna de conhecimento em relação à medição de desempenho e seus desafios nas organizações ágeis em escala, investigando como elas medem o desempenho em diferentes níveis organizacionais. Para isso, foi aplicado um estudo de caso exploratório buscando examinar objetivos e métricas de desempenho em duas unidades de negócios de uma organização ágil.

Destaca-se que o estudo sugere um *framework* para selecionar e estruturar objetivos de desempenho. Foram sugeridas cinco dimensões de medição (valor para o cliente, valor financeiro, eficiência de processos internos, colaboração e aprendizagem) que podem ser aplicadas para fornecer orientação para organizações ágeis de larga escala na seleção e priorização dos objetivos de desempenho. O *framework* criado deve ajudar profissionais na seleção e priorização das métricas e na compreensão dos desafios relacionados. Embora apresente resultados detalhados, o estudo focou apenas na estrutura SAFe com base em uma organização, tornando necessário verificar as possibilidades de observação desses *insights* em outras empresas que implementam SAFe em contextos similares.

O trabalho de Philipp *et al.* (PHILIPP *et al.*, 2023) apresentou um método baseado em objetivos, desenvolvido para uma grande organização ágil, com o intuito de apoiar os *stakeholders* na seleção, operação e escalabilidade de métricas. Além disso, com base nas lições aprendidas na organização estudada, os autores propõem princípios de design que podem potencialmente orientar o desenvolvimento de métodos adequados para outros contextos. A pesquisa foi conduzida utilizando uma abordagem de pesquisa-ação, combinada com a engenharia de métodos situacional.

Os resultados indicam que o método se mostrou eficaz para a organização em questão, ao combinar elementos consagrados de programas de medição projetados para a engenharia de software tradicional, como a orientação por objetivos e o suporte de ferramentas, com elementos exclusivos do método proposto, como atividades de escalabilidade de métricas e alinhamento com o desenvolvimento ágil de software. O estudo oferece *insights* aprofundados sobre a gestão de métricas em uma grande organização ágil, servindo como base para que pesquisadores e profissionais possam desenvolver programas de medição adequados a outras organizações ágeis em escala. Contudo, as limitações do estudo incluem a possibilidade de viés organizacional, já que o método foi desenvolvido com base em apenas duas organizações, sendo o design final especialmente adaptado ao ambiente da organização referenciada. Além disso, os princípios de design propostos provavelmente estão incompletos, e a aplicação do método em outros ambientes-alvo pode exigir a adição ou adaptação de design propostos. Portanto, uma visão prática e abrangente das estratégias de seleção de métricas nas grandes empresas pode ser usada como complemento para apoiar a generalização do método desenvolvido em (PHILIPP *et al.*, 2023).

Em outro trabalho relacionado relevante, seguindo a tendência crescente de interesse em métricas ágeis, Leal *et al.* (LEAL *et al.*, 2022) apresentaram um estudo com foco sobre como as métricas ágeis são usadas, o contexto de uso, como elas são selecionadas e classificadas, o público-alvo das métricas e o impacto de seus usos. Para reunir esse conhecimento, foi realizado um Mapeamento Sistemático de Literatura (MSL) sobre o uso de métricas em organizações de software ágeis.

Foi reportado que a maioria das organizações utiliza métricas ágeis com Scrum, *Lean* ou XP. Grande parte dos estudos definiu métricas de qualidade do produto, recursos e custos; dentre as quais algumas métricas atreladas ao desenvolvimento ágil em larga escala foram apresentadas. Para os pesquisadores o estudo apresenta um mapa detalhado da seleção, uso e resultados observados de métricas em contextos ágeis, servindo como um ponto de partida para investigar lacunas na aplicação de métricas em diferentes contextos organizacionais, métodos ágeis ou perfis de produtos. Por outro lado, os profissionais da indústria podem encontrar um catálogo organizado de métricas ágeis e escolher aquelas mais adequadas para melhorar a gestão ágil de acordo com seus contextos organizacionais ou de produto específicos.

Apesar de Leal *et al.* (LEAL *et al.*, 2022) abordarem aspectos pertinentes da análise de como organizações ágeis utilizam métricas e de terem trazido algumas evidências de medições em projetos de larga escala, o estudo não foi conduzido para cobrir o campo do escalonamento ágil. A *string* de busca utilizada nesse trabalho evidencia a intenção de não segmentar a pesquisa para o uso de métricas no ágil em larga escala, visto que não foram inclusos termos para orientar a busca por fontes que abordam o uso de medições relacionadas a *frameworks* de escalonamento ágil reconhecidos, como proposto no presente trabalho. Além disso, termos como “large-scale” e “scaling agile” também não foram aplicados. Em (LEAL *et al.*, 2022) também foi disponibilizado um catálogo métricas resultante do conhecimento extraído de cada fonte da revisão de literatura, o que difere do catálogo a ser apresentado nesta pesquisa principalmente pelos seguintes aspectos: i) as métricas listadas foram encontradas apenas na literatura voltada a contextos de larga escala; e ii) todas as métricas catalogadas passaram pela avaliação de popularidade na indústria a partir de um levantamento.

Britto *et al.* (BRITTO; USMAN; MENDES, 2014) relataram o estado da arte em estimativa de esforço no desenvolvimento de software no contexto global e ágil. Baseado em duas revisões da literatura, uma dentro do contexto de desenvolvimento ágil de software e outra no contexto de desenvolvimento de software global, esse artigo combinou os resultados dessas RSLs. Algumas métricas de tamanho foram identificadas em estudos primários como ‘pontos de função’, ‘linhas de código’ e ‘pontos de história’. Contudo, o trabalho é segmentado apenas para evidências de estimativa de esforço, deixando de fora outras etapas ou disciplinas relacionadas ao desenvolvimento de software tais como estimativa de custos e entrega de valor, por exemplo.

Os autores Edison *et al.* (EDISON; WANG; CONBOY, 2021) realizaram uma RSL a fim de comparar os diferentes métodos e *frameworks* no desenvolvimento ágil em larga escala com base na literatura. No âmbito das métricas, a contribuição mais importante desse trabalho é que ele apresenta-se como o primeiro estudo secundário que buscou destacar as medições existentes na literatura para todos os *frameworks* ágeis em

larga escala conhecidos. Porém, embora tenha registrado a existência de muitas métricas para o SAFe, não foram encontradas informações explícitas na literatura sobre métricas que poderiam ser usadas nas demais estruturas escalonadas, como LeSS, DA e Spotify. A Tabela 1 apresenta uma breve visão geral dos trabalhos relacionados, títulos, ano de publicação e métodos utilizados.

Tabela 1 – Visão geral dos trabalhos relacionados

Ano	Título	Método de Estudo	Referência
2014	<i>Effort estimation in agile global software development contex</i>	Revisão sistemática de revisões	(BRITTO; USMAN; MENDES, 2014)
2021	<i>Performance measurement in scaled agile organizations</i>	Estudo de caso/exploratório	(KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021)
2021	<i>Comparing methods for large-scale agile software development: A systematic literature review</i>	Revisão sistemática de literatura	(EDISON; WANG; CONBOY, 2021)
2022	<i>How agile organizations use metrics: A systematic literature mapping</i>	Mapeamento sistemático de literatura	(LEAL et al., 2022)
2023	<i>A method for metric management at a large-scale agile software development organization</i>	Pesquisa-ação	(PHILIPP et al., 2023)

A necessidade desta pesquisa decorre da diversidade de métricas e *frameworks* ágeis utilizados e da falta de estudos abrangentes em diferentes contextos organizacionais. Após a análise dos trabalhos relacionados à temática proposta foi possível verificar que a maior parte do conhecimento está fragmentado em estudos secundários ou em casos específicos limitados a determinados *frameworks* ou contextos. Portanto existe uma lacuna referente a relevância de revisão de literatura abrangente e *survey* para fornecer uma visão prática e ampla sobre a aplicação e adaptação dessas métricas, identificando desafios e benefícios reais. Assim, este trabalho contribui para o corpo teórico com dados empíricos atualizados e oferece aos praticantes *insights* para melhorar processos e selecionar métricas mais eficazes, promovendo um ambiente de desenvolvimento alinhado com os objetivos de negócio.

2.4 Considerações Finais do Capítulo

Neste capítulo foram mostrados os referenciais teóricos fundamentais para a compreensão deste trabalho como Métricas de Software e Desenvolvimento Ágil de Software em Larga Escala. A métrica de software foi abordada de modo geral no que diz respeito as razões para sua existência e utilizações partindo do desenvolvimento de software tradicional para o contexto ágil. Em relação ao desenvolvimento ágil de software em larga escala foram abordados os aspectos de sua natureza, definições técnicas, fatores considerados para escalar o ágil e os *frameworks* mais conhecidos comparando-os em relação ao SAFe, que é mais adotado.

Também foram apresentados os principais trabalhos relacionados à temática proposta desta dissertação. Após a análise comparativa e das limitações dessas iniciativas em relação ao presente trabalho foi possível verificar que existe uma lacuna referente a relevância de revisão de literatura abrangente e *survey* para ajudar a solidificar o conhecimento das medições implementadas em ambientes ágeis escaláveis. No próximo capítulo, serão apresentadas as etapas de pesquisa, os critérios metodológicos e o detalhamento dos procedimentos seguidos para buscar preencher as lacunas identificadas.

3 Metodologia

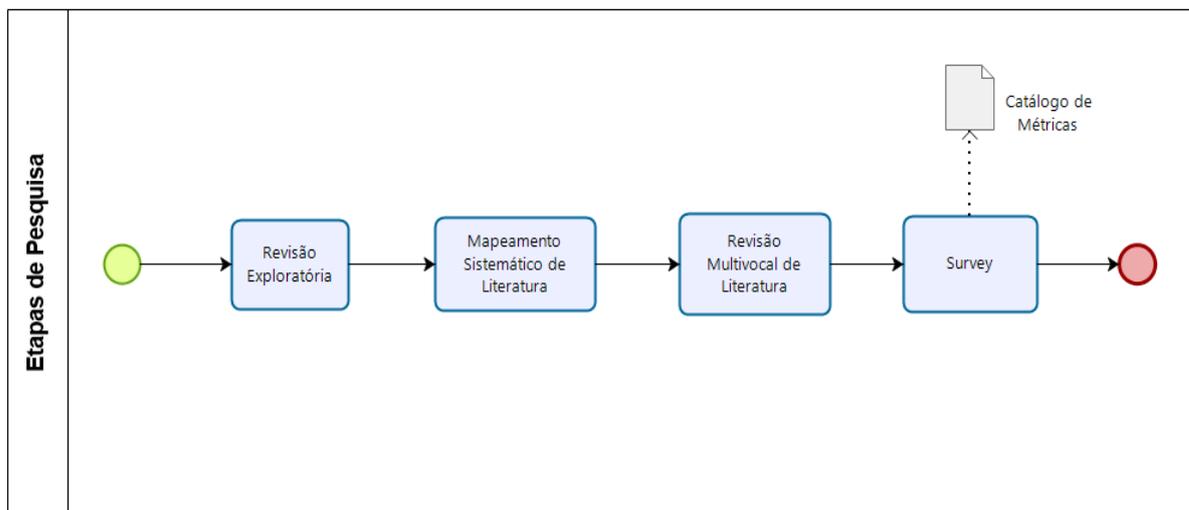
Neste capítulo serão apresentadas as etapas de pesquisa, a classificação e descrição dos critérios metodológicos bem como o detalhamento dos procedimentos técnicos seguidos em cada etapa que guiaram este estudo, considerando que a metodologia é aplicação de procedimentos e técnicas que devem ser observadas para construção do conhecimento com propósito de comprovar sua validade e utilidade (PRODANOV; FREITAS, 2013).

3.1 Etapas da Pesquisa

Nesta seção apresenta-se um resumo das etapas do percurso metodológico desta pesquisa que pode ser dividido em quatro etapas demonstradas na Figura 1. As etapas consistiram em:

- **Etapa 1:** como ponto de partida, foi realizada uma revisão exploratória inicial para verificar a existência da lacuna de conhecimento e áreas de interesse. O objetivo foi delinear os limites do campo e a identificar as principais questões, teorias e metodologias envolvidas para servir como uma base para a formulação de perguntas de pesquisa mais específicas.
- **Etapa 2:** a fim de responder as questões que dirigem esta pesquisa inicialmente foi realizado um Mapeamento Sistemático de Literatura, pois o escopo do tópico estudado necessitava descrever mais detalhadamente os resultados e a abrangência da pesquisa de métricas (MENEZES; MARINHO; SAMPAIO, 2022).
- **Etapa 3:** visando complementar o mapeamento sistemático, identificar lacunas no conhecimento e contribuir para o corpo de conhecimento em métodos ágeis de larga escala, foi realizada uma Revisão *Multivocal* de Literatura (RML) para reunir e sintetizar as informações limitadas encontradas na literatura acadêmica junto a outras descobertas da literatura cinza (MENEZES; MARINHO; SAMPAIO, 2024b). Foram extraídas informações sobre quais métricas são usadas no desenvolvimento ágil de software em larga escala, destacando por que são usadas e os resultados obtidos.
- **Etapa 4:** finalmente, foi necessário confrontar os resultados da revisão com as práticas do mercado. Para isso, foi realizado um *survey* na indústria a fim de verificar e enriquecer o conhecimento prévio. Os resultados, incluso um catálogo de métricas, são apresentados pela primeira vez neste trabalho.

Figura 1 – Resumo das etapas de pesquisa



3.2 Critérios Metodológicos

Este estudo pode ser classificado como uma abordagem qualitativa, pois se dispõe a explorar e compreender as percepções que os indivíduos ou grupos atribuem a práticas e métricas utilizadas no desenvolvimento ágil de software em larga escala. A análise e interpretação dos dados é construída indutivamente pelo pesquisador, com foco no significado individual e na importância de representar a complexidade do tema (CRESWELL; CRESWELL, 2014).

A pesquisa tem intuito de propor recomendações úteis para realização de medições em organizações de software que praticam agilidade de maneira escalonada, em razão disso, do ponto de vista de sua natureza é categorizada como uma pesquisa aplicada, em que o objetivo é gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos. Envolve verdades e interesses locais (PRODANOV; FREITAS, 2013).

Quanto a seus objetivos este estudo pode ser classificado como uma pesquisa exploratória e descritiva, visto que parte de uma fase preliminar com finalidade de proporcionar mais informações sobre o assunto em investigação, possibilitando sua definição e seu delineamento, em busca de delimitar o tema da pesquisa; orientar a fixação dos objetivos e a formulação das hipóteses ou descobrir um novo tipo de enfoque para o assunto (PRODANOV; FREITAS, 2013). Posteriormente, a pesquisa também se propõe a registrar e descrever os fatos observados sem interferir neles, visando detalhar as características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis por meio de técnicas padronizadas de coleta de dados.

Ainda observou-se as evidências do possível estado da arte da utilização efetiva de métricas no desenvolvimento ágil escalonado para desse modo verificar inferências. Por

isso, o método científico caracteriza-se por ser indutivo, no qual o objetivo dos argumentos indutivos é levar a conclusões prováveis sustentadas por premissas verdadeiras nas quais as conclusões se basearam (MARCONI; LAKATOS, 2003).

Já quanto aos procedimentos técnicos esta pesquisa é classificada como bibliográfica, documental e de levantamento. A pesquisa bibliográfica refere-se ao mapeamento sistemático de literatura realizado que buscou identificar, classificar e quantificar a literatura disponível, permitindo compreender a distribuição de pesquisas neste campo (PETERSEN et al., 2008). A revisão *multivocal* conduzida é uma forma de pesquisa que pode ser considerada tanto bibliográfica quanto documental, pois incorpora fontes acadêmicas e não acadêmicas para obter uma visão mais abrangente sobre um tópico específico (OGAWA; MALEN, 1991). Relacionada a pesquisa de levantamento está a coleta de uma amostragem para levantar características de uma população através de uma pesquisa do tipo *survey*, em que os dados dão suporte a construções, explicações, análise de tendências e relacionamentos (WAGNER et al., 2020). A Tabela 2 dispõe resumidamente as classificações para cada um dos critérios que ajudaram a construir esse estudo.

Tabela 2 – Critérios metodológicos da pesquisa

Critério	Classificação
Natureza	Aplicada
Objetivos do estudo	Pesquisa exploratória e descritiva
Abordagem	Qualitativa
Método científico	Indutivo
Procedimentos técnicos	Pesquisa bibliográfica, documental e levantamento

3.3 Mapeamento Sistemático de Literatura

O MSL fornece um procedimento sistemático e objetivo para identificar e classificar quais evidências estão disponíveis em uma área de pesquisa específica (PETERSEN et al., 2008). Ao coletar e analisar os principais estudos científicos sobre métricas aplicadas junto aos métodos ágeis em larga escala, pretendeu-se cobrir a amplitude da área de pesquisa. Os dados também foram sintetizados para identificar e classificar o conjunto de métricas usadas.

As diretrizes de Kitchenham e Charters (KITCHENHAM; CHARTERS, 2007) foram utilizadas como guia para o MSL. Assim, as seguintes perguntas orientaram o mapeamento deste estudo:

- (P1) “Quais são os estudos publicados a respeito da utilização de métricas no desenvolvimento ágil em ambiente escalável?”.

- **(P2)** “*Quais são as métricas adotadas no desenvolvimento ágil em ambiente escalável?*”.

Os trabalhos incluídos deveriam apresentar como critério o uso de métricas relacionadas a aplicação da agilidade em projetos de larga escala, soluções ou adaptações de métricas nesse cenário. Artigos não publicados em jornais ou conferências e que não estivessem publicados em inglês seriam excluídos. A string e mecanismos da busca, os critérios de inclusão, exclusão e processos de seleção para artigos que foram inclusos neste mapeamento serão melhor compreendidos na Seção 3.4; que também descreve a metodologia da revisão *multivocal* conduzida logo após, cujas fontes de literatura acadêmica e questões de pesquisa foram derivadas do mapeamento. Apenas um artigo incluso no mapeamento (KETTUNEN, 2014) não seguiu para revisão *multivocal*, pelo fato de apresentar somente diretrizes que sugerem a possível adoção de métricas de desempenho a serem utilizadas, as métricas em si não foram citadas explicitamente, apenas diretrizes de medição.

3.4 Revisão *Multivocal* de Literatura

Revisão Sistemática de Literatura e Mapeamento Sistemático de Literatura tornaram-se amplamente populares na engenharia de software. Embora esses procedimentos sejam importantes, alguns pesquisadores destacaram que os resultados de ambos podem fornecer um corpo padrão de conhecimento (AMPATZOGLU et al., 2015). Uma vez que esses estudos não incluem a literatura “cinza” (fontes de informação não publicadas e não revisadas por pares), que é produzida constantemente em grande volume por profissionais, eles não conseguem oferecer totalmente uma compreensão do campo prático.

A literatura *multivocal* é definida como toda a literatura acessível sobre um tópico (OGAWA; MALEN, 1991). Isso inclui, mas não limita-se a: sites, *white papers*, artigos e literatura acadêmica (OGAWA; MALEN, 1991). Recorrer a Revisão *Multivocal* de Literatura (RML) permite que os resultados deem uma visão mais diferenciada do tópico, uma vez que inclui as opiniões de acadêmicos, profissionais, pesquisadores e empresas de desenvolvimento (OGAWA; MALEN, 1991).

A RML deste estudo seguiu as diretrizes fornecidas por (GAROUSI; FELDERER; MÄNTYLÄ, 2019). Para direcionar a revisão, duas perguntas de pesquisa foram formuladas:

- **(P1)** “*Quais métricas são usadas no desenvolvimento ágil de software em larga escala?*”.
- **(P2)** “*Quais são as razões e os resultados do uso de métricas no desenvolvimento ágil em larga escala?*”.

3.4.1 Mecanismos de Busca

Os seguintes mecanismos de busca foram usados para encontrar a literatura relevante:

- bases ACM, IEEEExplore, Springer-Link, Scopus e ScienceDirect para localizar especificamente a literatura acadêmica;
- pesquisa do Google (<https://www.google.com>) para localizar a literatura cinza.

3.4.2 String de Busca

Para encontrar evidências de acordo com as questões de pesquisa, a *string* usada foi a seguinte: (Metric) AND (SAFe OR “Scaled agile framework” OR LeSS OR “Large scale scrum” OR “Scrum-at-Scale” OR “Scrum@Scale” OR DA OR “Disciplined agile” OR Spotify OR Scrum OR “Scaling agile” OR Nexus) AND (“Large-scale” OR Large OR Scaled).

3.4.3 Critérios de Inclusão e Exclusão

Os critérios de inclusão e exclusão (Tabela 3) foram definidos para garantir a inclusão de todas as fontes relevantes e a não inclusão de fontes fora do escopo.

Tabela 3 – Critérios de inclusão e exclusão

Critério	Descrição
Inclusão	(i) Literatura que discute explicitamente o uso de métricas relacionadas a projetos ágeis em larga escala, soluções ou adaptações de métricas nesse cenário; (ii) <i>papers</i> de conferências, revistas e páginas sobre métricas dos <i>frameworks</i> relacionados; (iii) literatura publicada após 2001; (iv) as primeiras 10 páginas válidas da busca do Google.
Exclusão	(i) Literatura inacessível; (ii) resultados que a busca do Google considera similares a outros resultados; (iii) anúncios do Google; (iv) artigos e páginas em um idioma diferente do inglês.

3.4.4 Seleção

Para apoiar o processo de busca e seleção da literatura acadêmica a ferramenta STArt¹ foi utilizada. Ela foi desenvolvida para auxiliar pesquisadores na organização, execução, e documentação de estudos secundários que buscam reunir, avaliar e sintetizar as evidências disponíveis sobre um determinado tema ou questão de pesquisa. As datas de publicação dos artigos foram filtradas nas bases em todas as buscas entre Janeiro de 2001 e Dezembro de 2022, considerando todo período desde a publicação do Manifesto Ágil. Retornaram um total de 3479 artigos na seleção automática. Após a remoção de 393 artigos duplicados, restaram 3086 artigos a serem classificados, nos quais nenhum artigo pertencia a ScienceDirect, pois todos os duplicados eram provenientes dessa base.

Iniciou-se o processo de seleção dos estudos científicos pela análise dos títulos e *abstracts* (etapa 1), em que 3027 artigos foram descartados por não atender os critérios de inclusão. Desses trabalhos, 59 foram encaminhados para a etapa 2, na qual foram lidas a introdução e conclusão dos artigos, logo 34 artigos foram excluídos e após leitura completa de 25 textos restantes 14 trabalhos que correspondiam aos critérios de inclusão seguiram para realização do estudo.

Antes de selecionar resultados de literatura cinza, foram separados previamente os links de páginas oficiais disponíveis sobre métricas dos *frameworks* presentes na *string* para garantir a inclusão das mesmas na pesquisa. Após, consideramos parar a busca inicial nos 10 primeiros links válidos que retornaram no resultado da *string* de busca, com exceção de anúncios do Google, conforme abordagem de parada sugeria em (GAROUSI; FELDERER; MÄNTYLÄ, 2019). Desse resultado, até 2 ciclos de *snowballing* para frente foram aplicados. A busca por literatura cinza foi realizada em Janeiro de 2023. Após a leitura dos títulos e metatextos (etapa 1), 13 resultados iniciais foram selecionados, considerando os dez principais links mais *snowballing* e três páginas oficiais de *frameworks*. Não foram encontrados links para artigos acadêmicos. Durante a leitura do texto completo da fonte, uma página foi adicionada por *snowballing*. Aplicando os critérios de inclusão às 14 fontes, cinco páginas foram elegíveis para o estudo. A Tabela 4 resume e quantifica os resultados.

3.4.5 Avaliação da Qualidade

Além disso, os estudos científicos também foram avaliados e classificados para os critérios de qualidade por Dyba *et al.* (DYBA; DINGSOYR; HANSSSEN, 2007), utilizando uma escala numérica composta pelos seguintes valores: (1) quando o critério era abundante ou explicitamente atendido; (0,5) quando o critério foi atendido de forma parcial ou implícita; (0) quando o critério estava ausente ou não aplicável. Todavia, nenhum estudo

¹ <https://www.lapes.ufscar.br/resources/tools-1/start-1>

Tabela 4 – Resumo dos resultados da busca e seleção para a literatura (RML)

Mecanismo de busca	Resultados iniciais	Títulos, metatexto, <i>abstracts</i> , introdução, conclusão	Texto completo
IEEEExplore	974	5	3
ACM	285	1	0
SpringerLink	184	14	7
Scopus	1643	5	4
ScienceDirect	0	0	0
Google	13	13	5

primário foi excluído por aplicação destes critérios. Os critérios em questão estão listados na Tabela 5.

Tabela 5 – Critérios de qualidade

Perguntas de avaliação de rigor

1. Existe uma definição clara dos objetivos do estudo?
2. Existe uma definição clara das justificativas do estudo?
3. Existe embasamento teórico sobre os temas do estudo?
4. Existe uma definição clara da questão de pesquisa (QP) e/ou da hipótese do estudo?
5. Existe uma descrição adequada do contexto em que a pesquisa foi realizada?
6. São utilizados e descritos métodos apropriados de coleta de dados?
7. Existe uma descrição adequada da amostra utilizada e dos métodos de identificação e recrutamento da amostra?
8. Existe uma descrição adequada dos métodos usados para analisar os dados e métodos apropriados para garantir que a análise dos dados seja fundamentada nos dados?

Perguntas de avaliação de credibilidade

9. É fornecida pelo estudo resposta ou justificativa clara sobre QP/hipótese?
10. O estudo fornece resultados claramente declarados com resultados confiáveis?

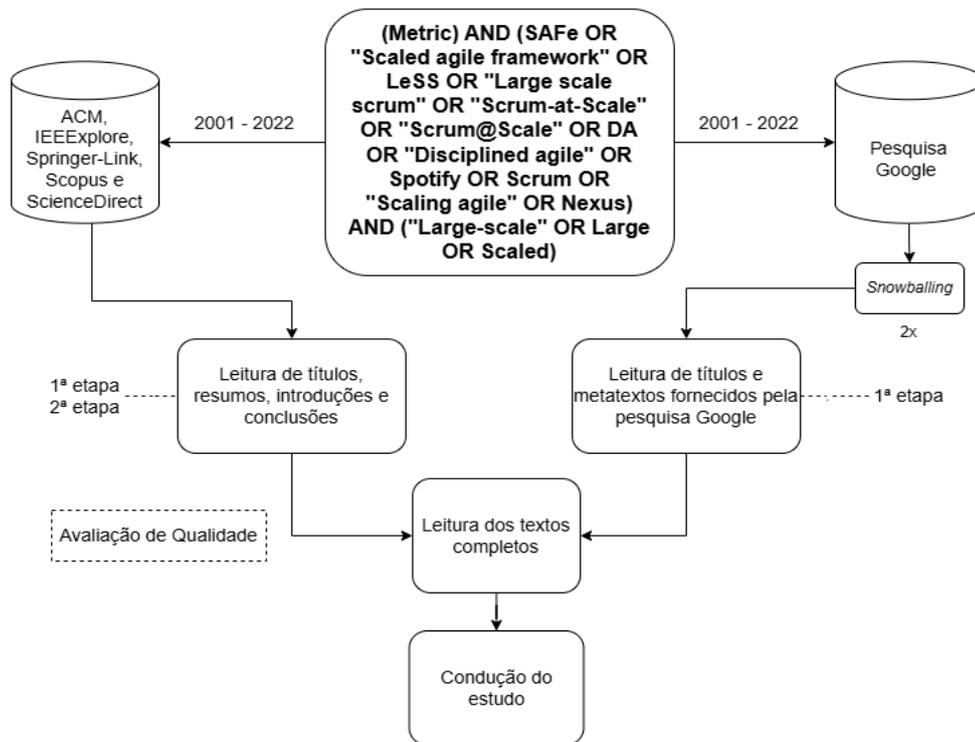
Perguntas de avaliação de relevância

11. O estudo fornece conclusões justificadas?
 12. É fornecido pelo estudo discussão sobre ameaças de validade?
-

3.4.6 Extração de Dados

Durante a condução da revisão, utilizou-se o Google Sheets como ferramenta de apoio na extração dos seguintes dados da literatura quando disponíveis: título, autores, país/local, base de dados, fonte de publicação, ano de publicação, a quais áreas de engenharia de software o estudo pertence, qual tipo, contribuição e métodos de pesquisa foram apresentados, onde a avaliação foi aplicada (indústria e/ou academia), tamanho e setores de atuação das organizações presentes. A visão geral do processo da revisão *multivocal* pode ser visualizada na Figura 2.

Figura 2 – Visão geral do processo (RML)



3.5 Survey

Após a realização da RML, fez-se necessário explorar empiricamente as circunstâncias sob as quais a utilização de métricas no desenvolvimento de software ágil em ambientes escaláveis têm sido eficaz. Para isso, foi desenvolvido um *survey* para investigar como os engenheiros de software têm utilizado as métricas ágeis em projetos de larga escala. A questão de pesquisa principal que o norteou foi: “*Como as métricas são utilizadas no desenvolvimento ágil de software em larga escala?*”. Por conseguinte, foram geradas três questões de pesquisa secundárias. A seguir será apresentado como as subquestões foram elaboradas.

O estudo de Leal *et al.* (LEAL *et al.*, 2022) verificou que as organizações ágeis adotam algumas estratégias majoritárias para selecionar as métricas a serem utilizadas. Em ordem decrescente de prevalência, as estratégias de seleção são: por meio de buscas direcionadas na literatura, por meio da experiência dos membros da equipe e por meio da definição de comitês externos de especialistas voltados aos objetivos da organização. Todavia, é necessário investigar se as organizações de grande porte possuem as mesmas estratégias.

- **Questão secundária 1:** “*Como e para quem as métricas são selecionadas?*”.

Uma outra constatação é a de que o público-alvo das métricas nas organizações ágeis costuma ser em primeiro lugar a própria a equipe ágil, seguida da gestão e do cliente, respectivamente (LEAL *et al.*, 2022). Por outro lado, aqueles que trabalham em grandes projetos podem destinar as medições de outro modo e necessitar de medições extras para impactar áreas diversas, que por sua vez, possuem diferentes objetivos específicos dentro do mesmo programa. Assim, surge mais uma questão secundária:

- **Questão secundária 2:** “*Quais são os objetivos e impactos da utilização das métricas?*”.

As equipes ágeis tem usado métricas para atingir os seguintes objetivos: melhoria de processos ágeis, cumprimento dos protocolos em agilidade, melhoria da qualidade do software durante o desenvolvimento, melhoria da estimativa e planejamento e aumento da produtividade (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021; CHLOROS; GEROGIANNIS; KAKARONTZAS, 2022; LEAL *et al.*, 2022). É razoável esperar que essas metas também estejam presentes em ambientes escaláveis, contudo, é provável que haja outros elementos a serem mensurados não identificados ou documentados. O que leva a última questão secundária:

- **Questão secundária 3:** “*Quais métricas utilizadas no desenvolvimento ágil em larga escala não foram identificadas na literatura?*”.

Buscando responder as questões de pesquisa apresentadas, este trabalho seguiu as diretrizes de Linåker *et al.* (LINÅKER *et al.*, 2015) para a condução de *surveys*. Esta seção apresenta em detalhes como estas diretrizes foram aplicadas.

3.5.1 Público-Alvo e População

O público-alvo do *survey* são pessoas do mercado brasileiro que trabalham dentro de um contexto focado em desenvolvimento ágil de software em larga escala. Tal público-alvo

é caracterizado como dependente em relação à coleta de dados, pois as questões do estudo exigem conhecimento técnico dos respondentes. A população em que se está interessada é composta por desenvolvedores de software, testadores, gerentes, Scrum Masters/*Agile Coaches*, designers e *Product Managers/Owners* em geral.

3.5.2 Amostragem

A amostragem desta pesquisa foi obtida por meio de diferentes canais nos quais foram compartilhados o estudo, sendo eles: LinkedIn (<https://acesse.one/lmkdpost>), grupos de WhatsApp com profissionais do mercado de software e grupos de e-mail das coordenações do curso de Bacharelado em Ciência da Computação e do Programa de Pós-Graduação em Informática Aplicada, ambos da Universidade Federal Rural de Pernambuco. Dessa forma, a amostragem é não probabilística por não poder garantir a sua aleatoriedade e não poder realizar o cálculo amostral com relação ao erro e ao nível de significância, pois não se sabe o tamanho da população atingida. Para este estudo foi obtida uma amostra com 145 respondentes válidos, sendo que o período de aplicação foi de 11/04/2024 a 15/05/2024.

3.5.3 Desenho do Estudo

Optou-se por uma pesquisa exploratória, pois se quer descobrir em algum grau as percepções dos profissionais do mercado a acerca do universo das métricas utilizadas no desenvolvimento de software ágil em larga escala (WOHLIN et al., 2012).

3.5.4 Instrumento e Avaliação

Neste trabalho foi desenvolvido um questionário para Web com vocabulário adaptado para o público-alvo. Este tipo de instrumento de estudo foi escolhido para que seja possível atingir uma quantidade suficiente de indivíduos que estejam inseridos no contexto da pesquisa em diferentes localizações geográficas.

A idealização da pesquisa foi discutida com os orientadores por meio de mensagens de e-mail e pelo Google Docs. Logo após, o questionário foi implementado usando o Google Forms e o formulário de pesquisa pode ser visualizado no Apêndice A. A elaboração do questionário seguiu uma abordagem baseada na literatura (LINÁKER et al., 2015) e na realização de validações com os orientadores, determinando quais informações seriam coletadas, o tipo de questionário, o método de execução, a duração da pesquisa e a sequência das questões da pesquisa. Além disso, planejou-se que a pesquisa pudesse ser respondida em no máximo 10 minutos. A pesquisa contém perguntas dos seguintes tipos: baseadas em escala *Likert*, caixas de seleção, múltipla escolha e abertas.

Para mitigar as ameaças à participação do público, as questões foram projetadas com linguagem simples, objetiva e sem buscar influenciar o respondente. Um teste piloto

foi liberado na data de 03/04/2024 para avaliar o formulário levando em consideração 3 respondentes e avaliando os comentários sobre a estrutura ou semântica do questionário. O profissional 1, coordenador de engenharia de software, levou 5 minutos para responder e reportou apenas que achou a variedade de métricas abrangente. Todavia, essa é a intenção da pesquisa. O profissional 2, Scrum Master, levou pouco menos do que 10 minutos para responder e também reportou que achou a variedade de métricas abrangente. Por fim, o profissional 3, gerente de projeto, levou pouco menos do que 10 minutos para responder. Ele solicitou que fosse realizada uma melhoria na descrição da seção “Dados Demográficos” e mudança de lugar da pergunta sobre o “papel” que o respondente desempenha na empresa. Tais ações de melhoria foram implementadas no instrumento para versão final.

3.5.5 Análise dos Dados

Para análise e extração de dados, foi escolhida como ferramenta facilitadora o RStudio², que é um ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos. A abordagem de análise aplicada aos dados coletados da pesquisa incluiu três etapas principais (LINÅKER et al., 2015), que estão apresentadas a seguir.

3.5.5.1 Validação e limpeza de dados

No total, a pesquisa coletou 164 respostas por meio do questionário. Assim, como primeiro passo, analisou-se o conjunto de dados completo inicial. Foi verificada a existência de 19 respostas em que os participantes sinalizaram que seu time ou projeto era composto por apenas uma equipe, caracterizando “pequena escala”. Portanto, a primeira ação foi removê-los do conjunto. Especificamente, uma atenção foi dada para muitos valores de NA nas respostas de uma determinada questão. Todavia, nesse caso indicou apenas que os participantes não forneceram informações por ser opcional. Além disso, antes de iniciar a análise, realizou-se a exclusão da coluna do conjunto de dados que continha os e-mails de contato dos participantes. Logo, concluiu-se que os 145 pontos de dados resultantes eram relevantes para inclusão no levantamento por conterem as informações adequadas para a análise demográfica e busca das respostas para as questões propostas, sem necessidade de mais interferências, pois todas as repostas estrategicamente obrigatórias no questionário foram fornecidas. Cada resposta forma um ponto de dados válido que consiste em 88 variáveis (mais metadados). O conjunto completo de dados pode ser encontrado em (MENEZES, 2024).

² <https://posit.co/download/rstudio-desktop/>

3.5.5.2 Análise quantitativa

A análise quantitativa empregou como instrumento principal o cálculo de percentuais, técnica estatística descritiva, que serve de base para análises inferenciais de proporções, comparações e tendências. O tratamento dos dados para apoiar esse instrumento consistiu em extrair os recortes adequados do conjunto completo para cada análise, calculando-se as quantificações e percentuais das entidades por meio da codificação. O processo de codificação compreendeu três etapas:

- (i) o autor preparou um modelo e organizou a codificação. Etapa na qual resultou em 16 códigos no total;
- (ii) o autor realizou uma revisão dos códigos para confirmar resultados e promover melhorias;
- (iii) o autor realizou as análises temáticas nos dados codificados.

3.5.5.3 Análise qualitativa

Foram investigadas as respostas obtidas por campos de texto livre que desviaram das opções pré-definidas em alguns poucos pontos de dados. Consistiam em respostas para as questões acerca de estratégias de seleção de métricas ágeis, finalidades pelas quais as empresas as utilizam, quem faz uso dos resultados da medição e que métricas utilizadas no ágil escalonado não foram apresentadas na pesquisa. Os códigos para análise quantitativa também contribuíram para a qualitativa, tornando relativamente fácil identificar as respostas que provocariam análises qualitativas adicionais. Foi necessário desenvolver apenas um novo código para uma análise temática mais robusta das quatorze novas medições apontadas utilizadas no ágil escalonado, distribuídas por sete pontos de dados. Além disso, três temas relevantes de como as métricas ágeis são selecionadas foram nomeados e um tema adicional em relação a quem faz uso dos resultados da medição foi nomeado por outro participante.

3.6 Considerações Finais do Capítulo

Neste capítulo foram apresentadas as etapas de pesquisa, os critérios e procedimentos metodológicos deste estudo. Inicialmente as etapas de pesquisa foram brevemente introduzidas e o fluxo metodológico foi representado visualmente. Logo após, foram classificados e descritos os critérios metodológicos que serviram como princípios norteadores do estudo. Desse modo, foi mostrado que esta pesquisa tem como critérios ser um estudo de natureza aplicada, ao passo que a respeito dos objetivos caracteriza-se como uma pesquisa tanto exploratória como descritiva de abordagem qualitativa e método científico indutivo,

por meio de procedimentos técnicos como pesquisa bibliográfica e documental bem como de levantamento. Por fim foi apresentado com detalhes os protocolos do mapeamento sistemático e da revisão *multivocal* de literatura e ainda, o protocolo do *survey*, que foram as diretrizes metodológicas desta pesquisa. No próximo capítulo, serão apresentados os resultados alcançados por meio da aplicação dessas metodologias.

4 Resultados

Neste capítulo serão descritos os resultados obtidos a partir da realização do Mapeamento Sistemático e da Revisão *Multivocal* de Literatura. Assim como os resultados alcançados a partir da condução do *survey* por meio dos dados obtidos com a aplicação do questionário.

4.1 Resultados do Mapeamento Sistemático de Literatura

Nesta seção constam os resultados que respondem as perguntas de pesquisa do mapeamento sistemático precedente com a visão geral dos estudos identificados sobre o tema e uma breve descrição do conjunto de métricas mapeadas na literatura científica.

4.1.1 Visão Geral dos Estudos Acadêmicos

Os estudos acadêmicos incluídos foram classificados segundo aspectos do tipo de pesquisa com base em (WIERINGA et al., 2006). Quanto às contribuições dos trabalhos, elas foram classificadas segundo (PETERSEN et al., 2008), que incluem teoria, modelo, *framework*, diretriz, lições aprendidas, adendo e ferramenta. A Tabela 6 apresenta uma visão geral dos estudos acadêmicos, tipo de pesquisa, aspectos de contribuição, ano de publicação e métodos utilizados.

4.1.2 Quais são os estudos publicados a respeito da utilização de métricas no desenvolvimento ágil em ambiente escalável?

O trabalho de Greening (GREENING, 2010) reportou a necessidade da definição de formas claras para mensurar as melhorias observadas na implementação organizacional do “Scrum Empresarial”, descrito em seu relato de experiência. Ele promove o pensamento ágil em toda a empresa, incentivando departamentos além dos de engenharia de software a adotarem o Scrum (GREENING, 2010). Tal abordagem possui pontos de semelhança com o *Scrum-at-Scale*, como a técnica de *Scrum-of-Scrums* para projetos maiores (GREENING, 2010). Além disso, algumas situações no “Scrum Empresarial” podem requerer soluções Kanban (GREENING, 2010). Brown (BROWN, 2011) discutiu a escalabilidade de abordagens ágeis (Scrum e XP) por meio de um estudo de caso detalhado, focado em maneiras pelas quais essa agilidade pode ser mensurada e adotada em uma fábrica de software do ponto de vista de uma organização. Tabib (TABIB, 2013) expôs em seu relato de experiência as práticas que ajudaram a lidar com os desafios comuns de grandes projetos empresariais, compartilhou as análises e métricas que permitiram que uma organização

Tabela 6 – Visão geral de estudos acadêmicos

Ano	Tipo de Pesquisa	Contribuição	Método de Estudo	Referência
2010	Experiência	Lições aprendidas	Relatório de experiência	(GREENING, 2010)
2011	Experiência	Lições aprendidas	Estudo de caso	(BROWN, 2011)
2012	Solução	Ferramenta	Pesquisa-ação/revisão de literatura	(STARON; MEDING; PALM, 2012)
2013	Experiência	Ferramenta	Relatório de experiência	(TABIB, 2013)
2013	Solução	Diretriz	Teoria fundamentada/quase-experimento	(HEIDENBERG et al., 2013)
2013	Solução	Ferramenta	Pesquisa-ação	(STARON; MEDING; CAIMAN, 2013)
2014	Solução	Ferramenta	Teoria fundamentada/quase experimento	(KETTUNEN, 2014)
2014	Filosófico	Adendo	Teoria fundamentada	(LAANTI, 2014)
2015	Experiência	Lições aprendidas	Relatório de experiência	(GREENING, 2015)
2015	Experiência	Lições aprendidas	Estudo de caso	(TRIPATHI et al., 2015)
2016	Validação	<i>Framework</i>	Estudo de caso	(GRIMALDI et al., 2016)
2017	Experiência	<i>Framework</i>	Observações	(RAZZAK et al., 2017)
2018	Experiência	Lições aprendidas	Estudo de caso múltiplo/exploratório	(STETTINA; SCHOEMAKER, 2018)
2020	Solução	Ferramenta	Estudo de caso	(THAWABA et al., 2020)
2022	Validação	<i>Framework</i>	Relatório de experiência/experimento	(TESSAROLO et al., 2022)

capacitasse a colaboração entre fusos horários e alinhamento por meio da transparência, fornecendo visão sobre qualidade e risco. Greening (GREENING, 2015) publicou em 2015 um estudo descrevendo um conjunto de métricas para o *framework* Scrum de forma escalável, discutindo como essas métricas forneceram a base para relatórios de *coaching*. Tassarolo *et al.* (TESSAROLO et al., 2022) apresentaram um *framework* baseado em conceitos de *Design Thinking*, *Lean* e Scrum dirigido para o desenvolvimento de tecnologias assistivas avançadas no setor de saúde e bem-estar. Além disso, o trabalho utilizou uma maneira de medir a eficácia percebida desse processo de desenvolvimento em larga escala, tanto pela equipe de desenvolvimento quanto pelas demais partes interessadas do projeto CAPTAIN.

Os autores Staron *et al.* (STARON; MEDING; PALM, 2012) investigaram em seu projeto, por meio de uma pesquisa-ação em uma organização que pratica *Agile* e *Lean* em larga escala, quais seriam os principais indicadores na área de prontidão de lançamento. Em outro estudo, Staron *et al.* (STARON; MEDING; CAIMAN, 2013) também apresentaram e avaliaram industrialmente padrões de medição para monitorar fluxos de trabalho no desenvolvimento ágil de software do tipo *Lean* baseados na qualidade da informação e na ISO/IEC 15939, que descreve como os processos de medição devem ser executados em

organizações de desenvolvimento de software e sistemas. Kettunen (KETTUNEN, 2014) propôs uma abordagem de análise de desempenho para organizações de software ágil de maneira abrangente, onde o objetivo geral é fornecer meios para identificar diferentes equipes de softwares de alto desempenho e, conseqüentemente, medidas práticas para estabelecer e sustentar o desempenho em diferentes transformações organizacionais, inclusive em empresas de software em larga escala. Heidenberg et al. (HEIDENBERG et al., 2013) catalogaram um modelo de métricas desenvolvido iterativamente em uma série de *workshops* com especialistas da indústria e pesquisadores. O modelo de métricas resultante pode ser utilizado em projetos de qualquer tamanho, complexidade e escopo, utilizando métricas que suportam quaisquer valores ágeis e enxutos (HEIDENBERG et al., 2013).

Laanti (LAANTI, 2014) examinou a agilidade em escala do ponto de vista dos aspectos ágeis e apresentou um conjunto de princípios, dentre eles, previsão contínua e planejamento adaptativo e métricas cumulativas baseadas em fluxo. Grimaldi et al. (GRIMALDI et al., 2016) trataram de um caso real e complexo de uma implementação do SAFe em uma grande empresa, enriquecido com métricas que impactam nos prazos de entrega do produto e na redução de custos. Razzak et al. (RAZZAK et al., 2017) analisaram Pequenas e Médias Empresas (PME) em transição para o SAFe estudando e medindo a taxa de adoção das práticas do SAFe a nível de equipe. Thawaba et al. (THAWABA et al., 2020) propuseram um novo mecanismo de medição para o desenvolvimento de software que melhora a avaliação da confiabilidade e a tomada de decisões durante os processos de desenvolvimento ágil de sistemas críticos de segurança utilizando SAFe. Stettina e Schoemaker (STETTINA; SCHOEMAKER, 2018) apresentaram resultados de um estudo de caso múltiplo sobre abordagens de relatórios, artefatos e métricas em grandes organizações aplicando métodos ágeis em seus portfólios de gestão ágil.

Por fim, os autores Tripathi et al. (TRIPATHI et al., 2015) encontraram em seu estudo de caso os desafios para dimensionar Kanban em organizações multi-sites, buscando as possíveis soluções para esses desafios.

4.1.3 Quais são as métricas adotadas no desenvolvimento ágil em ambiente escalável?

À medida que foi implementado o Scrum Empresarial, a princípio, definiu-se a produtividade da empresa como única medida razoável, que pode ser medida como ‘valor presente líquido por esforço’ de trabalho (GREENING, 2010). Essa métrica é uma estimativa de lucratividade, calculada para conhecer o valor atual do retorno de um investimento (GREENING, 2010).

Em (BROWN, 2011), duas dimensões de medição foram usadas para o processo de entrega ágil de uma organização bancária; relacionadas a negócios e a agilidade. Conside-

rando essas dimensões, foi possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessário por meio de dimensões de métricas classificadas como: (1) tempo de redução de ciclo, (2) qualidade, (3) otimização contínua e (4) produtividade (BROWN, 2011). Além disso, para auxiliar no processo de entrega ágil do banco, uma abordagem de pesquisa informal junto a equipe foi aplicada regularmente para medir o ‘pulso de equipe ágil’, constituído por medições da percepção de adoção da prática de desenvolvimento iterativo, planejamento em dois níveis, visão compartilhada, integração contínua e abordagem baseada em equipe (BROWN, 2011).

Métricas utilizadas pela HP facilitaram a habilitação da colaboração entre fusos horários e alinhamento das equipes por meio da transparência, fornecendo *insights* sobre qualidade e risco, foram elas: (1) linhas de código por história de usuário, (2) número de arquivos por história de usuário, (3) linhas de código para refatoração, (4) número de desenvolvedores por recurso, (5) cobertura de testes unitários por história de usuário, (6) taxa de aprovação em testes unitários por história de usuário e (7) número de defeitos por história de usuário (TABIB, 2013).

Ao trabalhar com grandes organizações contendo dezenas ou centenas de equipes, o mesmo autor que relatou o Scrum Empresarial (GREENING, 2010) descreveu várias métricas ágeis abrangentes para uma estrutura Scrum escalável (GREENING, 2015), são elas: (1) velocidade (*Velocity*), que mede o trabalho concluído em uma sprint, (2) desvio de velocidade (*Velocity Deviation*), que mede o desvio padrão percentual da velocidade, (3) horizonte de previsão (*Forecast Horizon*), que mede o ponto no *Product Backlog* pouco antes do primeiro item do *Backlog* do Produto não estimado, (4) *True Sprint Length*, que mede o tempo necessário para a equipe avaliar o valor, a qualidade e a aceitação do mercado de um recurso (5) *Lead Time* (incluindo algumas variantes, mas particularmente o *Lead Time*, que produz previsões do prazo de execução das entregas) e (6) contagem de dependências (*Dependency Count*), que mede potenciais ineficiências introduzidas por uma equipe em seus dependentes.

Sobre a estrutura ágil escalável baseada em Scrum adotada no projeto CAPTAIN (TESSAROLO et al., 2022), realizou-se uma avaliação do *framework* ágil proposto em dois níveis: (1) como uma avaliação pré-pós para comparar a metodologia de desenvolvimento CAPTAIN com outras metodologias existentes experimentadas pela equipe em projetos anteriores, (2) como uma avaliação longitudinal ao longo das iterações ágeis, usando métricas para detectar o moral da equipe e o engajamento e satisfação dos participantes (TESSAROLO et al., 2022). As métricas foram obtidas por meio de questionários aplicados cujas respostas foram classificadas de acordo com uma escala Likert. A avaliação da moral da equipe foi sugerida como mais apropriada do que outras métricas, como a felicidade da equipe (TESSAROLO et al., 2022). A felicidade foi considerada subjetiva e suscetível a variâncias possivelmente devido a outros fatores não relacionados às tarefas da equipe e

do projeto (TESSAROLO et al., 2022).

Um ‘indicador de prontidão para liberação’ importante foi utilizado estrategicamente em uma determinada empresa como indicador de desempenho no controle do desenvolvimento do produto e na melhora do desempenho organizacional (STARON; MEDING; PALM, 2012). O indicador previu em qual semana o lançamento seria possível dada a quantidade de defeitos conhecidos até o momento, quantos defeitos foram removidos em média nas últimas 4 semanas, e quantos defeitos eram esperados para serem descobertos. Trata-se do seguinte indicador:

$$\text{Indicador de pront.} = \frac{\# \text{Defeitos}}{\text{Taxa de remoção de defeitos} - (\text{Taxa de execução de testes} - \text{Taxa de aprovação de testes})}$$

Onde $\# \text{defeitos}$ é a quantidade de defeitos abertos para o produto, $\text{taxa de remoção de defeitos}$ é a quantidade média de defeitos removidos nas últimas 4 semanas, $\text{taxa de execução de testes}$ é a quantidade média de casos de teste executados nos últimos 4 pontos fracos e $\text{taxa de aprovação de testes}$ é a quantidade média de casos de teste passados nos últimos 4 fracos.

Um método baseado na ISO/IEC 15939 foi encontrado em um estudo primário (STARON; MEDING; CAIMAN, 2013). O método é adequado para avaliação da ‘completude de sistemas de medição’ para monitoramento de fluxos de trabalho, visto que são usadas dependências temporais e de processo entre as atividades de um processo (STARON; MEDING; CAIMAN, 2013). Em resumo, os principais princípios do método são: **(1)** criar um modelo de processo do fluxo de trabalho existente, **(2)** vincular as medidas apropriadas e necessárias às atividades neste processo, **(3)** vincular a escala de tempo ao processo e calcular a completude das informações fornecidas pelos sistemas de medição. Para realizar a avaliação de quão completo é o sistema de medição, foi proposta a seguinte fórmula:

$$\text{Indicador de integridade do sistema de med.} = \frac{\# \text{Atividades com medidas ou indicadores}}{\# \text{Atividades no total}} \times 100\%$$

Onde $\# \text{atividades com medidas ou indicadores}$ significa que também são consideradas atividades que possuem apenas medidas (ou seja, medidas de base ou derivadas). Isso se deve ao fato de que às vezes não é possível definir critérios de decisão estáticos para desenvolver um indicador, mas uma medida derivada ou base ainda fornece importantes informações sobre a atividade medida. O sistema de medição que fornece medidas ou indicadores para todas as atividades do processo monitorado é 100% completo. Essa quantidade de indicadores é dada pela soma da quantidade de atividades definidas pela necessidade de informação da parte interessada.

Em busca de modelos abrangentes para realizar transformações sistemáticas de métodos ágeis em empresas de software de larga escala, foi proposta uma análise holística da abordagem de desempenho para empresas de software ágeis baseadas em equipe (KETTUNEN, 2014). Com isso, as organizações de desenvolvimento de software podem avaliar suas diversas equipes para transformações organizacionais e as próprias equipes de software podem utilizá-lo para seu próprio gerenciamento de desempenho para apoiar a obtenção do estado de transformação desejado (KETTUNEN, 2014). As diretrizes apresentadas sugerem a possível adoção de métricas (não especificadas) a serem utilizadas nesse gerenciamento de transformação baseada em equipe, são elas: **(1)** classificar direcionadores de negócios e metas de agilidade, **(2)** traçar perfil de agilidade, **(3)** mapear posicionamento de desempenho da equipe (capacidades-chave), **(4)** criar quadro de orientação, **(5)** utilizar medidores de perfil de desempenho das equipes e **(6)** realizar análise de capacidade.

Um modelo de medições elaborado em *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo, com medições que podem ser usadas em projetos passados e em andamento, processos orientado a planos ou ágeis apresentou, em suma, as seguintes métricas essenciais (HEIDENBERG et al., 2013): **(1)** tempo de resposta da solicitação de atendimento ao cliente, **(2)** tempo de ciclo por recurso, **(3)** funcionalidade por esforço de trabalho, **(4)** valor do negócio por esforço de trabalho, **(5)** pulso de *commit* (mede como a integração contínua está dentro das *sprints* contando a quantidade de *check-ins* diários), **(6)** distribuição de fluxo (mede o fluxo em uma organização), **(7)** número de problemas externos relatados, **(8)** dias abertos com problemas externos relatados (mede a média de dias em que os relatórios de problemas externos não foram resolvidos desde a criação até ser resolvido).

Em (LAANTI, 2014), foram examinadas as características do escalonamento ágil, no qual o *feedback* do usuário é fundamental para o sucesso do produto e, portanto, a *Net Promoter Score* (NPS) também tem importância como métrica cumulativa primária baseada em fluxo que se correlaciona com sucesso empresarial no ágil em larga escala.

Temos na literatura um caso de implementação do *framework* SAFe 3.0 em uma grande empresa, centrado e enriquecido com métricas de eficiência, eficácia e produtividade (GRIMALDI et al., 2016), definidas como: **(1)** a precisão da estimativa delta que considera as horas estimadas para cada *sprint* para entregar o tema comprometido e horas reais gastas que produzem um resultado tangível, **(2)** eficácia interna, razão entre as horas reais e as horas estimadas para cada *sprint* para entregar o recurso comprometido, **(3)** eficiência técnica, razão entre as horas reais e a capacidade ideal de horas que uma equipe pode entregar, **(4)** custo financeiro de horas considerando as horas planejadas, **(5)** custo financeiro de horas considerando as horas reais, **(6)** custo financeiro de desperdício originado da precisão da estimativa delta, **(7)** horas de impedimentos que não produzem resultado tangível e **(8)** porcentagem originada da diferença entre os custos de horas reais

e planejadas.

Em uma PME em transição para SAFe, buscou-se saber a ‘taxa de adoção das práticas do SAFe’ a nível de equipe, identificando quaisquer alterações ao longo do tempo (RAZZAK et al., 2017). Perguntas foram enviadas para participantes, onde cada questão tinha um elemento quantitativo na escala Likert e um elemento qualitativo opcional (comentário) que permitiram que os participantes explicassem sua classificação, se necessário, para medir a frequência de uso da prática de acordo com cada área (*Product Ownership Health, PI/Release Health, Sprint Health, Team Health e Technical Health*) (RAZZAK et al., 2017).

O desenvolvimento ágil de software entrou no campo dos sistemas críticos de segurança, sendo necessário expandir princípios e adaptar novas técnicas de medição (THAWABA et al., 2020). Porém, uma lacuna na melhoria das técnicas de medição e interrupção de falhas nesse contexto ainda precisava ser preenchida com mais métodos de avaliação para obter segurança e confiabilidade do produto. Para isso, foi proposto um novo mecanismo de medição chamado *Package Metrics for Improving Software Development* (PM-ISD) (THAWABA et al., 2020). O PM-ISD auxiliou a equipe de gerenciamento ágil em todas as fases do projeto do *Medical Surgery Department* (MSD) para acompanhar a conclusão das tarefas dentro do SAFe. Os componentes do PM-ISD são o Plano do Projeto (PP), a tabela de testes do projeto e cinco medições (THAWABA et al., 2020). As cinco medições são: métricas de realização e métricas de realização de padrões, métricas de recursos, métricas de previsão para conclusão de tarefas e métricas de previsão para o alcance dos padrões finais (THAWABA et al., 2020). Todavia, o artigo citado não cobriu toda a discussão das cinco medidas, foram discutidas apenas métricas de realização (M1) e recursos (M2).

O Plano do Projeto do MSD contém cinco indicadores: custo, tempo, funções, mão de obra e padrões (THAWABA et al., 2020). O PM-ISD adicionou dois outros indicadores ao PP calculados por duas equações: custo necessário para realizar cada função pela divisão do custo da tarefa pela quantidade de funções da tarefa, tempo necessário para realizar cada função pela divisão do custo da tarefa pelo número de funções da tarefa. A métrica de realização para todas as tarefas (M1-1) exibe em um conjunto de oito equações que detalham teste de tarefas ou subtarefas como: tempo gasto, despesas de custo, funções desempenhadas, cumprimento de padrões e o tempo restante. A métrica de realização para tarefas principais usando subtarefas (M1-2) usa a realização de subtarefas para identificar a realização de suas tarefas principais. M1-2 contém dois procedimentos calculados por meio de cinco equações, (i) usando os resultados da conclusão da subtarefa e (ii) selecionando os dados de teste mais recentes para cada subtarefa. A métrica de recursos (M2-1) é composta por quatro equações e exibe o estado atual de todos os recursos e auxilia a equipe de gerenciamento para revisar o status do recurso após cada tarefa. A métrica de recursos

disponíveis (M2-2) identifica apenas os recursos disponíveis durante a implementação das subtarefas ou tarefas principais. Existem dois procedimentos em M2-2, primeiro usando dados de M2.1 e segundo verificando se o tempo disponível é maior que zero e as demais funções iguais a zero e imprimindo os dados dos recursos disponíveis. A métrica de recursos necessários (M2-3) identifica apenas os recursos necessários durante a implementação das subtarefas ou tarefas principais. Existem dois procedimentos em M2-3, primeiro usando dados de M2.1 e segundo verificando se o tempo disponível é menor que zero e as funções restantes igual ou maior que um e imprimindo os dados de recursos necessários.

Rotinas, métricas e artefatos para manter uma supervisão eficaz na gestão de portfólio ágil foram os resultados preliminares de um estudo de caso longitudinal em grandes organizações europeias onde, a respeito das métricas, 4 classificações foram identificadas para contextos ágeis de larga escala (STETTINA; SCHOEMAKER, 2018): (1) métricas de desempenho (SAFe), que medem a eficiência do trabalho, por exemplo, velocidade e tendências de tempo de ciclo, (2) métricas de qualidade do trabalho (SAFe), (3) métricas de progresso (*Scrum-of-scrums*), que medem até que ponto você está no trabalho, por exemplo, velocidade planejada e realizada e (4) métricas de status (Spotify), que medem o estado atual do trabalho, por exemplo, orçamento e recursos.

Focado nos desafios na fase de dimensionamento do Kanban em uma organização envolvida no desenvolvimento de software multi-site, foram definidos limites de ‘*Work In Progress*’ (WIP) para o *backlog* da área de produto e as equipes de desenvolvimento por acordo mútuo (TRIPATHI et al., 2015). As equipes visualizaram seu fluxo de trabalho em quadros físicos e aplicaram limites de WIP nas tarefas. O limite de WIP é definido para cada *backlog* de produto da área, definido por um acordo comum entre os proprietários do produto da área e o representante da equipe. Por exemplo, um limite de WIP de quatro é anotado no *backlog* de produto da área (ou seja, máximo de quatro histórias de usuário devem estar em cada *backlog* de área em um determinado momento). A estimativa de comum acordo entre as equipes de desenvolvimento deve ser fornecida durante a fase inicial de definição dos limites de WIP. Dessa forma, um limite adequado pode ser aplicado entre as equipes de desenvolvimento da organização. Além disso, uma média ‘*Lead Time*’ também seria uma métrica apropriada para medir o fluxo de trabalho no nível organizacional no Kanban em larga escala (TRIPATHI et al., 2015).

4.1.4 Discussão dos Resultados do Mapeamento

Nesta seção busca-se discutir as informações obtidas a partir dos resultados do MSL a fim de responder suas questões de pesquisa.

Considerando o pequeno número de artigos acadêmicos (15), pode-se inferir que, embora o interesse no desenvolvimento ágil de software esteja crescendo, a comunidade provavelmente ainda carece de conhecimento e compreensão abrangentes deste contexto,

dado o número limitado de estudos. Quanto à variedade de tipos de pesquisa, predominam os relatos de experiência (BROWN, 2011; GREENING, 2010; GREENING, 2015; RAZZAK et al., 2017; STETTINA; SCHOEMAKER, 2018; TABIB, 2013; TRIPATHI et al., 2015), seguidos de propostas de soluções (HEIDENBERG et al., 2013; STARON; MEDING; PALM, 2012; KETTUNEN, 2014; STARON; MEDING; CAIMAN, 2013; THAWABA et al., 2020), com 7 e 5 ocorrências, respectivamente. Foram registrados apenas 2 estudos de validação (GRIMALDI et al., 2016; TESSAROLO et al., 2022) e 1 estudo filosófico (LAANTI, 2014). Isso reforça a necessidade de pesquisas mais relevantes que abordem novas propostas de métricas ainda não implementadas na prática e avaliações extensas do uso de métricas em operações diárias usando agilidade em larga escala.

A presença do Scrum neste recorte da literatura é dominante. Os estudos abordaram métricas que impactaram sobre a necessidade de dimensionar a aplicação em larga escala do Scrum para implantação além dos times de desenvolvimento de software, promovendo o pensamento ágil em toda empresa (GREENING, 2010); como pode ser adotado em uma fábrica de software sob o ponto de vista organizacional (BROWN, 2011); como melhorar a colaboração entre times que trabalham em localizações geográficas e fuso horários distintos (TABIB, 2013); na definição de como as métricas escaláveis são úteis para promover melhorias (GREENING, 2015); como medir e validar a sustentabilidade e eficácia dessa metodologia (TESSAROLO et al., 2022).

A gestão puramente ágil foi referenciada quatro vezes na pesquisa (HEIDENBERG et al., 2013; KETTUNEN, 2014; STARON; MEDING; CAIMAN, 2013; STARON; MEDING; PALM, 2012). Nesses estudos, as métricas dão suporte para importância de melhorar o controle de desenvolvimento do produto e desempenho organizacional (STARON; MEDING; PALM, 2012); monitoramento de fluxos de trabalho (KETTUNEN, 2014); transformações organizacionais (STARON; MEDING; CAIMAN, 2013); e um modelo de medições para transformar qualquer organização que pratique agilidade, independente de escopo e complexidade (HEIDENBERG et al., 2013).

O *framework* SAFe surge na pesquisa cinco vezes sob diferentes aspectos (GRIMALDI et al., 2016; LAANTI, 2014; RAZZAK et al., 2017; STETTINA; SCHOEMAKER, 2018; THAWABA et al., 2020); um deles é o princípio da previsão contínua e planejamento adaptativo, onde foi apontada a importância de métricas de *feedback* (LAANTI, 2014). Foram encontradas novas medições agregadas ao SAFe, de eficiência, eficácia e produtividade, que impactam diretamente nas entregas do produto (GRIMALDI et al., 2016). Um mecanismo eficiente para medir a taxa de adoção de práticas do SAFe em Pequenas e Médias Empresas (PMEs) que desejam adaptar-se e realizar uma transição para usufruir dos benefícios da estrutura escalável (RAZZAK et al., 2017). Para o desenvolvimento ágil de sistemas críticos de segurança em determinadas áreas, como transporte, saúde, nuclear e defesa, as métricas nativas do SAFe não são suficientes, medições adicionais

necessárias foram apresentadas para garantir a obtenção de segurança e confiança nas tomadas de decisão durante o desenvolvimento desses sistemas (THAWABA et al., 2020). Em (STETTINA; SCHOEMAKER, 2018), foi possível observar 10 grandes organizações aplicando métodos ágeis em seus portfólios de gestão ágil no qual foram aplicadas métricas de desempenho, qualidade, progresso e status dentro de metodologias adotadas, como *Scrum-of-scrum*s, SAFe e Spotify.

O Kanban, além de aparecer na pesquisa como auxiliar do Scrum em determinadas situações de nivelamento de fluxo (GREENING, 2015; GREENING, 2010), também foi tema central de um estudo sobre os desafios de dimensioná-lo em organizações distribuídas, onde foi apresentada como solução a limitação do *Work In Progress*, sendo uma medida que deve ser acordada mutuamente entre as equipes para limitar a quantidade de tarefas visíveis em progresso para fins de melhoria do controle do fluxo de trabalho (TRIPATHI et al., 2015).

Nesta pesquisa, foram encontrados indícios de estudos que tratavam de métricas em abordagens próximas do *Scrum-at-Scale* (GREENING, 2015; GREENING, 2010; STETTINA; SCHOEMAKER, 2018) e dentro do modelo Spotify (STETTINA; SCHOEMAKER, 2018), diferentemente dos resultados relatados por Edison et al. (EDISON; WANG; CONBOY, 2021) em sua RSL mencionada. Todavia, de fato não foram encontrados estudos acerca de métricas especificadas para os *frameworks* LeSS e DAD.

4.2 Resultados da Revisão *Multivocal* de Literatura

Nesta seção constam os resultados da revisão de literatura *multivocal* com a apresentação dos contextos organizacionais e o conjunto de métricas identificadas na literatura cinza. Além disso, são apresentados resultados de maneira a responder as perguntas de pesquisa da própria RML, aprofundando-se em uma análise mais descritiva e direcionada da utilização de métricas.

4.2.1 Contexto Organizacional

A Tabela 7 apresenta os contextos organizacionais, na qual estão exibidas informações sobre o número de organizações (No. Org) referenciadas na literatura incluída. O símbolo “+” na coluna de escala de projeto próximo ao termo “larga escala” representa uma quantidade de equipes maior ou possivelmente maior que 9, e o termo “larga escala” representa uma quantidade entre 2 e 9 equipes (DINGSØYR; FÆGRI; ITKONEN, 2014). Na coluna de distribuição, “Localmente” é usado para quando as pessoas estão trabalhando remotamente na mesma cidade, “Nacionalmente” para trabalhando em cidades de um único país, “Continentalmente” para trabalhando em diferentes países dentro do mesmo continente e “Globalmente” para trabalhando em países de diferentes continentes.

Tabela 7 – Contexto organizacional dos estudos

No. Org	Escala	Framework/Método	Distribuição	Indústria	Referência
1	+ Larga escala	Scrum, Kanban	Globalmente	TI	(GREENING, 2010)
1	+ Larga escala	Scrum, XP	Continentalmente	Bancário	(BROWN, 2011)
1	Larga escala	Scrum	Globalmente	TI	(TABIB, 2013)
-	Indefinido	Scrum, Kanban	Indefinido	TI	(GREENING, 2015)
1	Larga escala	Scrum, Lean	Continentalmente	Saúde	(TESSAROLO et al., 2022)
1	+ Larga escala	Ágil, Lean	Nacionalmente	Telecomunicações	(STARON; MEDING; PALM, 2012)
1	+ Larga escala	Ágil, Lean	Nacionalmente	Telecomunicações	(STARON; MEDING; CAIMAN, 2013)
1	Indefinido	Ágil	Nacionalmente	TI	(HEIDENBERG et al., 2013)
-	Indefinido	SAFe	Indefinido	Indefinido	(LAANTI, 2014)
1	Indefinido	SAFe	Indefinido	Telecomunicações	(GRIMALDI et al., 2016)
1	Larga escala	SAFe	Globalmente	TI	(RAZZAK et al., 2017)
1	Indefinido	SAFe	Indefinido	Saúde	(THAWABA et al., 2020)
5	Indefinido	Scrum-of-scrums, SAFe, Spotify	Indefinido	Telecomunicações / Governo	(STETTINA; SCHOEMAKER, 2018)
2	Larga escala	Kanban	Globalmente	Telecomunicações	(TRIPATHI et al., 2015)
1	Larga escala	LeSS	Continentalmente	Financeiro	(KORSON, 2015) ¹

¹Fontes de literatura cinza.

4.2.2 Quais métricas são usadas no desenvolvimento ágil de software em larga escala?

A Tabela 8 apresenta todas as métricas encontradas nesta revisão. Elas estão agrupadas pelos estudos primários ou pela referência das páginas da internet onde foram mencionadas. Também se encontram classificadas de acordo com o *framework*/método ágil em que foram aplicadas ou relacionadas. Métricas com nomes idênticos que aparecem repetidamente foram identificadas em ambos os tipos de literatura e são equivalentes.

Também foram considerados os estágios do Ciclo de Vida de Desenvolvimento de Software (CVDS) para responder a P1. As métricas mapeadas aplicadas no ambiente de escalabilidade ágil foram vinculadas a algumas etapas do CVDS. Para tanto, considerou-se tanto as informações fornecidas nos estudos sobre as etapas do ciclo em que foram aplicadas quanto as descrições e fórmulas de cálculo das próprias métricas. A Tabela 9 apresenta uma visão geral das métricas agrupadas por estágio de desenvolvimento de software no ciclo de vida.

Tabela 8 – Métricas no desenvolvimento ágil de software em larga escala (RML)

Métricas	Framework/Método	Referência
Tempo até a entrega do primeiro incremento; Tempo até o encerramento do projeto; Velocidade; Tarefas bloqueadas; Taxa de defeitos por gravidade; Nível de maturidade do processo; Taxa de adoção de práticas ágeis; Pontos de função por homem-ano; Gráfico burndown; Gráfico de burnup; Pulso de equipe ágil	Scrum, XP	(BROWN, 2011)
Velocidade; Desvio de velocidade; Horizonte de previsão; True Sprint Length; Contagem de dependências; Lead Time	Scrum, Kanban	(GREENING, 2015)
Valor presente líquido por esforço	Scrum, Kanban	(GREENING, 2010)
Delta de precisão de estimativa; Eficácia; Eficiência; Custo por hora virtual planejado; Custo real virtual por hora; Desperdício; Impedimentos; Custo delta	SAFe	(GRIMALDI et al., 2016)
Tempo de resposta da solicitação de atendimento ao cliente; Funcionalidade por esforço de trabalho; Valor comercial por esforço; Pulso de commit; Distribuição de fluxo; Número de problemas externos relatados; Dias abertos com problemas externos relatados; Lead Time; Tempo de ciclo por recurso	Ágil	(HEIDENBERG et al., 2013)
NPS	SAFe	(LAANTI, 2014)
Taxa de adoção das práticas SAFe	SAFe	(RAZZAK et al., 2017)
Indicador de integridade do sistema de medição	Ágil, Lean	(STARON; MEDING; CAIMAN, 2013)
Indicador de prontidão para liberação	Ágil, Lean	(STARON; MEDING; PALM, 2012)
Velocidade planejada; Velocidade	Scrum, Scrum-of-Scrums	(STETTINA; SCHOEMAKER, 2018)
Felicidade da equipe	Scrum, Spotify	(STETTINA; SCHOEMAKER, 2018)
Tempo de ciclo por recurso; Velocidade	Scrum, SAFe	(STETTINA; SCHOEMAKER, 2018)
Linhas de código por história de usuário; Número de arquivos por história de usuário; Linhas de código para refatoração; Número de desenvolvedores por recurso; Cobertura de testes unitários por história de usuário; Taxa de aprovação em testes unitários por história de usuário; Número de defeitos por história de usuário	Scrum	(TABIB, 2013)
Eficácia percebida (Avaliação do trabalho em equipe; Levantamento de requisitos; Planejamento; Qualidade da metodologia; Cultura; Compartilhamento de conhecimento; Percepção geral do processo; Moral da equipe; Engajamento e satisfação dos participantes)	Scrum, Lean	(TESSAROLO et al., 2022)
Custo para cada função; Tempo por função; Tempo restante; Tempo gasto (todas as tarefas); Custo restante; Custo das despesas (todas as tarefas); Funções restantes; Funções concluídas (todas as tarefas); Padrões restantes; Padrões alcançados (todas as tarefas); Tempo gasto (tarefas principais usando a conclusão de subtarefas); Tempo disponível; Custo disponível	SAFe	(THAWABA et al., 2020)
Work In Progress (WIP); Lead Time	Kanban	(TRIPATHI et al., 2015)
Envolvimento dos funcionários; Distribuição de fluxo; Velocidade; Flow time; Carga de fluxo; Eficiência de fluxo; Previsibilidade de fluxo; Frequência de implantação; Tempo de espera pelas alterações; Tempo de restauração do serviço; Taxa de falhas de mudanças; Agilidade empresarial; Competências essenciais	SAFe	(SAFE, 2022) ¹
Taxa de aprovação de testes; Taxa de defeito de entrada para taxa de defeito de saída; Prontidão para liberação	SAFe	(COPADO, 2022) ¹
Valor do negócio pelo esforço; Felicidade da equipe; Pontos de função por homem-ano; Taxa de defeito; Tempo de inatividade do serviço	Scrum-at-scale	(SCRUM@SCALE, 2022) ¹
Índice de previsibilidade; Dívida técnica; Aceleração	Ágil, Lean	(LIYANAGE, 2014) ¹
Gráfico de burnup; WIP	LeSS	(KORSON, 2015) ¹

¹Fontes de literatura cinza.

Tabela 9 – Métricas vinculadas ao Ciclo de Vida de Desenvolvimento de Software

Estágio	Métricas	Referência
Viabilidade do produto	Custo por hora virtual planejado; Delta de precisão de estimativa; Eficácia; Custo real virtual por hora; Desperdício; Eficiência; Impedimentos; Custo delta	(GRIMALDI et al., 2016)
	Custo para cada função; Tempo por função	(THAWABA et al., 2020)
	Planned velocity	(STETTINA; SCHOEMAKER, 2018)
	Número de desenvolvedores por recurso	(TABIB, 2013)
	Previsibilidade de fluxo	(SAFE, 2022)
Desenvolvimento	Velocidade; Tarefas bloqueadas; Gráfico burndown; Gráfico de bur-nup	(BROWN, 2011)
	Linhas de código por história de usuário; Número de arquivos por história de usuário; Linhas de código para refatoração; Cobertura de testes unitários por história de usuário; Taxa de aprovação em testes unitários por história de usuário; Número de defeitos por história de usuário	(TABIB, 2013)
	Velocidade; Contagem de dependências	(GREENING, 2015)
	Funcionalidade por esforço de trabalho; Pulso de commit	(HEIDENBERG et al., 2013)
	Velocidade	(STETTINA; SCHOEMAKER, 2018)
	WIP	(TRIPATHI et al., 2015)
	Velocidade; Carga de fluxo	(SAFE, 2022)
	Gráfico de burnup; WIP	(KORSON, 2015)
Teste	Taxa de defeitos por gravidade	(BROWN, 2011)
	Número de problemas externos relatados; Dias abertos com problemas externos relatados	(HEIDENBERG et al., 2013)
	Número de defeitos por história de usuário	(TABIB, 2013)
	Indicador de prontidão para liberação	(STARON; MEDING; PALM, 2012)
	Tempo restante; Tempo gasto (todas as tarefas); Custo restante; Custo das despesas (todas as tarefas); Funções restantes; Funções concluídas (todas as tarefas); Padrões restantes; Padrões alcançados (todas as tarefas); Tempo gasto (tarefas principais usando a conclusão de subtarefas); Tempo disponível; Custo disponível	(THAWABA et al., 2020)
	Taxa de defeito	(SCRUM@SCALE, 2022)
Implantação	Taxa de aprovação de testes; Taxa de defeito de entrada para taxa de defeito de saída; Prontidão para liberação	(COPADO, 2022)
	Tempo de resposta da solicitação de atendimento ao cliente	(HEIDENBERG et al., 2013)
	NPS	(LAANTI, 2014)
	Frequência de implantação; Tempo de espera pelas alterações; Tempo de restauração do serviço	(SAFE, 2022)
Manutenção	Tempo de inatividade do serviço	(SCRUM@SCALE, 2022)
	Valor presente líquido por esforço	(GREENING, 2010)
	Tempo até a entrega do primeiro incremento; Tempo até o encerramento do projeto; Nível de maturidade do processo; Taxa de adoção de práticas ágeis; Pulso de equipe ágil ; Pontos de função por homem-ano	(BROWN, 2011)
	True Sprint Length; Lead Time; Desvio de velocidade; Horizonte de previsão	(GREENING, 2015)
	Eficácia percebida	(TESSAROLO et al., 2022)
	Indicador de integridade do sistema de medição	(STARON; MEDING; CAIMAN, 2013)
	Lead Time; Tempo de ciclo por recurso; Valor comercial por esforço; Distribuição de fluxo	(HEIDENBERG et al., 2013)
	Taxa de adoção das práticas SAFe	(RAZZAK et al., 2017)
	Tempo de ciclo por recurso; Felicidade da equipe	(STETTINA; SCHOEMAKER, 2018)
	Envolvimento dos funcionários; Distribuição de fluxo; Flow time; Eficiência de fluxo; Taxa de falhas de mudanças; Agilidade empresarial; Competências essenciais	(SAFE, 2022)
Valor comercial por esforço; Felicidade da equipe; Pontos de função por homem-ano	(SCRUM@SCALE, 2022)	
Índice de previsibilidade; Dívida técnica; Aceleração	(LIYANAGE, 2014)	

4.2.3 Quais são as razões e os resultados do uso de métricas no desenvolvimento ágil em larga escala?

Adotou-se a abordagem *Goal Question Metric* (GQM) para responder a P2, visando identificar padrões nas transformações causadas pelas métricas descobertas. A abordagem GQM fornece ferramentas para definir e interpretar medições de software orientadas por metas organizacionais, especificando um objetivo mensurável refinado em um conjunto de questões quantificáveis relacionadas aos objetivos (BASILI, 1992). Conforme apresentado pelo autor, essas questões, por sua vez, ajudam a definir um conjunto de métricas e dados para coleta. Em relação aos objetivos dentro da abordagem GQM, observa-se que grandes organizações ágeis buscam melhorar e medir pelo menos um dos seguintes fatores gerais: capacidade de resposta ao cliente, entrega de valor, produtividade, sustentabilidade de práticas e qualidade (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021; CHLOROS; GEROGIANNIS; KAKARONTZAS, 2022; LEAL et al., 2022; SCRUM@SCALE, 2022). Portanto, o modelo de métrica alvo tem os seguintes objetivos: melhorar a capacidade de resposta do cliente, a entrega de valor comercial, produtividade ponta a ponta, sustentabilidade das práticas tanto do ponto de vista organizacional quanto do cliente e qualidade do produto no desenvolvimento de software. Com base nesses objetivos, foram propostas cinco questões: **Q1:** Somos mais responsivos aos clientes na forma ágil de trabalhar? **Q2:** Entregamos mais valor aos clientes com a forma ágil de trabalhar? **Q3:** Somos mais produtivos trabalhando ágil? **Q4:** Temos melhores práticas de sustentabilidade no ambiente organizacional? **Q5:** Temos melhor qualidade de produto?

A etapa final da modelagem GQM foi determinar as 80 métricas da literatura *multivocal* (Tabela 8) a serem utilizadas de acordo com as questões levantadas. Algumas métricas relacionadas a cada questão apresentada Q1, Q2, Q3, Q4 e Q5 são apresentadas nas Tabelas 10, 11, 12, 13 e 14 respectivamente. Cada tabela lista o nome da métrica, descrições (quando necessário) e as fórmulas ou dados que as quantificam (Veja todas as métricas da modelagem e informações completas em inglês em (MENEZES; MARINHO; SAMPAIO, 2024a)).

É importante ressaltar que definir produtividade para a construção da Tabela 11, bem como no desenvolvimento de software, não foi uma atividade trivial. Inicialmente, o conceito de produtividade no desenvolvimento de software era medido em termos de ‘linhas de código por homem-hora’, mas hoje em dia uma definição mais abrangente que está sendo considerada é a relação entre os valores funcionais do software e o esforço, trabalho e custos envolvidos no desenvolvimento de um sistema (ROGER; BRUCE, 2015).

Nesta RML, entre os estudos acadêmicos, há relatos de grandes organizações que escalaram o Scrum para ir além das equipes de desenvolvimento de software, promovendo o pensamento ágil em toda a organização (GREENING, 2010), bem como para equipes de desenvolvimento trabalhando juntas (BROWN, 2011; GREENING, 2015), ou situados em

Tabela 10 – Somos mais responsivos aos clientes na forma ágil de trabalhar?

Nome	Descrição	Fórmula de Cálculo
Tempo até a entrega do primeiro incremento	Tempo gasto desde o início do projeto até a entrega do primeiro incremento	<i>Data de entrega do primeiro incremento - Data de início do projeto</i>
Tempo até o encerramento do projeto True Sprint Length	Tempo gasto desde o início do projeto até o encerramento do projeto	<i>Data de encerramento do projeto - Data de início do projeto</i> <i>Data real de entrega do incremento quando maior que a data de término da sprint definida anteriormente - Data de início da sprint</i>
Horizonte de previsão	Soma dos pontos de estimação, h , do topo do backlog do produto até um ponto dado	Se conhecermos a velocidade da equipe estimada $\mu(V)$ e o desvio padrão $\sigma(V)$, podemos expressar o horizonte de previsão em <i>sprints</i> como $h/\mu(V) \pm \sigma(V) * h/\mu(V)$ para 68% de confiança ou $h/\mu(V) \pm 2\sigma(V) * h/\mu(V)$ para 95% de confiança. Se conhecermos o comprimento do <i>sprint</i> l , podemos expressar o horizonte de previsão em unidades de tempo como $hl/\mu(V) \pm l\sigma(V) * h/\mu(V)$
Lead Time		<i>Data de conclusão de um processo - Data de início de um processo</i>
Tempo de resposta da solicitação de atendimento ao cliente	Mede o retorno de solicitações de serviço ao cliente	<i>Data de resolução da solicitação de serviço ao cliente - Data de criação da solicitação de serviço ao cliente</i>
Tempo de ciclo por recurso	Mede o tempo de ciclo para recursos selecionados para desenvolvimento	<i>Data de conclusão da prontidão do recurso - Data em que o recurso foi adicionado ao backlog</i>
Distribuição de fluxo	Mede a quantidade de cada tipo de trabalho no sistema ao longo do tempo	Uma medida simples é contar o número de cada tipo de item de trabalho em qualquer momento. Uma medida mais precisa pode considerar o tamanho de cada item de trabalho
Indicador de prontidão para liberação	Prevê em que semana o lançamento do produto seria possível, dado o número de defeitos conhecidos até o momento, quantos defeitos foram removidos em média nas últimas 4 semanas e quantos defeitos eram esperados serem descobertos	$\#Defeitos / Taxa de remoção de defeitos - (Taxa de execução de testes - Taxa de aprovação de testes)$. Onde $\#defeitos$ é o número de defeitos abertos para o produto, <i>taxa de remoção de defeitos</i> é a quantidade média de defeitos removidos nas últimas 4 semanas, <i>taxa de execução de testes</i> é o número médio de casos de teste executados nas últimas 4 semanas e <i>taxa de aprovação de testes</i> é o número médio de casos de teste aprovados nas últimas 4 semanas
Tempo por função		<i>Tempo de duração da tarefa / Quantidade de funções da tarefa</i>
Tempo restante	Tempo restante para concluir a tarefa	<i>Tempo de duração da tarefa / (Data do teste da tarefa - Data de início da tarefa)</i>
Tempo gasto (todas as tarefas)	Tempo gasto durante a implementação da tarefa	$((Data do teste da tarefa - Data de início da tarefa) / Tempo de duração da tarefa) * 100$
Tempo disponível	Tempo disponível para subtarefas ou tarefas principais	<i>Tempo gasto em subtarefas - (Data do teste da tarefa - Data de início da tarefa)</i>
Flow Time	Tempo total decorrido para todas as etapas de um fluxo de trabalho	Medido pelo período médio de tempo necessário para concluir um certo tipo de item de trabalho
Tempo de espera pelas alterações		Quantidade de tempo de espera para fazer uma mudança

diferentes localizações geográficas (TABIB, 2013; TESSAROLO et al., 2022), utilizando métricas para enfrentar desafios e promover melhorias. A gestão puramente ágil foi referenciada na pesquisa em estudos que abordaram medições mais focadas em aspectos de transformações organizacionais (HEIDENBERG et al., 2013), melhoria do controle do fluxo de trabalho (STARON; MEDING; CAIMAN, 2013) e metas de desempenho da organização

Tabela 11 – Entregamos mais valor aos clientes com a forma ágil de trabalhar?

Nome	Descrição	Fórmula de Cálculo
Valor presente líquido por esforço	Valor de lucratividade por esforço	$Valor\ Presente\ Líquido / Esforço$
Funcionalidade por esforço de trabalho	Quanta funcionalidade pode ser entregue em relação a um determinado esforço de trabalho	$Pontos\ de\ teste / Horas\ por\ pessoa$
Valor comercial por esforço NPS	Aqui é indicado como lançamentos principais mais frequentes em relação ao esforço de trabalho <i>Feedback</i> do usuário/cliente	$Número\ de\ lançamentos\ principais\ em\ um\ ano / Horas\ por\ pessoa$
Eficiência de fluxo	Quanto do tempo total de fluxo é gasto em atividades de trabalho que agregam valor versus esperando entre etapas	Calculado perguntando aos clientes se eles recomendariam o produto aos colegas em uma escala de 0 a 10. As respostas são categorizadas como detratores (0-6), neutros (7-8) e promotores (9-10). Por fim, a porcentagem total de detratores é subtraída da porcentagem de promotores para determinar o NPS. $Tempo\ total\ de\ atividade / Tempo\ de\ fluxo$
Previsibilidade de fluxo	Mede quão bem as equipes, os <i>Agile Release Trains</i> (ARTs) e os <i>Solution Trains</i> podem planejar e atingir seus objetivos do <i>Program Increment</i> (PI)	Razão entre o valor de negócio planejado alcançado e o valor de negócio real entregue em um PI

(STARON; MEDING; PALM, 2012). O SAFe aparece em pesquisas com propostas de enriquecimento de suas métricas nativas para atender demandas de velocidade nas entregas medindo recursos e custos no processo (GRIMALDI et al., 2016) e melhorando a confiança no desenvolvimento de sistemas críticos de segurança (THAWABA et al., 2020). Também foi encontrado um mecanismo eficiente para medir a taxa de adoção de práticas SAFe em PMEs que desejam se adaptar a estruturas escaláveis (RAZZAK et al., 2017). No (STETTINA; SCHOEMAKER, 2018) foram citadas métricas de desempenho, qualidade, progresso e status em metodologias como *Scrum-of-scrums*, *SAFe* e *Spotify*. A importância do *Kanban* no controle e escalonamento de fluxos de trabalho entre equipes utilizando métricas como limites de trabalho e tempos de espera foi destacada (GREENING, 2010; GREENING, 2015; TRIPATHI et al., 2015).

Na literatura cinza, a maioria dos achados está relacionada ao SAFe, que possui um sistema robusto de métricas sugeridas (SAFE, 2022). Além das métricas nativas do SAFe, encontramos uma proposta para adicionar medição de qualidade ao seu monitoramento de fluxo em um ambiente DevOps (COPADO, 2022), perspectiva atualmente não abordada em (SAFE, 2022). O *Scrum-at-scale* não se compromete oficialmente com um sistema de métricas predefinido, mas enfatiza claramente a necessidade de monitorizar a produtividade, a qualidade, a entrega de valor e a sustentabilidade, influenciando diretamente a tomada de decisões e promovendo a transparência (SCRUM@SCALE, 2022).

A sinergia entre Ágil e *Lean* é vital para melhorar a agilidade em sistemas ágeis de grande escala (LIYANAGE, 2014). Em (LIYANAGE, 2014), foram apresentadas métricas para medir a agilidade dos negócios com base na previsibilidade dos riscos/custos de

Tabela 12 – Somos mais produtivos trabalhando ágil?

Nome	Descrição	Fórmula de Cálculo
Velocidade		Quantidade de pontos de história completados para itens de trabalho de um tipo ao longo de um período de tempo
Tarefas bloqueadas		Número de tarefas bloqueadas durante um período de tempo específico
Pontos de função por homem-ano		Quantidade de trabalho realizado por um indivíduo ao longo do ano
Gráfico burn-down	Mede o progresso da <i>sprint</i> e fornece indicadores do processo de trabalho da equipe	Marca os dias da <i>sprint</i> no eixo horizontal e os pontos planejados para compor a <i>sprint</i> no eixo vertical, começando dos pontos máximos da <i>sprint</i> (velocidade da equipe) até zero
Gráfico de burnup	Mede o progresso com base nas horas ou pontos restantes de cima para baixo. Mede o progresso do lançamento e fornece indicadores do processo de trabalho da equipe	Marca os dias da <i>sprint</i> no eixo horizontal e os pontos planejados para compor a <i>sprint</i> no eixo vertical, começando dos pontos máximos da <i>sprint</i> (velocidade da equipe) até zero
Desvio de velocidade	Mede a estabilidade da velocidade	$\sigma(V)/\mu(V)$. Onde $\mu(V)$ é a velocidade esperada (velocidade média com base em um certo número de <i>sprints</i> anteriores) e $\sigma(V)$ é o desvio padrão da velocidade
Contagem de dependências	Dependência entre equipes onde <i>bugs</i> e atrasos afetam dependentes imediatos	Número de dependentes imediatos
Eficiência	Indica o quão bem usamos a equipe em comparação com a capacidade máxima	AH / C . Onde AH representa as horas reais gastas produzindo resultados tangíveis e C é o número ideal de horas que uma equipe pode entregar, dependendo do tamanho da equipe, número de equipes, dias não trabalhados e dias gastos em cerimônias. $C = S*(DS - (NWD + (DS*KO/20))) * TS*DH$, onde DS é a duração da <i>sprint</i> , NWD representa dias de folga, KO significa dias de transferência de conhecimento/planejamento/estimativa, TS é o tamanho da equipe e DH é a capacidade diária de desenvolvimento em horas
Impedimentos	Qualquer ‘perda de tempo’ devido a um defeito ou obstáculo que prejudique a produtividade	Número de horas que não produzem resultados tangíveis
Pulso de commit	Mede como a integração contínua está dentro das <i>sprints</i>	Número de dias entre <i>commits</i>
Velocidade planejada		Quantidade de trabalho (pontos de história) que uma equipe espera completar durante um <i>sprint</i>
Custo para cada função		$Custo da tarefa / Número de funções da tarefa$
Custo restante	Mede o custo restante para a conclusão da tarefa	$Custo da tarefa - Custo do gasto$

desenvolvimento, mantendo a confiabilidade do produto e a adaptabilidade às condições que impactam a entrega de valor e associadas diretamente aos princípios de gestão *Lean* e *Ágil*. O LeSS surgiu na pesquisa por meio de um resumo de estudo de caso em que utilizar o gráfico de *burnup* como indicador de aumento de escopo e estabelecer um limite de WIP para realização de correções foram essenciais para o andamento e conclusão do projeto (KORSON, 2015). Nenhuma evidência foi encontrada em relação às métricas usadas para implementar a estrutura ágil escalável *Disciplined Agile*.

Tabela 13 – Temos melhores práticas de sustentabilidade no ambiente organizacional?

Nome	Descrição	Fórmula de Cálculo
Pulso de equipe ágil	Envolve avaliações casuais regulares da equipe para entender suas visões sobre a adoção de práticas de desenvolvimento iterativo, planejamento em dois níveis, visão compartilhada, integração contínua e uma abordagem baseada em equipe	Coletado através de pesquisas informais
Taxa de adoção das práticas SAFe	Pesquisa de autoavaliação enviada aos participantes. Cada pergunta inclui uma classificação numérica (escala <i>Likert</i>) e uma seção de comentários opcional	A escala <i>Likert</i> tem seis opções de resposta (de ‘nunca’ a ‘sempre’) para medir a frequência de uso da prática em áreas como saúde do proprietário do produto, saúde do PI/liberação, saúde do <i>sprint</i> , saúde da equipe e saúde técnica
Indicador de integridade do sistema de medição	Avaliação da completude do sistema de medição para monitoramento do fluxo de trabalho, já que dependências temporais e de processo são utilizadas entre atividades do processo	$(\#Atividades\ com\ medidas\ ou\ indicadores / \#Atividades\ no\ total) * 100\%$
Envolvimento dos funcionários	Mede o quanto os indivíduos se sentem motivados e engajados em apoiar os objetivos e valores da organização	Existem diversos métodos para medir o engajamento dos funcionários, e cada organização deve escolher o que melhor se adequa a ela. Alguns utilizam uma pesquisa anual, enquanto outros confiam em um <i>Employee Net Promoter Score</i> (eNPS), perguntando aos funcionários o quão provavelmente eles recomendariam seu empregador em uma escala de 10 pontos
Agilidade empresarial	Avaliação em alto nível que resume o quão ágil o negócio está em determinado momento	Dois métodos de avaliação estão disponíveis: (i) Os participantes preenchem avaliações de forma independente, seguidas por discussão em grupo e análise; (ii) Todos os participantes discutem coletivamente e concordam com pontuações (de 1 a 5) para cada afirmação. O relatório de avaliação inclui visualizações acompanhando o progresso nas sete competências principais do SAFe

4.2.4 Discussão dos Resultados da *Multivocal*

Nesta seção busca-se sintetizar as informações encontradas na literatura cinza como resposta para as questões de pesquisa da RML.

Ao separar as métricas por estágios do ciclo de vida, conclui-se que 33 métricas podem ser utilizadas durante a etapa de manutenção, definida como a última etapa do ciclo de vida. 20 métricas podem ser usadas na fase de testes. As quantidades de 20, 13 e 6 métricas podem ser aplicadas nas etapas de desenvolvimento, viabilidade do produto e implantação. Não foram encontradas métricas aplicadas ou definidas especificamente para as etapas requisitos e design.

Para cada questão do modelo métrico resultante, existe um conjunto de métricas associadas do total de 80 identificadas nesta RML. Para Q1, relacionado ao monitoramento da capacidade de resposta do cliente, foram identificadas 18 métricas. As métricas de valor entregue aos clientes (Q2) têm o menor conjunto, com apenas 6 métricas. Este número pode indicar a necessidade de investigar a viabilidade de alternativas adicionais de medição de entrega de valor. O conjunto maior (Q3), com 21 métricas, está relacionado ao

Tabela 14 – Temos melhor qualidade de produto?

Nome	Descrição	Fórmula de Cálculo
Taxa de defeitos por gravidade	Mede a taxa de defeitos de software	<i>Número de defeitos (gravidade 1 e 2) em produção / 100</i>
Número de problemas externos relatados	Mede o número total de relatórios de problemas externos durante um período de tempo específico	Número de relatórios de problemas externos originários de uma versão específica
Dias abertos com problemas externos relatados	Mede o número médio de dias que os relatórios de problemas externos permaneceram não resolvidos da criação à resolução	<i>Data de resolução dos relatórios de problemas - Data de criação dos relatórios de problemas</i>
Linhas de código por história de usuário		<i>Linhas de código / História do usuário</i>
Número de arquivos por história de usuário		<i>Número de arquivos / História do usuário</i>
Linhas de código para refatoração		Quantidade de linhas de código para refatoração
Número de desenvolvedores por recurso		<i>Número de desenvolvedores / Recurso</i>
Cobertura de testes unitários por história de usuário		<i>Cobertura de testes unitários / História do usuário</i>
Taxa de aprovação em testes unitários por história de usuário		<i>Taxa de aprovação dos testes unitários / História do usuário</i>
Número de defeitos por história de usuário		<i>Número de defeitos / História do usuário</i>
Taxa de falhas de mudanças	Mede a eficiência do fluxo	Percentual de mudanças que exigem correção após a entrada em produção
Tempo de restauração do serviço	Métrica de tempo de fluxo	Quantidade de tempo de espera para restauração do serviço
Taxa de aprovação de testes	Mede tendências na taxa de aprovação para conjuntos de testes automatizados	Percentual da taxa de aprovação para conjunto de testes automatizados
Taxa de defeito de entrada para taxa de defeito de saída	Determina se a equipe pode corrigir mais bugs do que apenas aqueles descobertos durante os testes	<i>Taxa de defeitos de entrada / Taxa de defeitos de saída</i>
Taxa de defeito		<i>Número de defeitos / 100</i>
Tempo de inatividade do serviço		Quantidade de tempo que um serviço específico ficou fora do ar

desempenho da equipe. Existem 18 métricas que avaliam a sustentabilidade nas práticas (Q4) e 17 que medem a qualidade do produto (Q5).

A maioria das fontes literárias aborda métricas para apoiar grandes organizações que escalaram estrategicamente o Scrum para melhorar resultados ou mitigar desafios em situações adversas (BROWN, 2011; GREENING, 2010; GREENING, 2015; TABIB, 2013; TESSAROLO et al., 2022). A gestão ágil pura foi referenciada na pesquisa, juntamente com métricas que apoiaram o controle de fluxo (STARON; MEDING; CAIMAN, 2013), as transformações organizacionais (HEIDENBERG et al., 2013) e a busca pela agilidade dos negócios (STARON; MEDING; PALM, 2012; LIYANAGE, 2014). Foi descoberto que o SAFe possui métricas citadas em artigos como propostas para enriquecer suas métricas nativas no monitoramento de recursos e custos (GRIMALDI et al., 2016), atender avaliações de confiabilidade (THAWABA et al., 2020) e demandas de qualidade (COPADO, 2022) e aplicar transformações de modelo organizacional (RAZZAK et al., 2017). Também foi notada a relevância do *Kanban* no controle e escalonamento de fluxos de trabalho

entre equipes utilizando métricas como limites e tempos de espera (GREENING, 2010; GREENING, 2015; TRIPATHI et al., 2015). A estrutura *Scrum-at-Scale* sugere medidas para facilitar a tomada de decisões e promover a transparência (SCRUM@SCALE, 2022).

4.3 Resultados do *Survey*

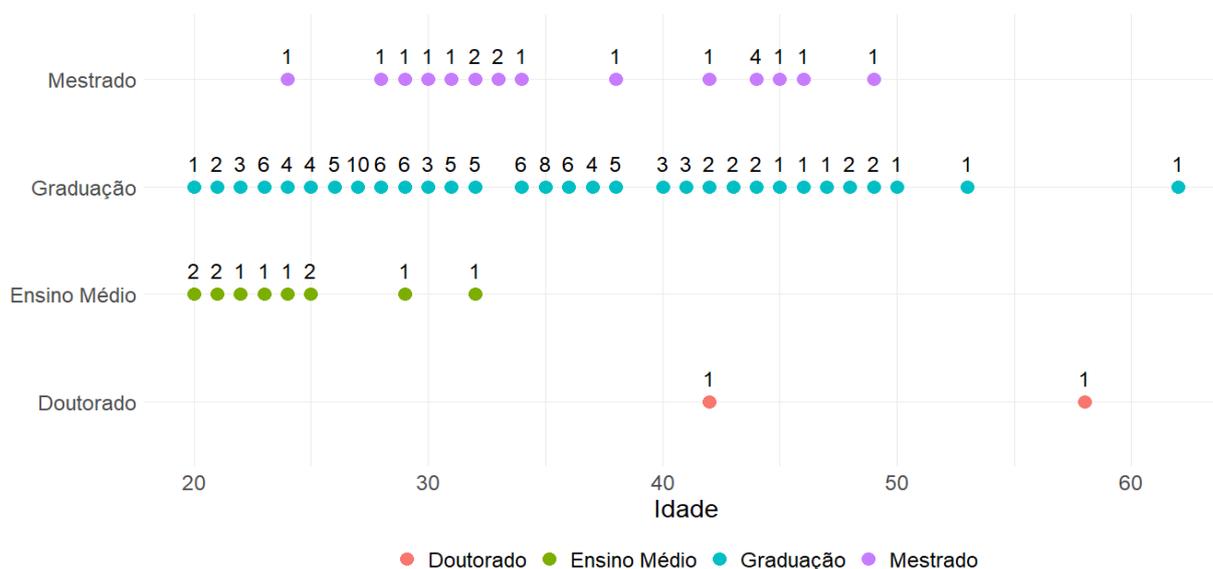
Nesta seção são apresentados os dados coletados através de questionário aplicado durante os meses de Abril e Maio/2024 com uma amostra de 145 participantes da população atuante no mercado de software brasileiro que trabalham em equipes ou projetos inseridos em contexto ágil escalado. Também destaca-se os resultados para as perguntas de pesquisa do *survey* conduzido em função de solidificar o conhecimento adquirido a partir do mapeamento e revisão *multivocal* que o antecedeu.

4.3.1 Dados Demográficos

A maioria dos respondentes (71 - 48.96%) possui entre 30 e 45 anos, enquanto 60 (41.38%) possui menos de 30 anos, 11 (7.59%) entre 46 e 60 anos, e apenas um (0.69%) possui mais de 60 anos. Isto significa que aproximadamente 90.34% dos respondentes possui até 45 anos, o que representa uma fatia considerável da amostra em relação à faixa etária. 2 participantes forneceram informações incorretas no campo da idade. Com relação ao nível de escolaridade, a grande maioria desses profissionais possui graduação (111 - 76.55%), mestrado ou pós-graduação vêm em seguida (20) com 13.79% e posteriormente os que possuem até o ensino médio (12) representam 8.27%. Apenas dois (1.38%) possuem doutorado. A Figura 3 mostra a relação entre a idade e escolaridade concluída de cada participante. Pode-se notar uma considerável quantidade de profissionais experientes que possui um grau de instrução como mestrado ou pós-graduação. Uma pequena parcela deles, mais jovem, possui apenas o ensino médio.

A maior parte dos respondentes (47 - 32.41%) trabalha como Scrum Masters/*Agile Coaches*, seguidos por desenvolvedores ou função equivalente (43 - 29.65%). Líderes de projeto e analistas/engenheiros de requisitos compõem a amostra com 12 (8.27%) e 11 (7.59%), respectivamente. Logo após constam 10 *Product Managers/Owners* (6.9%), 6 gerentes do projeto (4.14%) e 4 testadores ou equivalente (2.6%). As demais funções específicas correspondem a 12 respondentes (8.27%). Os profissionais possuem em sua maioria (61 - 42.07%) entre 1 e 5 anos de experiência, seguidos do grupo dos com mais de 15 anos (35 - 24.14%) e dos que possuem pelo menos entre 6 e 10 anos de experiência (22 - 15.17%). Profissionais com experiência entre 10 e 15 anos (20) representam 13.79% da amostra. Portanto, os que alegaram possuir mais de 6 anos de experiência totalizam 53.1%, o maior percentual. Os que possuem menos de 1 ano de experiência (7) representam apenas 4.83% da amostra.

Figura 3 – Relação entre idade e escolaridade concluída



Considerando o entendimento dos respondentes a respeito do contexto das organizações, a maioria delas são grandes empresas que possuem 100 ou mais colaboradores (136 - 93.79%), seguidas por empresas médias que possuem entre 50 e 99 (6 - 4.14%) e empresas pequenas que possuem entre 10 e 49 (3 - 2.07%). Em relação aos setores nos quais pertencem as empresas, 50 (34.48%) respostas indicam consultorias, 37 são do setor bancário (25.52%), 14 representam fábricas de software (9.65%), 9 serviços financeiros (6.21%), 6 varejo (4.14%), 5 telecomunicações (3.45%), 4 são do setor público/governamental (2.76%) e 4 de saúde (2.76%). As organizações pertencentes a setores mais nichados (16) representam em conjunto 11.03% da amostra.

A Figura 4 apresenta a distribuição de respondentes em relação ao tamanho das empresas, cargos e tempo de experiência desses. É possível observar que a maior diversidade de cargos da amostra está no grupo dos indivíduos que possuem entre 1 e 5 anos de experiência. Também pode-se destacar que a maior parte das pessoas mais experientes trabalham em grandes empresas.

Para categorização da escala de projetos ágeis foi coletada a quantidade de equipes pertencentes ao projeto ou time no qual cada profissional atua, a maioria das iniciativas possui entre 2 e 9 equipes (78 - 53.79%), seguidas das que possuem uma quantidade maior do que 9 equipes (67 - 46.21%), conforme a Figura 5.

Cada projeto ou time é composto por equipes cujas pessoas encontram-se trabalhando distribuídas geograficamente das seguintes maneiras (Figura 6): nacionalmente distribuídas (101 - 69.65%) foram maioria, em seguida estão as iniciativas cujas pessoas trabalham globalmente distribuídas (21 - 14.48%) e as que trabalham localmente distribuídas (12 - 8.27%). Pessoas trabalhando distribuídas continentalmente (11) representam

Figura 4 – Distribuição de respondentes por cargo, anos de experiência e tamanho da empresa

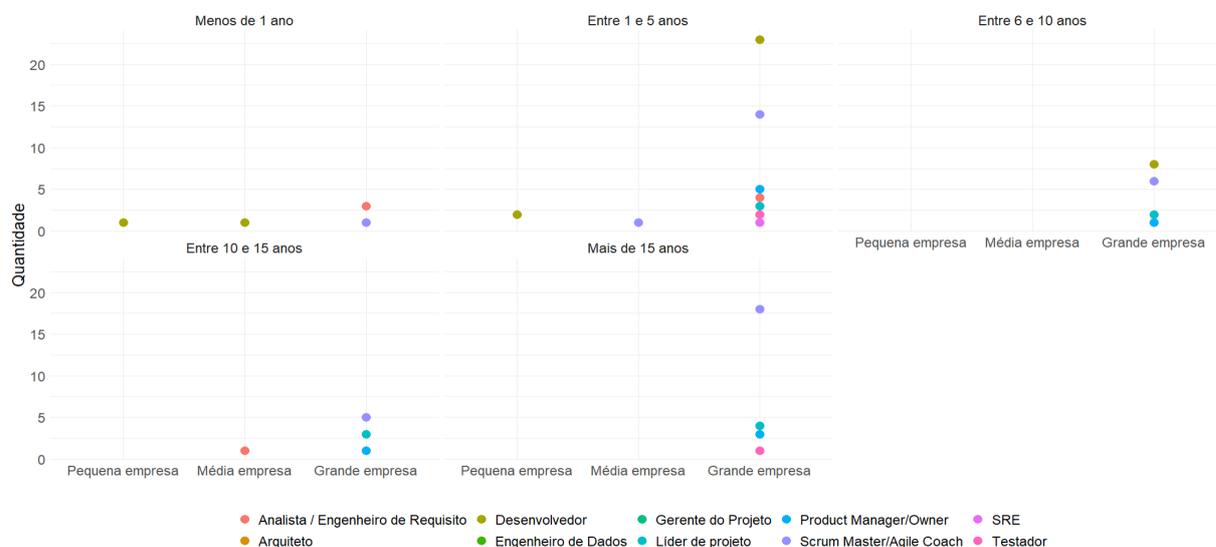
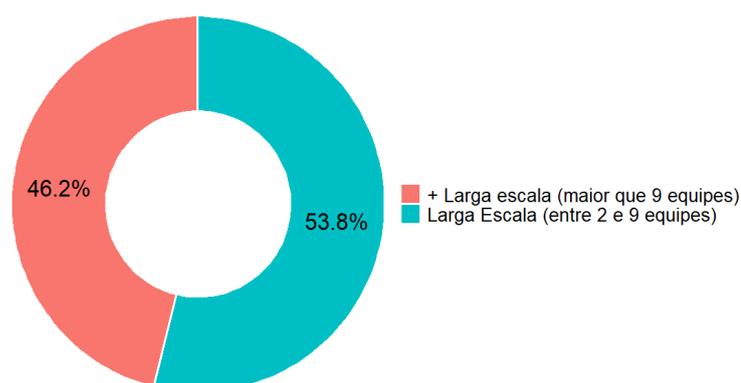


Figura 5 – Categorização da escala de projetos ágeis



7.59% da amostra.

A respeito dos métodos ou *frameworks* ágeis escalonados, combiná-los é o estado da prática e isso se confirma na pesquisa. Os métodos, estruturas e práticas são usados para formar métodos híbridos nos projetos de acordo com 128 respostas (88.27%). Os métodos mais frequentemente usados como ingredientes em métodos híbridos são apresentados na Figura 7. Constata-se que a metodologia Scrum é utilizada na maioria dos projetos ou times (134 - 92.41%), seguida do método de gestão visual *Kanban* (110 - 75.86%). Posteriormente, a gestão Ágil foi destacada por 66 profissionais (45.51%), o *framework* SAFe por 49 (33.79%) e a filosofia *Lean* por 42 (28.96%). A prática *Scrum-of-scrums* foi apontada por 22 participantes (15.17%), logo após a metodologia XP (19 - 13.10%). O modelo do Spotify foi mencionado por 5 respondentes (3.44%). Finalmente, a metodologia *Scrum-at-scale* e o *framework* Nexus surgiram na amostra 5 (3.44%) e 2 (1.37%) vezes, respectivamente.

Figura 6 – Distribuição geográfica das equipes

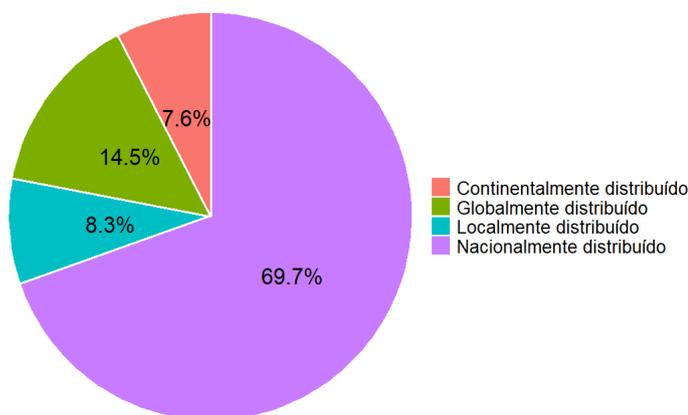
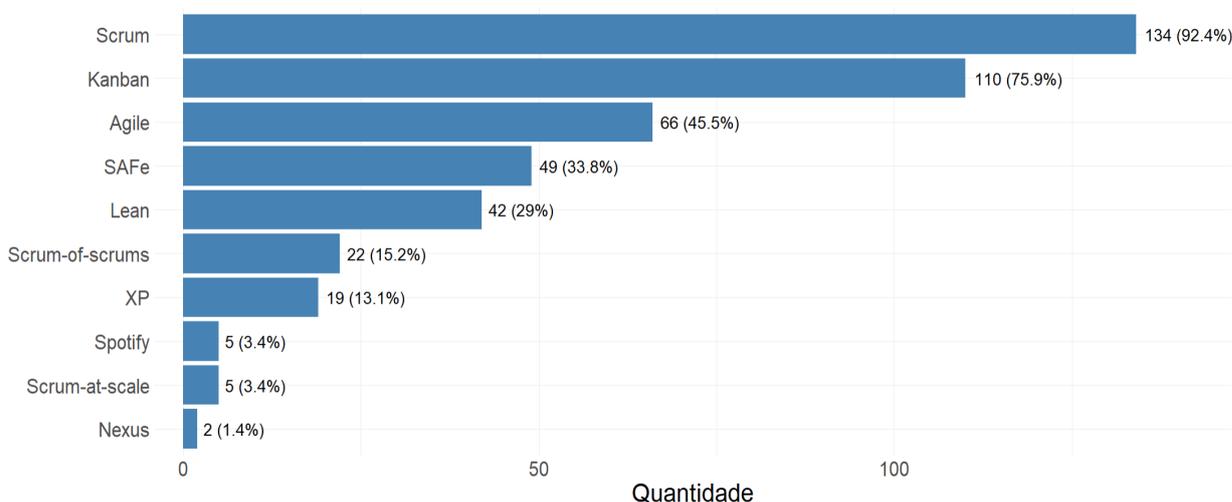


Figura 7 – Utilização de métodos/frameworks ágeis escalonados



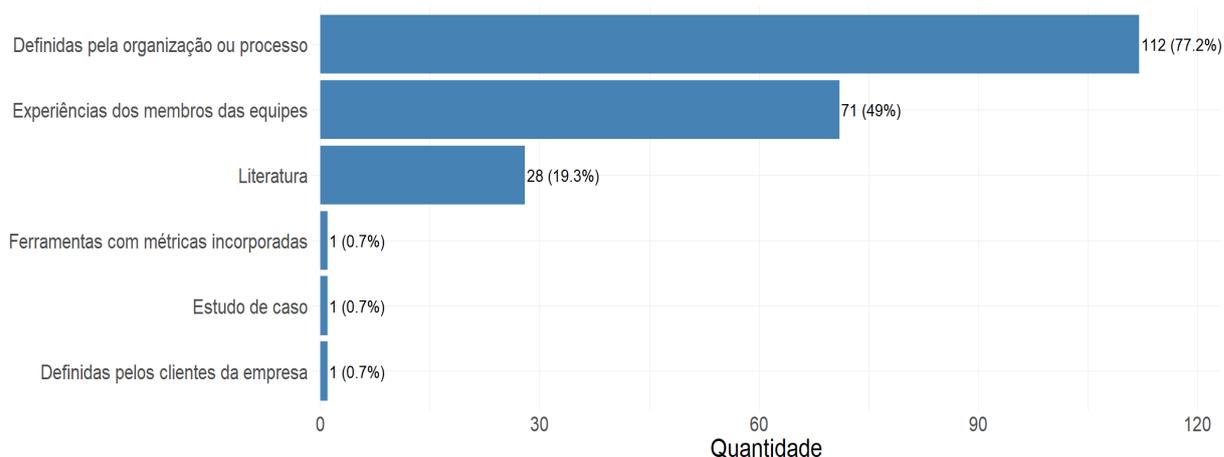
4.3.2 Seleção das Métricas Ágeis

Acerca de como as métricas ágeis são selecionadas nos projetos ou times (Figura 8), em sua maioria, são definidas pela organização ou processo (112 - 77.24%). 48.96% dos profissionais revelaram que as experiências dos membros de cada equipe são consideradas para seleção, um forte subconjunto. 19.31% dos participantes apontaram a relevância da literatura como fonte de seleção. Houveram menções únicas adicionais a formas de seleção mais específicas, por exemplo: métricas definidas “pelos clientes da empresa”, por meio de “estudo de caso” ou de “ferramentas” com métricas incorporadas.

4.3.3 Público-Alvo das Métricas

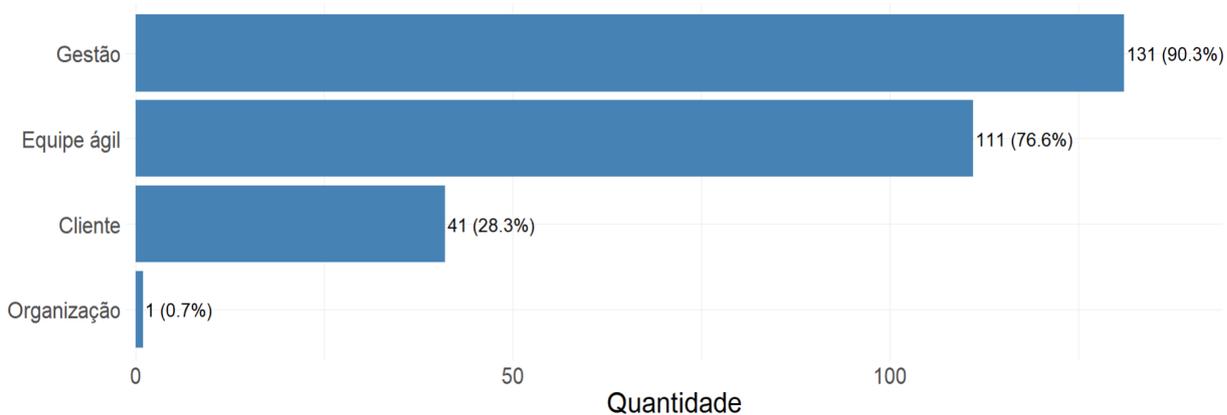
Sobre o público-alvo que faz uso dos resultados das medições (Figura 9), a grande maioria dos participantes (131 - 90.34%) apontou que os profissionais envolvidos com a gestão são os principais utilizadores, em seguida, a equipe ágil foi ressaltada por 111 (76.55%) participantes como importante destinatária das medições e o cliente foi referenciado por

Figura 8 – Estratégias de seleção das métricas ágeis



41 (28.27%) profissionais como um alvo significativo. Uma menção relevante adicional de público-alvo foi: “a organização, através de metas corporativas”.

Figura 9 – Público-alvo das métricas



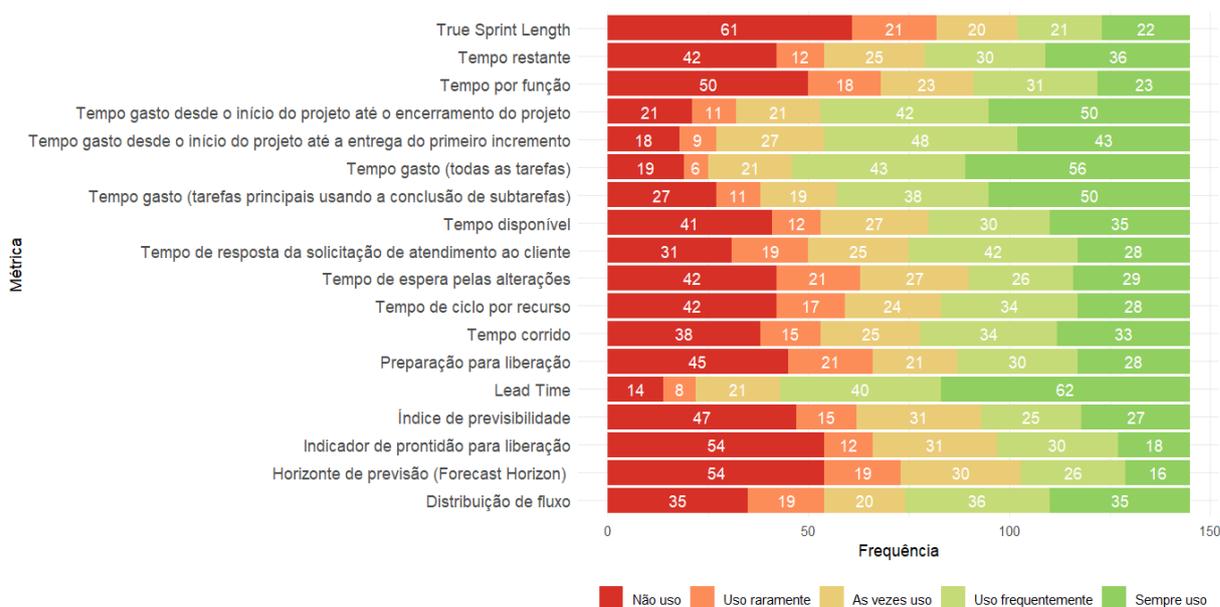
4.3.4 Objetivos e Impactos da Utilização de Métricas

Cinco perguntas do questionário questionavam sobre o uso de 71 métricas no ágil em larga escala, categorizadas e descritas na modelagem GQM (Subseção 4.2.3) guiada por objetivos de negócio resultante da revisão *multivocal*. Como requisito para compor essas perguntas, foi realizada uma triagem das métricas ou instrumentos de medição mais conhecidos, referenciados, bem descritos ou que pudessem ser genericamente aplicáveis em grandes projetos independente do setor ou localidade geográfica. Além da maioria das métricas terem sido traduzidas, a nomenclatura de algumas delas sofreram ajustes em relação às suas fontes originais para facilitar que o público-alvo respondente assimilasse com mais facilidade do que se tratavam ao ler. Os respondentes avaliaram se usam ou não as medições achadas na literatura, bem como a frequência de uso por meio de uma escala

Likert, na qual as respostas foram elaboradas na seguinte escala ordenada: “Não uso”, “Uso raramente”, “As vezes uso”, “Uso frequentemente” e “Sempre uso”. Foi obrigatória a seleção de apenas uma das alternativas para cada métrica. As seguintes descrições para os resultados dessa avaliação foram quantificadas percentualmente considerando o montante de todas as marcações afirmativas de frequência de uso da escala, ou seja, desconsiderando apenas a opção “Não uso”.

A avaliação de frequência das medições relacionadas a melhoria da responsividade aos clientes pode ser visualizada na Figura 10. Os resultados apontam que ‘*Lead Time*’ (90.34%), ‘tempo gasto desde o início do projeto até a entrega do primeiro incremento’ (87.59%), ‘tempo gasto (todas as tarefas)’ (86.9%), ‘tempo gasto desde o início do projeto até o encerramento do projeto’ (85.52%), ‘tempo gasto (tarefas principais usando a conclusão de subtarefas)’ (81.38%), ‘tempo de resposta da solicitação de atendimento ao cliente’ (78.62%), ‘distribuição de fluxo’ (75.86%), ‘tempo corrido’ (73.79%), ‘tempo disponível’ (71.72%), ‘tempo restante’ (71.03%), ‘tempo de espera pelas alterações’ (71.03%) e ‘tempo de ciclo por recurso’ (71.03%) são as métricas do conjunto mais usadas, auxiliando nas monitorações contínuas do fluxo de entregas. As métricas que apresentaram um percentual de uso abaixo de 70% foram: ‘*True Sprint Length*’, ‘horizonte de previsão’, ‘indicador de prontidão para liberação’, ‘tempo por função’, ‘índice de previsibilidade’ e ‘preparação para liberação’; das quais buscam detalhar mais profundamente o tempo real ou previsto para as entregas, conforme documentado em (MENEZES; MARINHO; SAMPAIO, 2024b).

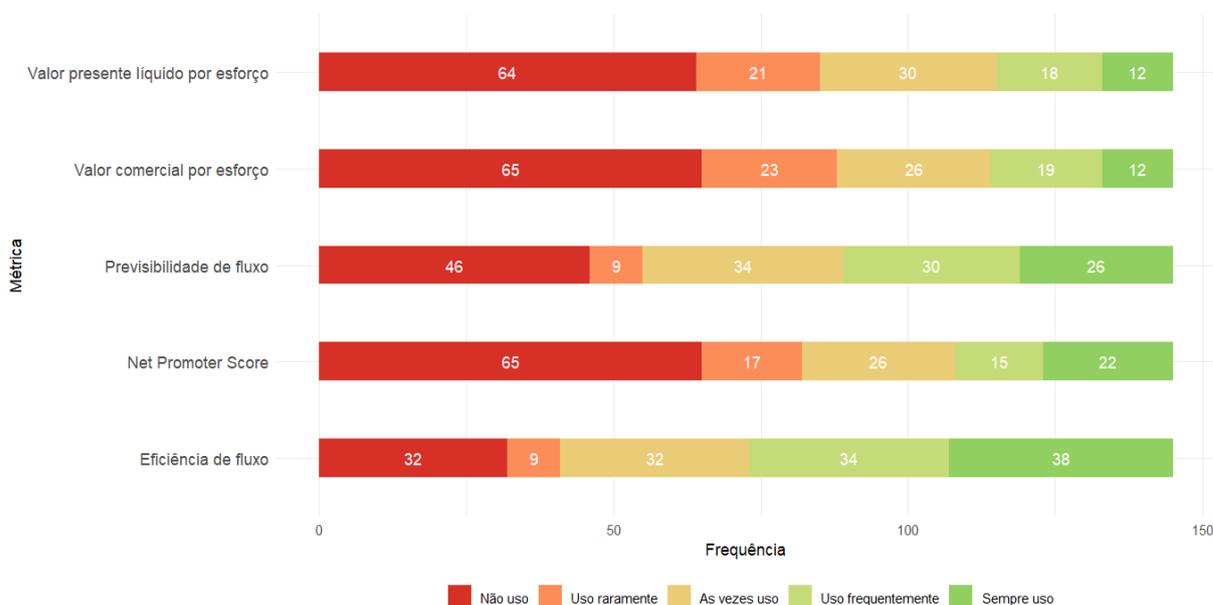
Figura 10 – Uso de métricas relacionadas a responsividade aos clientes



O resultado para as medições voltadas à geração de valor ao cliente está apresentado na Figura 11. ‘Eficiência de fluxo’ (77.93%) e ‘previsibilidade de fluxo’ (68.27%) são as medições mais recorrentes segundo a avaliação. O que expõe a preocupação em medir

quanto do tempo total do fluxo as equipes gastam em atividades de trabalho de valor agregado e o quão bem as equipes podem planejar e atingir seus objetivos de entrega de valor comercial (SAFE, 2022). As métricas ‘valor presente líquido por esforço’ (55.86%), ‘valor comercial por esforço’ (55.17%) e ‘Net Promoter Score’ (55.17%) tiveram resultados aproximados, referenciando práticas de mensuração do valor entregue em termos de lançamentos importantes ou lucratividade por esforço de trabalho e de acompanhamento da satisfação e fidelidade dos clientes, como visto em (MENEZES; MARINHO; SAMPAIO, 2024b).

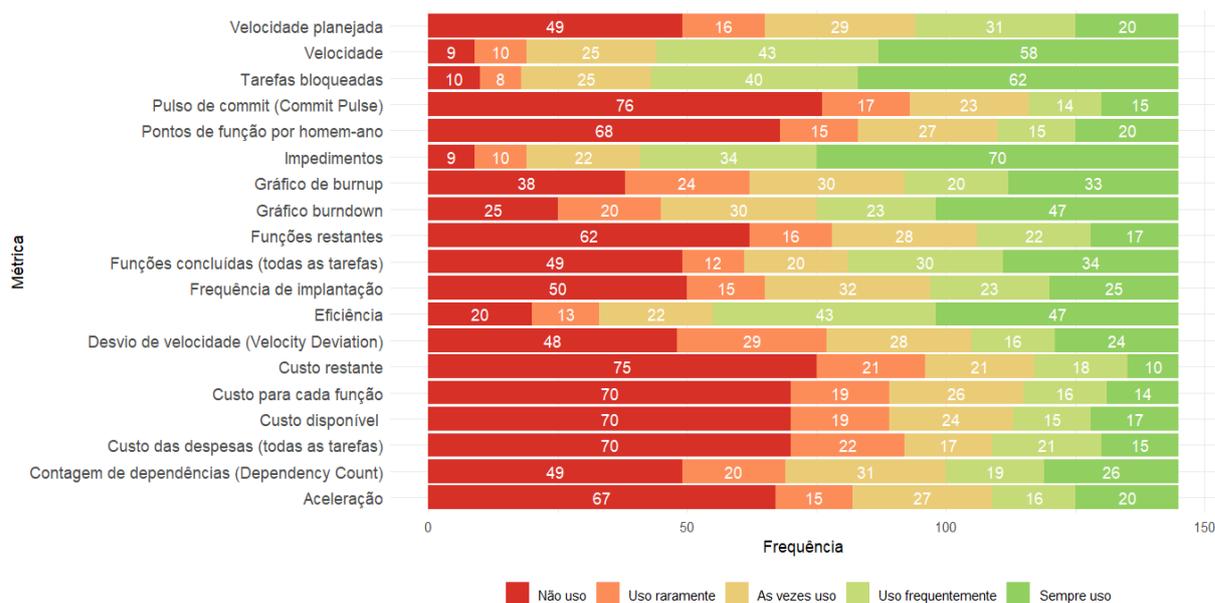
Figura 11 – Uso de métricas relacionadas a entrega de valor aos clientes



É possível observar na Figura 12 que ‘velocidade’ (93.79%), ‘impedimentos’ (93.79%), ‘tarefas bloqueadas’ (93.10%), ‘eficiência’ (86.21%), ‘gráfico *burndown*’ (82.76%) e ‘gráfico *burnup*’ (73.79%) são algumas das métricas relacionadas a produtividade mais frequentemente usadas. Em geral, estão ligadas ao acompanhamento constante do desempenho das equipes. As demais métricas apresentaram um percentual de uso abaixo de 70%, foram elas: ‘pulso de *commit*’, ‘custo restante’, ‘custo das despesas’, ‘custo disponível’, ‘custo para cada função’, ‘aceleração’, ‘pontos de função homem-ano’, ‘funções restantes’, ‘contagem de dependências’, ‘desvio de velocidade’, ‘frequência de implantação’, ‘funções concluídas’ e ‘velocidade planejada’; dentre as quais uma parte está relacionada à medição de custos das tarefas, outra está ligada a quantificação acerca de funções das tarefas e outra parcela aborda a velocidade das equipes sob outras perspectivas (MENEZES; MARINHO; SAMPAIO, 2024b).

A frequência do uso de métricas ligadas a sustentabilidade de práticas está exibida na Figura 13. Este grupo apresenta a menor aderência de acordo com a avaliação, das 12 métricas fornecidas no questionário, apenas a ‘*Work In Progress*’ (94.48%) é usada por mais

Figura 12 – Uso de métricas relacionadas a produtividade



de 70% dos participantes. Isso reafirma a necessidade de monitorar os limites sustentáveis de trabalho para as equipes (TRIPATHI et al., 2015). Entre as métricas que atingiram no mínimo 50% de usabilidade, constam: ‘eficácia’, ‘carga de fluxo’, ‘envolvimento dos funcionários’, ‘desperdício’, ‘pulso da equipe ágil’, ‘delta de precisão da estimativa’ e ‘eficácia percebida’; dentre as quais parte trata de medir o quão bem as equipes planejam as tarefas e controlam a entrada e saída de itens do fluxo de trabalho ao longo do tempo, outra parte está direcionada a medir o quanto as equipes estão engajadas com as metas e processos ágeis da organização (MENEZES; MARINHO; SAMPAIO, 2024b). Para as demais medições de sustentabilidade que atingiram uma frequência de uso menor do que 50%, a maioria mede a adequação das equipes aos modelos de custo aplicados nas empresas (GRIMALDI et al., 2016).

Por fim, na Figura 14, as métricas para alçar a qualidade do produto como ‘cobertura de testes unitários por histórias’ (73.79%), ‘dívida técnica’ (73.10%), ‘número de problemas externos’ (73.10%) e ‘dias abertos com problemas externos’ (72.41%) são as mais apontadas do grupo. O que pode indicar necessidades essenciais das equipes, tais como: monitorar as decisões erradas tomadas acerca dos requisitos, design ou código ao longo dos incrementos (LIYANAGE, 2014); garantir a qualidade funcional do código implementado (TABIB, 2013); e monitorar a sustentação dos sistemas mediante solicitações externas referentes a problemas (HEIDENBERG et al., 2013). As demais medições apresentaram um percentual de uso abaixo de 70%.

Os participantes também foram questionados em relação as finalidades pelas quais as empresas que fazem parte utilizam métricas. A questão foi elaborada a fim de sondar os possíveis resultados alcançados ao utilizar as métricas, ou seja, quais impactos têm

Figura 13 – Uso de métricas relacionadas a sustentabilidade de práticas ágeis

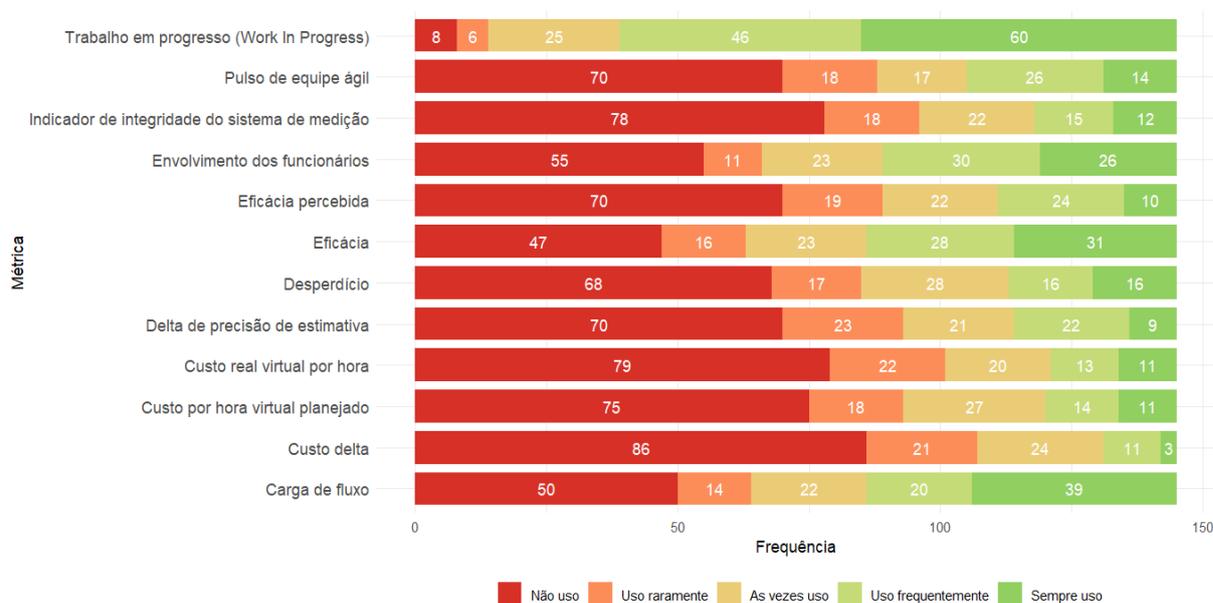
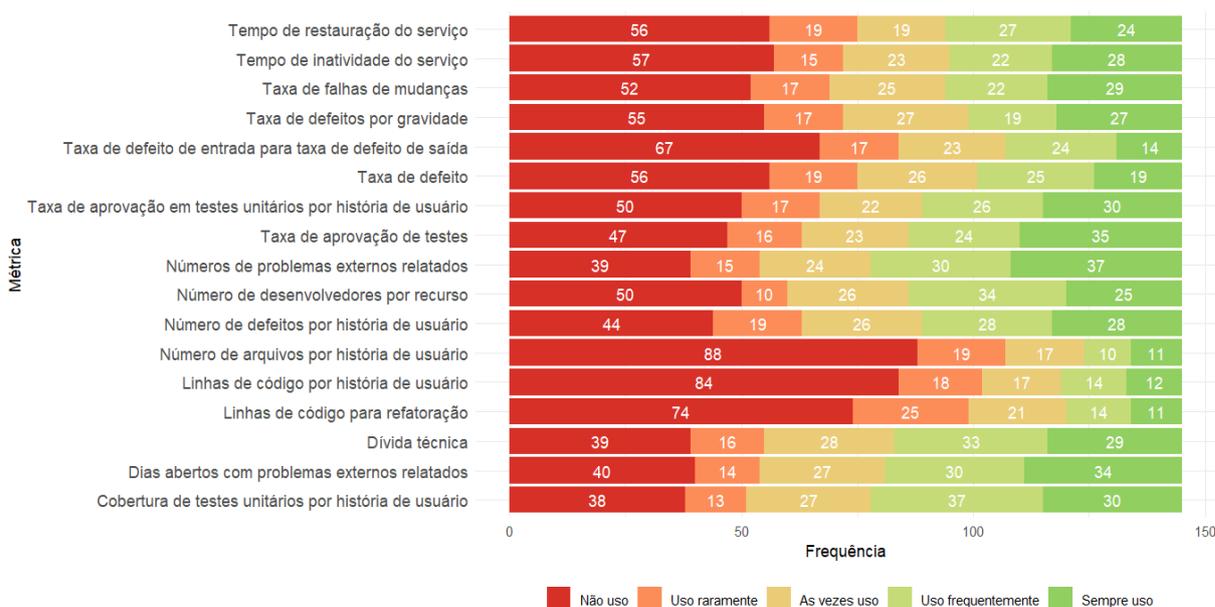
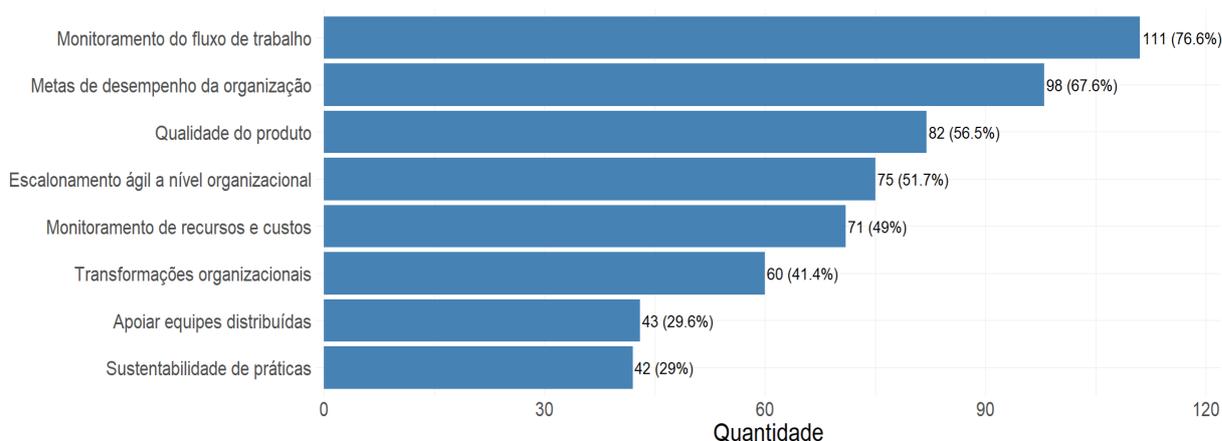


Figura 14 – Uso de métricas relacionadas a qualidade do produto



sejam causados. Na Figura 15 estão exibidas as áreas de impacto que elas têm perseguido: monitoramento do fluxo de trabalho com 111 menções, metas de desempenho da organização com 98, qualidade do produto com 82, escalonamento ágil a nível organizacional com 75 e monitoramento de recursos e custos com 71 menções. Transformações organizacionais, apoio a equipes de desenvolvimento distribuídas geograficamente e sustentabilidade de práticas tiveram, respectivamente, 60, 43 e 42 menções.

Figura 15 – Áreas de impacto da utilização de métricas



4.3.5 Métricas Relatadas na Pesquisa

Foram obtidos relatos do uso de métricas emergentes não apresentadas no questionário, são elas: ‘qualidade de refinamento técnico’, ‘*Aging Backlog*’, ‘*MTTR*’, ‘atividades canceladas’, ‘histogramas de entrega versus quantidade de dias’, ‘atividades entregues versus construídas’, ‘*Health Check*’, ‘porcentagem de retrabalho’, ‘*Aging WIP*’, ‘*Heat Map* por fase do fluxo’, ‘densidade de defeitos por cenário’, ‘índice de reteste’ e ‘completude dos cenários’. Cada métrica foi citada uma única vez. A Tabela 15 apresenta todas as métricas agrupadas junto ao *framework* e método ágil, escala de projeto, distribuição da equipe, setor da indústria e tamanho da empresa nas quais foram relacionadas nas respostas.

A partir da análise da Tabela 15, sobre as medições emergentes, pode-se observar que relacionadas ao Scrum e *Kanban* são maioria, representam juntas as 13 ocorrências. Junto ao Ágil também são bastante utilizadas, aparecendo em 8 das ocorrências. Junto ao *Lean* e SAFe são menos prevalentes, 7 com *Lean* e 5 vezes com SAFe, respectivamente. Menos medições para XP e *Scrum-of-scrums*, XP em 3 e *Scrum-of-scrums* em uma das ocorrências. 7 das métricas são aplicadas em projetos de “+ larga escala”, enquanto 6 em projetos de “larga escala”. 12 métricas são aplicadas por equipes distribuídas nacionalmente. Apenas uma é aplicada por equipes distribuídas globalmente. O setor bancário surge como o principal utilizador dessas novas métricas, com 7 métricas informadas por quatro respondentes. Para os demais setores de telecomunicações, consultoria e financeiro estão vinculadas 3, 1 e 2 métricas respectivamente, informadas por um respondente de cada setor. Todas as métricas analisadas são usadas por grandes empresas.

4.4 Discussão dos Resultados do *Survey*

Nesta seção busca-se responder a questão de pesquisa principal através da realização de uma discussão sobre as respostas de cada questão secundária do *survey*.

Tabela 15 – Métricas relatadas na pesquisa

Métricas	Framework/Método	Escala	Distribuição	Indústria	Tam. da empresa
Densidade de defeitos por cenário, Índice de reteste, Completude dos cenários	Scrum, Kanban, Ágil	+ Larga escala	Nacionalmente	Telecomunicações	Grande empresa
Aging Backlog	Scrum, Kanban, Ágil	Larga Escala	Nacionalmente	Bancário	Grande empresa
Health check	Scrum, Kanban, Ágil, SAFe, Scrum-of-scrams	Larga Escala	Globalmente	Consultoria	Grande empresa
Histogramas de entrega x quantidade de dias, Atividades canceladas, Atividades entregues x construídas, Porcentagem de retrabalho	Scrum, Kanban, Lean, SAFe	+ Larga escala	Nacionalmente	Bancário	Grande empresa
Aging WIP, Heat map por fase do fluxo	Scrum, Kanban, XP, Ágil, Lean	Larga Escala	Nacionalmente	Financeiro	Grande empresa
MTTR	Scrum, Kanban, Ágil	Larga Escala	Nacionalmente	Bancário	Grande empresa
Qualidade de refinamento técnico	Scrum, Kanban, XP, Lean	Larga Escala	Nacionalmente	Bancário	Grande empresa

4.4.1 Como e para quem as métricas são selecionadas?

Para entender como as métricas ágeis são utilizadas nas empresas é importante entender de que forma são selecionadas. A pesquisa buscou investigar tal questão com base inicial em três estratégias complementares para seleção de métricas: (i) definidas pela organização ou processo; (ii) por meio de experiências dos membros da equipe; ou (iii) através da literatura. Dentre os resultados, 77.2% (112) dos participantes apontou que determinadas métricas são definidas pelas organizações ou processos, o que indica que são estratégicas para o negócio e podem ser selecionadas externamente as equipes, muitas vezes por intermédio de grupos de especialistas ou comitês (HOSSAIN; AHMED; ARAFAT, 2021; LEAL et al., 2022; PHILIPP et al., 2023). 49% (71) das respostas indicam que as experiências dos membros de cada equipe também são consideradas para selecionar as medições utilizadas, estratégia favorável na adaptação as constantes mudanças que surgem internamente nos projetos ou equipes ágeis no decorrer dos processos (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021). 19.3% (28) dos profissionais recorrem a pesquisas das métricas mais adequadas na literatura. Alguns profissionais nomearam outros três meios de seleção de métricas importantes: (i) a partir de ferramentas com métricas incorporadas, nas quais pode-se obter medições nativas geradas automaticamente a partir de dados; (ii) definidas pelos clientes da empresa, prática presente em consultorias; e (iii) por meio de estudo caso, método que pode funcionar como um meio para a organização e as equipes

analisarem e identificarem o conjunto de métricas adequado. É provável que tais estratégias adicionais específicas também sejam consideradas em outras empresas. De forma geral, pode se considerar que ao escalar o ágil as organizações de grande porte tendem a selecionar e prescrever às equipes mais métricas para dar suporte em primeiro lugar aos objetivos de negócio, não abandonando as demais estratégias de seleção, mas deixando-as em segundo plano. Esta é a primeira descoberta deste estudo.

- **Descoberta 1:** As métricas utilizadas no ágil escalado são predominantemente selecionadas e definidas pelas organizações ou processos, todavia, pode-se utilizar de outras estratégias complementares.

Também é importante compreender a quem se destinam as métricas definidas nesse contexto, ou seja, para quais públicos as informações geradas pela métrica são destinadas ou por quem são usadas. A pesquisa também investigou essa questão com base inicial em três possíveis públicos-alvo: (i) a gestão; (ii) a equipe ágil; e (iii) o cliente. Enfim, obteve-se o seguinte resultado: 131 (90.3%) relataram que as métricas tinham como objetivo serem aproveitadas pela gestão. 76.6% (111) dos respondentes relataram que as métricas tinham como objetivo focar a equipe ágil, alinhando-se com os princípios e valores ágeis (BECK et al., 2020). Em seguida, 28.3% destacaram o cliente como utilizadores da métrica, resultado coerente para o modelo de negócios de organizações de software dos setores de consultoria e fábrica de software presentes na amostra. A organização foi destacada uma vez adicionalmente como entidade destinatária das métricas “através de metas corporativas”. Essa perspectiva alinha-se ao meio de seleção de métricas dominante e ressalta a importância das medidas de desempenho como ferramentas indispensáveis pelas empresas para orientar e incentivar os membros a alcançarem os resultados almejados (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021). Pode-se inferir que os indicadores destinados à organização passam também pela gestão, que possui prioridade como público-alvo em razão dos ambientes ágeis de escala envolverem várias equipes, ao invés das configurações de uma única equipe ágil (PHILIPP et al., 2023). Portanto, a gestão necessita prioritariamente de métricas que preencham as lacunas do *feedback* qualitativo com as equipes (PHILIPP et al., 2023). Isto pode causar uma contradição com os valores ágeis ao inverter-se a lógica do foco na equipe e nos clientes. Contudo, as necessidades de orientar a tomada de decisão e manter a consciência situacional adequada entre as equipes são tarefas desafiadoras e por tanto, tais indícios apontam para a segunda descoberta:

- **Descoberta 2:** As métricas utilizadas no ágil escalado são predominantemente destinadas a serem utilizadas para gestão monitorar o desempenho das equipes, embora também estão acessíveis à equipe ágil e aos clientes.

4.4.2 Quais são os objetivos e impactos da utilização das métricas?

Na avaliação da frequência de uso das medições, diversas métricas ágeis foram apresentadas junto com um guia de fórmulas de cálculo. As métricas extraídas da literatura foram organizadas em grupos com diferentes escopos e objetivos. Dessa forma, buscou-se observar os elementos mais e menos tendenciosos a serem mensurados para cada objetivo, considerando determinadas faixas de percentual de uso relevantes percebidas na avaliação. A partir disso, para as métricas relacionadas ao objetivo de melhorar a responsividade ao clientes, pode-se iniciar a resposta da questão secundária 2 e gerar a terceira descoberta:

- **Descoberta 3:** As métricas de responsividade aos cliente mais utilizadas no ágil escalado buscam medir continuamente o fluxo de entregas.

Do resultado para as métricas relacionadas ao objetivo de gerar valor para os clientes, a quarta descoberta foi então identificada:

- **Descoberta 4:** As métricas de entrega de valor para os clientes mais utilizadas no ágil escalado buscam medir quanto do tempo total do fluxo as equipes gastam em atividades de trabalho de valor agregado e o quão bem as equipes podem planejar e atingir seus objetivos de entrega de valor comercial.

Em relação as métricas relacionadas ao objetivo de garantir a produtividade, a quinta descoberta foi identificada:

- **Descoberta 5:** As métricas de produtividade mais utilizadas no ágil escalado estão ligadas ao acompanhamento constante do desempenho das equipes.

A respeito das métricas relacionadas ao objetivo de promover a sustentabilidade de práticas, a sexta descoberta foi constatada:

- **Descoberta 6:** As métricas de sustentabilidade de práticas ágeis demonstram ter pouca aderência nas grandes organizações. Apenas ‘*Work In Progress*’ mostrou-se ser usada com frequência.

Por fim, sobre as métricas relacionadas ao objetivo de elevar a qualidade do produto, a sétima descoberta foi detectada:

- **Descoberta 7:** As métricas de qualidade do produto mais utilizadas no ágil escalado destacam a importância de monitorar problemas ao longo dos incrementos, qualidade do código e sustentação dos sistemas.

Acerca dos resultados da utilização das métricas, as respostas revelam que no mínimo metade dos profissionais da amostra compreendem que as medições impactam sobretudo no monitoramento do fluxo de trabalho (76.6%), no alcance das metas de desempenho da organização (67.6%), na busca pela qualidade do produto (56.5%) e no escalonamento ágil para além das equipes de desenvolvimento de software (51.7%). Menos da metade dos participantes enxergam que as métricas geram impacto sobre o monitoramento de recursos e custos no processo, transformações organizacionais, equipes trabalhando distribuídas e sustentabilidade de práticas ágeis. Assim, a oitava descoberta a ser destacada é:

- **Descoberta 8:** O monitoramento do fluxo de trabalho, as metas de desempenho da organização, a qualidade do produto e o escalonamento ágil para além das equipes de desenvolvimento são as áreas mais impactadas pela utilização das métricas nas organizações.

4.4.3 Quais métricas utilizadas no desenvolvimento ágil em larga escala não foram identificadas na literatura?

Após a avaliação dos objetivos para aplicação das métricas com base na frequência de uso, os respondentes relataram 13 novas métricas utilizadas por eles que não foram listadas. Dentre elas, estão presentes medições de qualidade do produto, desempenho e monitoramento do fluxo de trabalho. A análise mostrou serem adotadas em projetos de larga escala conduzidos em grandes empresas, principalmente no setor bancário, com equipes trabalhando remotamente no mesmo país e uma forte preferência por métodos como Scrum, *Kanban* e pela abordagem *Lean Ágil*.

4.5 Análise Comparativa dos Resultados da Revisão *Multivocal* e do *Survey* sobre Métricas Ágeis Escalonadas

Esta seção tem como objetivo comparar os dados da literatura e os resultados do *survey*, visto que estudos secundários estão sujeitos a limitações associadas à disponibilidade, qualidade e relevância dos estudos existentes, assim como, podem não revelar ideias ou tendências emergentes.

Uma análise comparativa das métricas encontradas na revisão da literatura e as métricas efetivamente utilizadas nas empresas é necessária. O presente trabalho cobre a literatura cinza publicada entre Janeiro de 2001 e Janeiro de 2023, somente 19 fontes relevantes foram selecionadas, no entanto, 80 métricas foram reunidas. Durante a elaboração do questionário para o *survey*, 71 métricas da literatura foram escolhidas para compor,

considerando como critério de inclusão apenas as que poderiam ser aplicadas em ambientes e contextos abrangentes de projetos ágeis de larga escala.

Algumas das métricas mais citadas na literatura foram ‘velocidade’ (5), ‘*Work In Progress*’ (2), ‘*Lead Time*’ (3), ‘tempo de ciclo por recurso’ (2) e ‘distribuição de fluxo’ (2). Tais medições também apresentaram bons percentuais de usabilidade afirmativa na pesquisa, 93.79%, 94.48%, 90.34%, 71.03% e 75.86%, respectivamente. O ‘gráfico de *burnup*’, instrumento com medidas incorporadas, também foi citado duas vezes e na pesquisa apresentou um percentual de usabilidade afirmativa de 73.79%. ‘Pontos de função por homem-ano’ também surgiu na literatura duas vezes, contudo, atingiu o percentual de apenas 53.10% de usabilidade afirmativa. As demais métricas encontradas na literatura foram citadas apenas uma vez.

O grupo com métricas que atingiram os percentuais mais altos de usabilidade na avaliação foi o ligado aos objetivos de desempenho das equipes, com ‘velocidade’ (93.79%), ‘impedimentos’ (93.79%) e ‘tarefas bloqueadas’ (93.10%); o que sustenta o acompanhamento primordial da produtividade. Na revisão da literatura, foi o grupo com a maior quantidade de medições. Todavia, o grupo que possui a maior variedade de métricas que alcançaram percentuais altos na avaliação foi o relacionado ao monitoramento da capacidade de resposta ao cliente, com ‘*Lead Time*’ (90.34%), ‘tempo gasto desde o início do projeto até a entrega do primeiro incremento’ (87.59%), ‘tempo gasto (todas as tarefas)’ (86.9%), ‘tempo gasto desde o início do projeto até o encerramento do projeto’ (85.52%), ‘tempo gasto (tarefas principais usando a conclusão de subtarefas)’ (81.38%), ‘tempo de resposta da solicitação de atendimento ao cliente’ (78.62%), ‘distribuição de fluxo’ (75.86%), ‘tempo corrido’ (73.79%), ‘tempo disponível’ (71.72%), ‘tempo restante’ (71.03%), ‘tempo de espera pelas alterações’ (71.03%) e ‘tempo de ciclo por recurso’ (71.03%). Tal resultado pode indicar que os profissionais possuem mais familiaridade com métricas para monitorar a responsividade ou ainda, que esse objetivo compreende uma maior variedade de métricas aplicadas na indústria.

Cinco métricas encontradas na literatura relacionadas a entrega de valor foram submetidas à avaliação. Exceto ‘eficiência de fluxo’ (77.93%) e ‘previsibilidade de fluxo’ (68.27%), os demais resultados indicam uma ocorrência mediana de usabilidade. Isso sugere a princípio que a supervisão do impacto dos projetos e das atividades de desenvolvimento em termos de benefícios para os *stakeholders*, clientes e a própria organização não é algo recorrente. Porém, há de se considerar que essa demanda não costuma fazer parte do domínio de atuação dos cargos em que a maioria dos participantes da pesquisa atuam diretamente.

Embora a literatura tenha apresentado uma variada diversidade de métricas que avaliam a sustentabilidade de práticas, esse não foi um grupo que demonstrou força de usabilidade. Com exceção da ‘*Work In Progress*’, a maior parte delas (7) atingiu um

percentual de uso mínimo entre 50% e 68%, as demais ficaram abaixo. Com base na avaliação das escalas de frequência, o objetivo de garantir a sustentação de práticas nas grandes organizações ágeis demonstra ser preferencialmente apoiado pelo *feedback* qualitativo com as equipes.

Para as medições voltadas a busca pela qualidade do produto ou processo, uma vasta e relevante variedade coletada da literatura foi submetida à avaliação. Com ressalva a ‘cobertura de testes unitários por histórias’ (73.79%), ‘dívida técnica’ (73.10%), ‘número de problemas externos’ (73.10%) e ‘dias abertos com problemas externos’ (72.41%); todo o restante resultou em percentuais de uso menores do que 70%, porém, trata-se de uma vasta diversidade de medições capazes de fornecer transparência e *insights* valiosos sobre qualidade do produto e risco para iniciativas que demandem por velocidade sem comprometer a qualidade. Contudo, também é preciso considerar que os profissionais especializados em qualidade de software representam uma pequena parcela da amostra.

Analisar discrepâncias e convergências entre a literatura cinza e o *survey* acerca da utilização de cada método e *framework* ágil escalonado também é possível. Nas 19 fontes de literatura, observou-se as seguintes ocorrências: Scrum e SAFe (7 vezes cada), Ágil e Lean (4 vezes cada), Kanban (3 vezes). Houve apenas uma única ocorrência para XP, *Scrum-of-Scrums*, Spotify, *Scrum-at-scale* e LeSS. Para a amostra da pesquisa, a presença do Scrum nos projetos também conservou-se hegemônica (92.41%), porém, foi seguida pelo Kanban (75.86%) e Ágil (45.51%). O SAFe e o Lean surgem após, presentes em 33.79% e 28.96% dos projetos, respectivamente.

4.6 Considerações Finais do Capítulo

Este capítulo apresentou uma análise abrangente sobre os resultados obtidos a partir da revisão de literatura *multivocal* que suplementa o mapeamento sistemático preliminar, além dos achados provenientes do *survey* conduzido para consolidar o conhecimento adquirido. A identificação dos estudos acadêmicos proporcionou uma visão inicial geral das produções sobre o tema, dos contextos organizacionais relatados e das variadas métricas usadas no ágil escalado. A revisão de literatura *multivocal*, por meio da integração dos estudos acadêmicos e evidências da literatura cinza, trouxe uma compreensão mais detalhada e prática das métricas identificadas bem como a análise sobre os objetivos e resultados da sua utilização no desenvolvimento de software. Em seguida, foram apresentados os resultados do *survey* realizado com uma amostra de 145 profissionais do mercado de software brasileiro, que forneceram dados empíricos relevantes para validar e complementar os achados da revisão de literatura. A coleta de dados, realizada entre abril e maio de 2024, permitiu evidenciar práticas comuns, objetivos, áreas de impacto e a frequência do uso eficaz das métricas aplicadas em ambientes ágeis escalados. Além disso,

também foram evidenciadas métricas emergentes da pesquisa e uma análise comparativa dos dados resultantes da revisão *multivocal* e do *survey* foi efetuada. No próximo capítulo, uma coleção de métricas procedente dos resultados alcançados será compartilhada como ferramenta destinada a ser utilizada por pesquisadores e profissionais.

5 Catálogo de Métricas para Desenvolvimento Ágil em Larga Escala

Este capítulo tem o objetivo de compartilhar um catálogo de métricas validadas para orientar futuras pesquisas e práticas organizacionais no domínio do desenvolvimento ágil de software em larga escala. Ele é uma ferramenta útil para pesquisadores e profissionais envolvidos em projetos nesse âmbito, incluindo desenvolvedores, gerentes de projeto, profissionais ágeis e de garantia de qualidade.

As tendências de utilização de métricas conectam-se as principais áreas beneficiadas em empresas que praticam ágil escalonado, tais como desempenho e colaboração entre as equipes e qualidade do produto. Além disso, grande parte das medições mostram forte influência na identificação de gargalos no fluxo de trabalho.

5.1 Visão Geral das Métricas

Métricas são medições, variáveis que quantificam propriedades de entidades que são mensuráveis (FENTON; BIEMAN, 2014; LEAL et al., 2022).

Este compilado é um refinamento após avaliação pelos profissionais do mercado da frequência de uso das métricas listadas no modelo derivado da revisão de literatura. As métricas inclusas neste trabalho tiveram como critério ter atingido o percentual mínimo de 70% de usabilidade. Dessa forma, a afirmação de utilização recorrente pelos participantes do levantamento sustenta uma maior probabilidade de eficácia e sucesso das medições selecionadas.

Antes de exibir o catálogo, o modelo utilizado para apresentar as métricas foi inspirado por (BURKARD et al., 2024) e será definido como um *template* na Tabela 16, que contém os seguintes campos específicos: nome, id, descrição, fórmula de cálculo, etapa do ciclo de vida, popularidade na indústria, exemplos de aplicação e referências. O objetivo deste *template* é proporcionar uma estrutura clara para facilitar a identificação, compreensão e aplicação das métricas.

5.2 Métricas

Esta seção apresenta apenas as 24 métricas mais populares do catálogo seguindo o modelo proposto. Elas estão categorizadas em subseções por cada objetivo de negócio identificado na modelagem (GQM) deste estudo. Todavia, o catálogo completo com as 71

Tabela 16 – Modelo utilizado para descrição de métricas ágeis

Nome	Nome da métrica.
ID	Identificador único da métrica.
Descrição	Explicação detalhada do que a métrica mede.
Fórmula de Cálculo	Fórmula matemática ou lógica usada para calcular a métrica.
Etapa do Ciclo de Vida	Etapa do Ciclo de Vida de Desenvolvimento de Software favorável para mensurar.
Popularidade na Indústria	Valor percentual que denota o grau de relevância da métrica de acordo com a avaliação de frequência de uso do <i>survey</i> .
Exemplos de Aplicação	Se acessível, exemplos práticos de como as métricas foram aplicadas em organizações de software.
Referências	Lista de fontes utilizadas para desenvolver o catálogo.

métricas postas em avaliação também foi disponibilizado digitalmente e pode ser acessado, compreendido e utilizado em Catálogo de Métricas para Desenvolvimento Ágil em Larga Escala¹.

5.2.1 Responsividade aos Clientes

5.2.1.1 *Lead Time*

- **ID:** M001.
- **Descrição:** É uma métrica ágil que mede o tempo total desde o início de uma tarefa ou pedido até a sua conclusão. Ele abrange todo o ciclo de vida do trabalho, incluindo o tempo de espera e o tempo ativo de execução, proporcionando uma visão clara da eficiência dos processos e permitindo a identificação de gargalos e oportunidades de melhoria.
- **Fórmula de Cálculo:** *Data de conclusão de um processo - Data de início de um processo.*
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 90.34%.

¹ <https://renatoviei.github.io/metricas-para-desenvolvimento-agil-em-larga-escala/index.html>

- **Exemplos de Aplicação:** Ao trabalhar com grandes organizações contendo dezenas ou centenas de equipes, Greening (GREENING, 2015) apontou o ‘*Lead Time*’ como métrica essencial para uma estrutura Scrum escalável, incluindo algumas variantes, pois produz previsões do prazo de execução das entregas. Focado nos desafios na fase de dimensionamento do Kanban em uma organização envolvida no desenvolvimento de software *multisite*, o trabalho de Tripathi *et al.* (TRIPATHI *et al.*, 2015) ressalta que uma ‘*média Lead Time*’ também seria uma métrica apropriada para medir o fluxo de trabalho no nível organizacional no Kanban em larga escala.
- **Referências:** (HEIDENBERG *et al.*, 2013; GREENING, 2015; TRIPATHI *et al.*, 2015).

5.2.1.2 Tempo gasto desde o início do projeto até a entrega do primeiro incremento (Em inglês: *Time to delivery first increment*)

- **ID:** M002.
- **Descrição:** Avalia a rapidez do tempo de ciclo em que uma equipe consegue entregar valor ao cliente ou usuário final pela primeira vez, destacando a eficiência no início do projeto e a capacidade de lançar produtos ou funcionalidades no mercado de forma ágil.
- **Fórmula de Cálculo:** *Data de entrega do primeiro incremento - Data de início do projeto.*
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 87.59%.
- **Exemplos de Aplicação:** Um estudo de caso em uma organização bancária relatou uma dimensão de medição de tempo de ciclo relacionada a negócio usada para o processo escalável de entrega ágil (Scrum e XP) de uma organização bancária. Ao considera-la, foi possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessários por meio de métricas como esta (BROWN, 2011).
- **Referências:** (BROWN, 2011).

5.2.1.3 Tempo gasto desde o início do projeto até o encerramento do projeto (Em inglês: *Time to project closure*)

- **ID:** M003.
- **Descrição:** Avalia a duração total do ciclo de vida do projeto, desde a concepção e planejamento inicial até a entrega final e encerramento. O ‘*time to project closure*’ é

útil para entender a eficiência do processo de desenvolvimento e entrega de projetos completos.

- **Fórmula de Cálculo:** *Data de encerramento do projeto - Data de início do projeto.*
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 85.52%.
- **Exemplos de Aplicação:** Um estudo de caso em uma organização bancária relatou uma dimensão de medição de tempo de ciclo relacionada a negócio usada para o processo escalável de entrega ágil (Scrum e XP) de uma organização bancária. Ao considera-la, foi possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessários por meio de métricas como esta (BROWN, 2011).
- **Referências:** (BROWN, 2011).

5.2.1.4 Tempo gasto (todas as tarefas)

- **ID:** M004.
- **Descrição:** Avalia o tempo gasto durante a implementação da tarefa com base na data de teste.
- **Fórmula de Cálculo:** $((Data\ do\ teste\ da\ tarefa - Data\ de\ início\ da\ tarefa) / Tempo\ de\ duração\ da\ tarefa) * 100.$
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 86.9%.
- **Exemplos de Aplicação:** Thawaba *et al.* (THAWABA *et al.*, 2020) propuseram um novo mecanismo de medição para o desenvolvimento de software que melhora a avaliação da confiabilidade e a tomada de decisões durante os processos de desenvolvimento ágil de sistemas críticos de segurança utilizando SAFe. Desse modo, para auxiliar a equipe de gerenciamento ágil em todas as fases do projeto em um departamento de cirurgia médica, foi proposto um novo mecanismo de medição no qual métricas de realização como esta foram definidas para demarcar a realização segura de todas as tarefas após o teste.
- **Referências:** (THAWABA *et al.*, 2020).

5.2.1.5 Tempo gasto (tarefas principais usando a conclusão de subtarefas)

- **ID:** M005.
- **Descrição:** Percentagem de tempo gasto durante a implementação das tarefas principais pela soma da percentagem de tempo gasto nas suas subtarefas e dividida pelo número das suas subtarefas.
- **Fórmula de Cálculo:** $Soma(Tempo\ gasto\ em\ subtarefas) / Número\ de\ subtarefas$.
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 81.38%.
- **Exemplos de Aplicação:** Thawaba *et al.* ([THAWABA et al., 2020](#)) propuseram um novo mecanismo de medição para o desenvolvimento de software que melhora a avaliação da confiabilidade e a tomada de decisões durante os processos de desenvolvimento ágil de sistemas críticos de segurança utilizando SAFe. Desse modo, para auxiliar a equipe de gerenciamento ágil em todas as fases do projeto em um departamento de cirurgia médica, foi proposto um novo mecanismo de medição no qual métricas de realização como esta foram definidas para identificar a realização segura das tarefas principais usando a realização de subtarefas.
- **Referências:** ([THAWABA et al., 2020](#)).

5.2.1.6 Tempo de resposta da solicitação de atendimento ao cliente

- **ID:** M006.
- **Descrição:** Mede a capacidade de retorno de solicitações de serviço ao cliente.
- **Fórmula de Cálculo:** $Data\ de\ resolução\ da\ solicitação\ de\ serviço\ ao\ cliente - Data\ de\ criação\ da\ solicitação\ de\ serviço\ ao\ cliente$.
- **Etapa do Ciclo de Vida:** Implantação.
- **Popularidade na Indústria:** 78.62%.
- **Exemplos de Aplicação:** Foi incluída em um modelo de medições elaborado durante *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo ([HEIDENBERG et al., 2013](#)). De acordo com Reinertsen ([REINERTSEN, 2009](#)), em áreas onde o tempo de resposta é de alto valor às partes interessadas, como operações de suporte, esta é a principal métrica que deve ser usada para melhorar o serviço.
- **Referências:** ([REINERTSEN, 2009](#); [HEIDENBERG et al., 2013](#)).

5.2.1.7 Distribuição de fluxo

- **ID:** M007.
- **Descrição:** Mede a quantidade de cada tipo de trabalho no sistema ao longo do tempo.
- **Fórmula de Cálculo:** Uma medida simples é contar o número de cada tipo de item de trabalho em qualquer momento. Uma medida mais precisa pode considerar o tamanho de cada item de trabalho.
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 75.86%.
- **Exemplos de Aplicação:** Foi incluída em um modelo de medições elaborado durante *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo (HEIDENBERG et al., 2013). Ter um fluxo contínuo e sem gargalos permite que a organização de desenvolvimento responda melhor e rapidamente às solicitações dos clientes, conforme argumentam Petersen e Wohlin (PETERSEN; WOHLIN, 2011).
- **Referências:** (PETERSEN; WOHLIN, 2011; HEIDENBERG et al., 2013; SAFE, 2022).

5.2.1.8 Tempo corrido

- **ID:** M008.
- **Descrição:** Tempo total decorrido para todas as etapas de um fluxo de trabalho.
- **Fórmula de Cálculo:** Medido pelo período médio de tempo necessário para concluir um certo tipo de item de trabalho.
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 73.79%.
- **Exemplos de Aplicação:**
- **Referências:** (SAFE, 2022).

5.2.1.9 Tempo disponível

- **ID:** M009.
- **Descrição:** Tempo disponível para subtarefas ou tarefas principais com base na data de teste.
- **Fórmula de Cálculo:** *Tempo gasto em subtarefas - (Data de teste da tarefa - Data de início da tarefa).*
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 71.72%.
- **Exemplos de Aplicação:** Thawaba *et al.* (THAWABA *et al.*, 2020) propuseram um novo mecanismo de medição para o desenvolvimento de software que melhora a avaliação da confiabilidade e a tomada de decisões durante os processos de desenvolvimento ágil de sistemas críticos de segurança utilizando SAFe. Desse modo, para auxiliar a equipe de gerenciamento ágil em todas as fases do projeto em um departamento de cirurgia médica, foi proposto um novo mecanismo de medição no qual métricas de recursos como esta foram definidas para identificar com segurança após o teste apenas o tempo disponível para a implementação das subtarefas ou tarefas principais.
- **Referências:** (THAWABA *et al.*, 2020).

5.2.1.10 Tempo restante

- **ID:** M010.
- **Descrição:** Tempo restante para concluir a tarefa com base na data de teste.
- **Fórmula de Cálculo:** *Tempo de duração da tarefa / (Data do teste da tarefa - Data de início da tarefa).*
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 71.03%.
- **Exemplos de Aplicação:** Thawaba *et al.* (THAWABA *et al.*, 2020) propuseram um novo mecanismo de medição para o desenvolvimento de software que melhora a avaliação da confiabilidade e a tomada de decisões durante os processos de desenvolvimento ágil de sistemas críticos de segurança utilizando SAFe. Desse modo, para auxiliar a equipe de gerenciamento ágil em todas as fases do projeto em um departamento de cirurgia médica, foi proposto um novo mecanismo de medição

no qual métricas de realização como esta foram definidas para demarcar o tempo restante para realização segura de todas as tarefas após o teste.

- **Referências:** ([THAWABA et al., 2020](#)).

5.2.1.11 Tempo de espera pelas alterações

- **ID:** M011.
- **Descrição:** Avalia o tempo que uma tarefa ou item de trabalho permanece em espera para alterações antes de ser iniciado ou retomado, identificando ineficiências e gargalos no fluxo de trabalho.
- **Fórmula de Cálculo:** Quantidade de tempo de espera para fazer uma alteração.
- **Etapa do Ciclo de Vida:** Implantação.
- **Popularidade na Indústria:** 71.03%.
- **Exemplos de Aplicação:**
- **Referências:** ([SAFE, 2022](#)).

5.2.1.12 Tempo de ciclo por recurso

- **ID:** M012.
- **Descrição:** Monitora o tempo de ciclo para recursos selecionados para desenvolvimento durante o decorrer do projeto.
- **Fórmula de Cálculo:** *Data de conclusão da prontidão do recurso - Data em que o recurso foi adicionado ao backlog.*
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 71.03%.
- **Exemplos de Aplicação:** Foi incluída em um modelo de medições elaborado durante *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo ([HEIDENBERG et al., 2013](#)). O tempo de ciclo rápido foi essencial para vantagens competitivas, conforme destacado no estudo de Petersen ([PETERSEN, 2010](#)) acerca de uma empresa de larga escala.
- **Referências:** ([PETERSEN, 2010](#); [HEIDENBERG et al., 2013](#); [STETTINA](#); [SCHONEMAKER, 2018](#))

5.2.2 Entrega de Valor

5.2.2.1 Eficiência de fluxo

- **ID:** M019.
- **Descrição:** Mede quanto do tempo total de fluxo é gasto em atividades de trabalho que agregam valor versus esperando entre etapas.
- **Fórmula de Cálculo:** *Tempo ativo total / tempo de fluxo.*
- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 77.93%.
- **Exemplos de Aplicação:**
- **Referências:** ([SAFE](#), 2022).

5.2.3 Produtividade

5.2.3.1 Velocidade

- **ID:** M024.
- **Descrição:** Mede a quantidade de trabalho que uma equipe de desenvolvimento consegue concluir em um ciclo de trabalho. Ela é calculada somando os pontos de história ou outras unidades de medida de todas as tarefas concluídas durante o ciclo de trabalho. A velocidade fornece uma visão da produtividade da equipe e ajuda a estimar a capacidade de entrega futura, permitindo um planejamento mais preciso e a identificação de tendências ao longo do tempo.
- **Fórmula de Cálculo:** Quantidade de pontos de história concluídos para itens de trabalho de um tipo durante um período de tempo.
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 93.79%.
- **Exemplos de Aplicação:** Um estudo de caso em uma organização bancária relatou uma dimensão de medição de tempo de ciclo relacionada a agilidade usada para o processo escalável de entrega ágil (Scrum e XP) de uma organização bancária. Ao considerá-la, foi possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessários por meio de métricas de produtividade como esta ([BROWN](#), 2011). Ao trabalhar com grandes organizações contendo dezenas ou centenas de equipes, Greening ([GREENING](#), 2015) apontou ‘velocidade’ como

a métrica de produtividade mais comum, pois trata-se da taxa na qual a equipe conclui recursos ([GREENING, 2015](#)).

- **Referências:** ([BROWN, 2011](#); [GREENING, 2015](#); [STETTINA](#); [SCHOEMAKER, 2018](#); [SAFE, 2022](#)).

5.2.3.2 Impedimentos

- **ID:** M025.
- **Descrição:** Qualquer ‘perda de tempo’ devido a um defeito ou obstáculo que prejudique a produtividade.
- **Fórmula de Cálculo:** Número de horas que não produzem resultados tangíveis.
- **Etapa do Ciclo de Vida:** Viabilidade do produto.
- **Popularidade na Indústria:** 93.79%.
- **Exemplos de Aplicação:** Grimaldi *et al.* ([GRIMALDI et al., 2016](#)) trataram de um caso real e complexo de uma implementação do SAFe em uma grande empresa, centrado e enriquecido com métricas de produtividade que monitoraram e geraram impactos nos prazos de entrega do produto e na redução de custos tais como esta.
- **Referências:** ([GRIMALDI et al., 2016](#)).

5.2.3.3 Tarefas bloqueadas

- **ID:** M026.
- **Descrição:** Esta métrica rastreia quantas tarefas foram impedidas de avançar devido a bloqueios ou impedimentos, proporcionando uma visão clara dos obstáculos frequentes que afetam o fluxo de trabalho e a produtividade da equipe. Ela é útil para identificar áreas problemáticas que precisam de atenção para melhorar a eficiência e o fluxo do processo.
- **Fórmula de Cálculo:** Número de tarefas bloqueadas durante um período de tempo específico.
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 93.10%.
- **Exemplos de Aplicação:** Um estudo de caso em uma organização bancária relatou uma dimensão de medição relacionada a agilidade usada para o processo escalável de entrega ágil (Scrum e XP) de uma organização bancária. Ao considera-la, foi

possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessários por meio de métricas como esta (BROWN, 2011).

- **Referências:** (BROWN, 2011).

5.2.3.4 Eficiência

- **ID:** M027.
- **Descrição:** Indica o quão bem usa-se a equipe em comparação com a capacidade máxima.
- **Fórmula de Cálculo:** AH / C . Onde AH representa as horas reais gastas produzindo resultados tangíveis e C é o número ideal de horas que uma equipe pode entregar.
- **Etapa do Ciclo de Vida:** Viabilidade do produto.
- **Popularidade na Indústria:** 86.21%.
- **Exemplos de Aplicação:** Grimaldi *et al.* (GRIMALDI *et al.*, 2016) trataram de um caso real e complexo de uma implementação do SAFe em uma grande empresa, centrado e enriquecido com métricas de produtividade que monitoraram e geraram impactos nos prazos de entrega do produto e na redução de custos tais como esta.
- **Referências:** (GRIMALDI *et al.*, 2016).

5.2.3.5 Gráfico *burndown*

- **ID:** M028.
- **Descrição:** Mede o progresso da *sprint* e fornece indicadores do processo de trabalho da equipe.
- **Fórmula de Cálculo:** Marca os dias da *sprint* no eixo horizontal e os pontos planejados para compor a *sprint* no eixo vertical, começando dos pontos máximos da *sprint* (velocidade da equipe) até zero.
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 82.76%.
- **Exemplos de Aplicação:** Um estudo de caso em uma organização bancária relatou uma dimensão de medição de produtividade relacionada a agilidade usada para o processo escalável de entrega ágil (Scrum e XP) de uma organização bancária. Ao considera-la, foi possível obter *feedback* se os projetos estivessem começando a enfrentar desafios desnecessários por meio de métricas de produtividade como esta (BROWN, 2011).

- **Referências:** (BROWN, 2011).

5.2.3.6 Gráfico de *burnup*

- **ID:** M029.
- **Descrição:** Mede o progresso com base nas horas ou pontos restantes de cima para baixo. Mede o progresso do lançamento e fornece indicadores do processo de trabalho da equipe.
- **Fórmula de Cálculo:** Marca os dias da *sprint* no eixo horizontal e os pontos planejados para compor a *sprint* no eixo vertical, começando dos pontos máximos da *sprint* (velocidade da equipe) até zero.
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 73.79%.
- **Exemplos de Aplicação:** Em um resumo de estudo de caso relatado por Korson (KORSON, 2015), utilizar o ‘gráfico de *burnup*’ como indicador de aumento de escopo para realização de correções foi essencial para o andamento e conclusão do projeto de um produto que consiste em um conjunto de aplicações web externas usadas diretamente por clientes e uma empresa do setor financeiro.
- **Referências:** (BROWN, 2011; KORSON, 2015).

5.2.4 Sustentabilidade de Práticas Ágeis

5.2.4.1 *Work In Progress*

- **ID:** M043.
- **Descrição:** A equipe pode usar a métrica ‘*Work In Progress*’ (WIP) para fornecer transparência em seu progresso, ajustando potencialmente seu WIP de acordo com as necessidades e melhorando seu fluxo.
- **Fórmula de Cálculo:** Quantidade máxima de trabalho que pode existir em cada status de um fluxo de trabalho.
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 94.48%.
- **Exemplos de Aplicação:** Focado nos desafios na fase de dimensionamento do Kanban em uma organização envolvida no desenvolvimento de software *multisite*, o trabalho de Tripathi *et al.* (TRIPATHI *et al.*, 2015) relatou a definição de limites de

WIP para o *backlog* da área de produto e as equipes de desenvolvimento por acordo mútuo. O limite de WIP é definido para cada *backlog* de produto da área, definido por um acordo comum entre os proprietários do produto da área e o representante da equipe (TRIPATHI et al., 2015). Também no resumo de estudo de caso relatado por Korson (KORSON, 2015), estabelecer um limite de WIP para realização de correções foi essencial para o andamento e conclusão do projeto de um produto que consiste em um conjunto de aplicações web externas usadas diretamente por clientes e uma empresa do setor financeiro.

- **Referências:** (TRIPATHI et al., 2015; KORSON, 2015).

5.2.5 Qualidade do Produto

5.2.5.1 Cobertura de testes unitários por histórias

- **ID:** M055.
- **Descrição:** Mede a extensão em que os testes unitários cobrem as funcionalidades descritas nas histórias de usuário. Esta métrica avalia a qualidade e a robustez do código ao garantir que as funcionalidades implementadas tenham testes unitários associados, validando seu comportamento esperado.
- **Fórmula de Cálculo:** *Cobertura de testes unitários / história de usuário.*
- **Etapa do Ciclo de Vida:** Desenvolvimento.
- **Popularidade na Indústria:** 73.79%.
- **Exemplos de Aplicação:** Tabib (TABIB, 2013) expôs em seu relato de experiência métricas essenciais como ‘cobertura de testes unitários por histórias’ que levaram uma grande organização a capacitar a colaboração e alinhamento por meio da transparência, fornecendo visão sobre qualidade e risco.
- **Referências:** (TABIB, 2013).

5.2.5.2 Dívida técnica

- **ID:** M056.
- **Descrição:** Indicador de decisões erradas tomadas em relação aos requisitos, design ou código.
- **Fórmula de Cálculo:** Duas maneiras: (i) Abordagens baseadas em ferramentas como *SonarQube* e outros; ou (ii) a equipe deve discutir e selecionar um conjunto

dos resultados de dívida técnica mais dolorosos e começar a rastreá-los como parte de suas reuniões retrospectivas.

- **Etapa do Ciclo de Vida:** Manutenção.
- **Popularidade na Indústria:** 73.10%.
- **Exemplos de Aplicação:**
- **Referências:** ([LIYANAGE, 2014](#)).

5.2.5.3 Número de problemas externos

- **ID:** M057.
- **Descrição:** Mede o número total de relatórios de problemas externos gerados durante um período de tempo após o lançamento de uma versão específica.
- **Fórmula de Cálculo:** Número de problemas externos relatados de uma determinada versão.
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 73.10%.
- **Exemplos de Aplicação:** Foi incluída em um modelo de medições elaborado durante *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo ([HEIDENBERG et al., 2013](#)). ‘Número de problemas externos’ leva em consideração o aspecto da qualidade do produto desenvolvido, melhorias em outras áreas não devem ocorrer em detrimento da qualidade do produto ([HEIDENBERG et al., 2013](#)).
- **Referências:** ([HEIDENBERG et al., 2013](#)).

5.2.5.4 Dias abertos com problemas externos

- **ID:** M058.
- **Descrição:** Mede o número médio de dias que os relatórios de problemas externos permaneceram não resolvidos da criação à resolução.
- **Fórmula de Cálculo:** *Data de resolução dos relatórios de problemas - Data de criação dos relatórios de problemas.*
- **Etapa do Ciclo de Vida:** Teste.
- **Popularidade na Indústria:** 72.41%.

- **Exemplos de Aplicação:** Foi incluída em um modelo de medições elaborado durante *workshops* com especialistas da indústria e pesquisadores para ser aplicável a projetos de qualquer tamanho, complexidade e escopo (HEIDENBERG et al., 2013). ‘Dias abertos com problemas externos’ leva em consideração que se os problemas levarem consistentemente mais tempo para serem resolvidos, são indícios de que a qualidade do produto se deteriorou (HEIDENBERG et al., 2013).
- **Referências:** (HEIDENBERG et al., 2013).

5.3 Considerações Finais do Capítulo

Este capítulo compartilhou, como lições aprendidas, um catálogo de métricas populares e eficazes para orientar futuras pesquisas na área e grandes organizações de software que implementam agilidade em larga escala. A coleção foi construída considerando e traçando paralelos entre os resultados do levantamento e da revisão de literatura conduzidas. No próximo capítulo, serão realizadas discussões sobre os resultados obtidos, limitações e ameaças à validade da pesquisa. Além disso, serão pontuadas as oportunidades de pesquisas futuras.

6 Discussão

Este capítulo tem por finalidade apresentar uma discussão focada em interpretar os resultados obtidos, situá-los no contexto da literatura existente e explorar suas implicações. Também serão discutidas as principais limitações e ameaças à validade deste estudo. Ao final do capítulo, serão sugeridas direções para pesquisas futuras, visando contribuir para o avanço do campo de estudo e inspirar trabalhos subsequentes.

Os resultados obtidos na pesquisa sobre métricas ágeis em larga escala revelam uma diversidade de práticas e percepções entre os profissionais da indústria de software brasileira. A revisão de literatura *multivocal* identificou várias métricas utilizadas no desenvolvimento ágil, como ‘velocidade’, ‘eficiência de fluxo’, ‘WIP’ e ‘dívida técnica’, refletindo as tendências de adoção de métricas orientadas por diferentes categorias de objetivos de negócio como resposta aos clientes, entrega de valor, produtividade, sustentabilidade de práticas e qualidade do produto (HEIDENBERG *et al.*, 2013; LEAL *et al.*, 2022). Tais medições podem gerar impactos em diversas áreas e suportam a agilidade em contextos complexos e escaláveis (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021; LEAL *et al.*, 2022).

A pesquisa de campo, realizada com 145 profissionais do setor, avaliou a relevância dessas métricas e evidenciou a aplicação prática em ambientes organizacionais diversos. Os participantes afirmaram que a seleção de métricas pode ser proveniente de diversas estratégias, contudo é fortemente influenciada pela definição de metas corporativas e busca pela melhoria contínua de processos ágeis. Estudos como os de Dingsøyr e Moe (DINGSØYR; MOE, 2014) e Korpivaara *et al.* (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021) já haviam apontado a necessidade de métricas adequadas para avaliar o desempenho e promover a transparência em projetos ágeis de grande escala. De maneira similar, Kettunen (KETTUNEN, 2014) argumenta que as organizações podem ser aprimoradas medindo o desempenho de acordo com os estados desejados. Métricas estratégicas para o negócio podem ser selecionadas por intermédio de grupos de especialistas ou comitês (HOSSAIN; AHMED; ARAFAT, 2021; PHILIPP *et al.*, 2023). Em segundo lugar, o levantamento também indicou que as experiências dos membros de cada equipe também são consideradas para selecionar as medições. Tal estratégia mostra-se favorável para as necessidades internas e recorrentes de adaptação das equipes (KORPIVAARA; TUUNANEN; SEPPÄNEN, 2021). Finalmente, pesquisas na literatura para encontrar métricas adequadas também são realizadas, conforme evidenciado por Leal *et al.* (LEAL *et al.*, 2022). A presente pesquisa corrobora esses achados ao identificar métricas amplamente utilizadas e valorizadas no contexto da indústria de software brasileira, oferecendo evidências adicionais de que a seleção e aplicação de métricas são fundamentais para o sucesso de projetos ágeis em larga escala.

Os achados deste estudo também estão em consonância com a literatura existente ao enfatizar o forte interesse da gestão nas métricas ágeis, pois sua aplicação eficaz pode transformar a gestão de projetos de software em larga escala, proporcionando *insights* valiosos para assegurar a tomada de decisões estratégicas (PHILIPP et al., 2023). Além disso, a gestão deve ter o compromisso de disseminar as métricas entre as demais partes interessadas da organização, pois somente uma abordagem colaborativa da organização, gestão e as equipes pode resultar no uso bem sucedido das métricas (ERTABAN; SARIKAYA; BAGRIYANIK, 2018; PHILIPP et al., 2023). Para além do protagonismo da gestão, a pesquisa também assegurou a existência de métricas que tinham como destinatária a própria equipe ágil. O foco na equipe ágil alinha-se com os princípios e valores ágeis do manifesto (BECK et al., 2020). Por fim, também houve destaque para os clientes como utilizadores da métrica, o que além de ser coerente com o modelo de negócios de certas empresas de consultoria na amostra, também alinha-se ao princípio da satisfação do cliente.

Os profissionais sinalizaram que as empresas que adotam essas práticas e métricas atingem melhorias significativas no monitoramento do fluxo de trabalho, metas de desempenho organizacionais e qualidade do produto. A pesquisa realizada por Greening (GREENING, 2010) já havia discutido como métricas podem fornecer a base para relatórios usados por treinadores para melhorar o desempenho das equipes. Grimaldi et al. (GRIMALDI et al., 2016) também apresentaram um caso de implementação do SAFe em uma grande empresa, enriquecido com métricas que impactaram nos prazos de entrega do produto e na redução de custos. Ainda, Tabib (TABIB, 2013) elencou métricas essenciais para colaboração e alinhamento, que forneceram visão sobre qualidade do produto e risco sem comprometer a velocidade no ágil em larga escala. As métricas identificadas também servem como ferramentas para a identificação de gargalos e ineficiências, possibilitando ações corretivas mais precisas. Além disso, a utilização de métricas permite uma melhor comunicação e alinhamento entre as equipes e a gestão, promovendo uma cultura de transparência e responsabilidade. Isso é especialmente relevante em ambientes de desenvolvimento ágil em larga escala, onde a coordenação e a integração entre múltiplas equipes são desafiadoras. A adoção de métricas apropriadas pode facilitar a implementação de metodologias ágeis, assegurando que os objetivos do projeto sejam atingidos de maneira eficiente e eficaz.

6.1 Limitações e Ameaças à Validade

Nesta seção são elencadas e discutidas as limitações e ameaças à validade desta pesquisa:

- a principal limitação da revisão *multivocal* é que, exceto pela seleção de artigos acadêmicos, todas as outras fases foram conduzidas exclusivamente pelo autor

e validadas pelos orientadores. Para mitigar esse desafio, seguimos as diretrizes de Garousi *et al.* (GAROUSI; FELDERER; MÄNTYLÄ, 2019) para incorporar literatura cinza e conduzir uma revisão *multivocal* em engenharia de software, que foram adotadas e validadas;

- seguindo a estrutura proposta por (WOHLIN *et al.*, 2012), a ameaça geral à validade do construto do *survey* é que algumas perguntas do questionário incluíram uma lista de respostas pré-definidas, como, por exemplo, sobre estratégias de seleção, quem faz uso e para quais finalidades as métricas são usadas. Assim, os respondentes podem ter optado pela facilidade de escolher uma dessas respostas, em vez de optar por preencher um campo de texto aberto, para relatar e descrever outras opções;
- para avaliação da frequência de uso das métricas foi disponibilizado no questionário um guia de métricas (Apêndice B) no Google Docs por link contendo as fórmulas de cálculo caso houvesse dúvidas sobre do que se tratavam apenas pela nomenclatura. Contudo, parte dos respondentes podem ter hesitado em usá-lo, gerando algumas respostas imprecisas. Por isso, nas análises optou-se por focar nas tendências para os tipos de medições mais utilizadas para cada objetivo de negócio (ameaça à validade de constructo);
- como ameaça à validade externa do estudo, a generalização de um único estudo para todos os casos é sempre uma possível ameaça, não se pode afirmar que a amostra representa a população brasileira de profissionais que atuam com agilidade escalada. No entanto, alcançou-se uma distribuição relativamente ampla de domínios e funções de participantes. Compreende-se que a amostra é relevante e isso permite fazer observações independentes desses fatores;
- também não é possível generalizar os resultados para desenvolvedores de outros países, visto existem modelos de negócio diferentes. No entanto, o principal objetivo foi detectar em um estudo exploratório evidências preliminares relevantes (ameaça à validade externa do estudo);
- a preparação e limpeza dos dados podem representar uma ameaça à validade interna do estudo pois erros podem ser introduzidos. Visto que, todas as etapas de código das análises foram realizadas e revisadas apenas pelo autor do trabalho. Devido a esse processo de revisão, há segurança de que o método é confiável e reproduzível. Ameaças à análise qualitativa podem ser a avaliação incompleta de temas relevantes.

6.2 Trabalhos Futuros

As seguintes linhas de base para trabalhos futuros podem ser consideradas:

- realização de estudos de caso em organizações que implementam o *Disciplined Agile* como um *framework* ágil escalável para coletar dados sobre métricas utilizadas, pois não foram relatadas na pesquisa;
- estudos futuros são importantes para detalhar e analisar todas as métricas em dois grupos, um no qual podem ser utilizadas em projetos ágeis de qualquer escala e outro no qual são aplicáveis apenas a contextos escalados, além de destacar quais tipos de equipes ou níveis organizacionais podem fazer uso de cada tipo de medição;
- nesta pesquisa, as métricas de sustentabilidade de práticas ágeis relacionadas a fatores humanos encontradas na literatura apresentaram baixa popularidade na indústria. Essa primeira constatação aponta para oportunidade de investigar mais profundamente na literatura e indústria a existência, escopo e frequência de uso nas organizações de métricas que envolvam apenas fatores emocionais e humanos no ágil em grande escala;
- futuramente, também planeja-se analisar o *feedback* dos respondentes para verificar se eles concordam com os resultados deste trabalho, por meio de um estudo longitudinal. Isso poderá ser feito repetindo a pesquisa, apresentando os resultados obtidos neste estudo e avaliando o que mudou, o que piorou e o que melhorou.

6.3 Considerações Finais do Capítulo

Este capítulo apresentou uma discussão da revisão *multivocal* focada nos resultados obtidos, delineando o conhecimento extraído da literatura cinza sobre a utilização de métricas no desenvolvimento ágil de software em larga escala. Foi destacada a relevância dos objetivos comuns para o uso das métricas e resultados alcançados documentados. As respostas coletadas com o questionário permitiram discutir o *survey* com base nas percepções dos participantes sobre como tais medições são utilizadas de fato em seus contextos de trabalho. Os dados coletados levaram à descobertas das principais práticas, métricas efetivamente utilizadas com base em objetivos de negócio e impactos relacionados as medições adotadas no desenvolvimento ágil escalonado. Necessariamente, todos os pontos em discussão foram fundamentados à literatura existente. Também foram abordadas as principais limitações e ameaças da pesquisa, proporcionando uma visão crítica e abrangente da validade dos resultados obtidos. Finalmente, foram sugeridos os possíveis direcionamentos para pesquisas futuras sobre o tema.

7 Conclusões

O desenvolvimento de software está presente na maioria dos setores, incluindo serviços, manufatura, saúde e até segurança. Grandes projetos estão cada vez mais adotando práticas de desenvolvimento ágil. Logo, a indústria de desenvolvimento de software de larga escala necessita de métricas para dar suporte à coordenação de várias equipes trabalhando em conjunto, promover melhoria contínua e monitorar o progresso.

Além disso, as medições são essenciais para identificar desafios, promover a transparência e orientar a tomada de decisões. Suas análises impactam significativamente na melhoria e no sucesso das iniciativas ágeis empresariais em diversas áreas, proporcionando uma visão clara do progresso, facilitando a colaboração entre equipes e garantindo consistência nas entregas. Ao fornecer um contexto valioso para a tomada de decisões baseada em dados, essas métricas representam uma ferramenta fundamental para o sucesso e a eficácia do desenvolvimento ágil em larga escala. Considerando os diversos resultados e público-alvo, a seleção e personalização de métricas por parte de uma empresa muitas vezes depende de fatores como o tamanho organizacional, aliança com seus objetivos de negócios, estrutura operacional e requisitos específicos. Os *insights* deste estudo equiparão pesquisadores e profissionais com o conhecimento necessário para aprofundar os desafios relacionados às métricas em ambientes ágeis de grande escala, promovendo o desenvolvimento de soluções mais direcionadas.

O presente trabalho teve como objetivo consolidar descobertas chave sobre métricas de agilidade em larga escala, melhorando a visibilidade e acessibilidade desse conhecimento para a comunidade científica e equipes ágeis. Previamente, foram identificadas 19 fontes de literatura cinza por meio de uma Revisão *Multivocal* da Literatura que englobou a literatura científica reunida por um mapeamento sistemático antecedente. A revisão preliminar forneceu informações sobre métricas, motivações e resultados em processos ágeis escaláveis. Posteriormente, foram apresentados os resultados de um *survey* conduzido com a contribuição de 145 profissionais do mercado de software brasileiro, buscando entender empiricamente o estado da prática da utilização de métricas nas organizações de software que implementam agilidade escalada. Essa pesquisa foi orientada por descobertas da revisão *multivocal* e buscou, além de solidificar o achados, realizar novas descobertas relevantes. Este é um dos poucos estudos de que se tem conhecimento até agora que explora o universo prático de utilização das métricas no escalonamento ágil de maneira abrangente com base na percepção dos profissionais.

7.1 Contribuições

As principais contribuições desse trabalho são:

- buscou explorar e dar visibilidade à comunidade científica das evidências do campo das métricas, soluções e adaptações de métricas no desenvolvimento ágil de software em larga escala. Um resumo do estado da arte sobre o tema com informações sintetizadas de artigos publicados entre 2010 e 2022 foi fornecido (MENEZES; MARINHO; SAMPAIO, 2022);
- contribuiu para o entendimento de métodos e *frameworks* ágeis em larga escala, realizando uma compilação e análise de 80 métricas vinculadas identificadas nas fontes de literatura (MENEZES; MARINHO; SAMPAIO, 2024b);
- foram consideradas as etapas do ciclo de vida do desenvolvimento de software para mapear todas as métricas (MENEZES; MARINHO; SAMPAIO, 2024b);
- o conjunto das métricas foi categorizado usando o modelo *Goal Question Metric*, no qual foram relacionadas a objetivos organizacionais, incluindo resposta e entrega de valor ao cliente, melhoria da produtividade da equipe, garantia da qualidade do produto e promoção de práticas sustentáveis (MENEZES; MARINHO; SAMPAIO, 2024b). Na modelagem, foram apresentadas as descrições e fórmulas de cálculo das métricas. Notavelmente, a maioria das métricas catalogadas no GQM foca no monitoramento da funcionalidade geral do processo e no aumento da produtividade, com apenas uma minoria abordando a medição da entrega de valor;
- também foram comentados os resultados descritos na literatura, obtidos pela aplicação das métricas no ágil escalado (MENEZES; MARINHO; SAMPAIO, 2024b);
- para o *survey* apresentado neste trabalho, a questão de pesquisa principal foi dividida em três questões secundárias e os resultados apresentados geraram um conjunto de oito descobertas. Essas descobertas geram os seguintes indícios: as métricas utilizadas no ágil escalado são predominantemente selecionadas e definidas pelas organizações ou processos e destinadas a serem utilizadas para orientar a tomada de decisão da gestão;
- a avaliação de frequência de usabilidade de métricas usando escala *Likert* destacou a força de métricas com foco em medir o fluxo de entregas pontualmente. O valor entregue demonstra ser mensurado internamente a partir do tempo total do fluxo em que as equipes gastam em atividades de trabalho de valor agregado, além do quanto bem podem planejá-las e atingir seus objetivos de valor. Medir produtividade nesse ambiente ainda significa principalmente monitorar o desempenho contínuo das equipes. Grandes organizações pouco priorizam medir a eficiência de suas práticas

ágeis, formas de compreender os limites possíveis de trabalho em progresso são exceção. Monitorar problemas ao longo dos incrementos, qualidade do código e a resolução de problemas dos sistemas são os aspectos prioritários das métricas de qualidade do produto. É importante ressaltar que essas foram as medidas mais tendenciosas no ágil em larga escala, outros aspectos também são mensurados, porém com menos frequência;

- resultados apontam que as áreas mais impactadas pela utilização das métricas nas grandes organizações são: o monitoramento do fluxo de trabalho, as metas de desempenho da organização, a qualidade do produto e o escalonamento ágil para além das equipes de desenvolvimento;
- também foram coletadas e expostas novas métricas usadas no contexto que não foram identificadas nas fontes de literatura. Algumas análises a respeito de suas aplicações nas empresas de software foram apresentadas e discutidas;
- finalmente, como um resumo das lições aprendidas neste estudo, a análise dos resultados de ambas as etapas de pesquisa resultou na sugestão de um catálogo de métricas eficazes para ser utilizado por grandes organizações de software que implementam agilidade em larga escala e orientar futuras pesquisas.

Referências

- AMBLER, S.; LINES, M. Going beyond scrum: disciplined agile delivery. *Disciplined Agile Consortium. White Paper Series*, p. 1–16, 2013. Citado na página 5.
- AMBLER, S. W.; LINES, M. *Disciplined agile delivery: A practitioner's guide to agile software delivery in the enterprise*. [S.l.]: IBM press, 2012. Citado na página 6.
- AMPATZOGLU, A. et al. The financial aspect of managing technical debt: A systematic literature review. *Information and Software Technology*, Elsevier, v. 64, p. 52–73, 2015. Citado na página 14.
- BARCELLOS, M. P.; ROCHA, A. R. C. da. Avaliação de bases de medidas considerando sua aplicabilidade ao controle estatístico de processos de software. In: SBC. *Anais do VII Simpósio Brasileiro de Qualidade de Software*. [S.l.], 2008. p. 75–89. Citado na página 4.
- BASILI, V. R. *Software modeling and measurement: the Goal/Question/Metric paradigm*. [S.l.], 1992. Citado na página 37.
- BECK, K. et al. Agile manifesto, 2001. URL <http://www.agilemanifesto.org>, 2020. Citado 2 vezes nas páginas 54 e 76.
- BRITTO, R.; USMAN, M.; MENDES, E. Effort estimation in agile global software development context. In: SPRINGER. *Agile Methods. Large-Scale Development, Refactoring, Testing, and Estimation: XP 2014 International Workshops, Rome, Italy, May 26-30, 2014, Revised Selected Papers 15*. [S.l.], 2014. p. 182–192. Citado 2 vezes nas páginas 8 e 9.
- BROWN, A. W. A case study in agile-at-scale delivery. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2011. p. 266–281. Citado 16 vezes nas páginas 24, 25, 26, 27, 32, 34, 35, 36, 37, 42, 62, 63, 68, 69, 70 e 71.
- BURKARD, M. F. et al. Catálogo de melhores práticas de gestão de pessoas em projetos de desenvolvimento de software. Universidade Federal de Santa Maria, 2024. Citado na página 60.
- CAMARA, R. et al. Agile global software development: A systematic literature review. In: *Proceedings of the XXXIV Brazilian Symposium on Software Engineering*. [S.l.: s.n.], 2020. p. 31–40. Citado na página 5.
- CAMARA, R.; MARINHO, M.; MOURA, H. Agile tailoring with scaling agile frameworks in distributed large-scale settings. In: SBC. *Anais Estendidos do XX Simpósio Brasileiro de Qualidade de Software*. [S.l.], 2021. p. 12–17. Citado 2 vezes nas páginas 1 e 2.
- CHLOROS, D.; GEROGIANNIS, V. C.; KAKARONTZAS, G. Use of software and project management metrics in agile software development methodologies: A systematic mapping study. In: *Proceedings of the 2022 European Symposium on Software Engineering*. [S.l.: s.n.], 2022. p. 25–32. Citado 3 vezes nas páginas 4, 19 e 37.

- COPADO, T. *How to Optimize Flow Metrics Quality in Scaled Agile Framework (SAFe)*. 2022. <<https://www.copado.com/devops-hub/blog/how-to-optimize-flow-metrics-quality>>. (accessed 4 January 2023). Citado 4 vezes nas páginas 35, 36, 39 e 42.
- CRESWELL, J. W.; CRESWELL, J. *Research design, qualitative, quantitative, and mixed methods approaches (I)*. Sage Publications. [S.l.]: Inc, 2014. Citado na página 12.
- DINGSØYR, T.; FÆGRI, T. E.; ITKONEN, J. What is large in large-scale? a taxonomy of scale for agile software development. In: SPRINGER. *Product-Focused Software Process Improvement: 15th International Conference, PROFES 2014, Helsinki, Finland, December 10-12, 2014. Proceedings 15*. [S.l.], 2014. p. 273–276. Citado 2 vezes nas páginas 5 e 33.
- DINGSØYR, T.; MOE, N. B. Towards principles of large-scale agile development. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2014. p. 1–8. Citado 3 vezes nas páginas 1, 2 e 75.
- DYBA, T.; DINGSOYR, T.; HANSEN, G. K. Applying systematic reviews to diverse study types: An experience report. In: IEEE. *First international symposium on empirical software engineering and measurement (ESEM 2007)*. [S.l.], 2007. p. 225–234. Citado na página 16.
- EDISON, H.; WANG, X.; CONBOY, K. Comparing methods for large-scale agile software development: A systematic literature review. *IEEE Transactions on Software Engineering*, IEEE, 2021. Citado 5 vezes nas páginas 2, 6, 8, 9 e 33.
- ERTABAN, C.; SARIKAYA, E.; BAGRIYANIK, S. Agile performance indicators for team performance evaluation in a corporate environment. In: *Proceedings of the 19th International Conference on Agile Software Development: Companion*. [S.l.: s.n.], 2018. p. 1–3. Citado 5 vezes nas páginas 1, 2, 4, 5 e 76.
- FENTON, N.; BIEMAN, J. *Software metrics: a rigorous and practical approach*. [S.l.]: CRC press, 2014. Citado 3 vezes nas páginas 1, 4 e 60.
- GAROUSHI, V.; FELDERER, M.; MÄNTYLÄ, M. V. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology*, Elsevier, v. 106, p. 101–121, 2019. Citado 3 vezes nas páginas 14, 16 e 77.
- GREENING, D. R. Enterprise scrum: Scaling scrum to the executive level. In: IEEE. *2010 43rd Hawaii International Conference on System Sciences*. [S.l.], 2010. p. 1–10. Citado 14 vezes nas páginas 24, 25, 26, 27, 32, 33, 34, 35, 36, 37, 39, 42, 43 e 76.
- GREENING, D. R. Agile enterprise metrics. In: IEEE. *2015 48th Hawaii International Conference on System Sciences*. [S.l.], 2015. p. 5038–5044. Citado 15 vezes nas páginas 1, 25, 27, 32, 33, 34, 35, 36, 37, 39, 42, 43, 62, 68 e 69.
- GRIMALDI, P. et al. An agile, measurable and scalable approach to deliver software applications in a large enterprise. *International Journal of Agile Systems and Management*, Inderscience Publishers (IEL), v. 9, n. 4, p. 326–339, 2016. Citado 16 vezes nas páginas 1, 2, 5, 25, 26, 29, 32, 34, 35, 36, 39, 42, 50, 69, 70 e 76.

- HEIDENBERG, J. et al. A metrics model to measure the impact of an agile transformation in large software development organizations. In: SPRINGER. *Int. Conference on Agile Software Development*. [S.l.], 2013. p. 165–179. Citado 17 vezes nas páginas 25, 26, 29, 32, 34, 35, 36, 38, 42, 50, 62, 64, 65, 67, 73, 74 e 75.
- HOSSAIN, S. S.; AHMED, P.; ARAFAT, Y. Software process metrics in agile software development: A systematic mapping study. In: SPRINGER. *Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part IX 21*. [S.l.], 2021. p. 15–26. Citado 4 vezes nas páginas 4, 5, 53 e 75.
- KETTUNEN, P. Realizing agile software enterprise transformations by team performance development. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2014. p. 285–293. Citado 7 vezes nas páginas 1, 14, 25, 26, 29, 32 e 75.
- KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Citeseer, 2007. Citado na página 13.
- KNIBERG, H.; IVARSSON, A. Scaling agile@ spotify. *online*], *UCVOF, ucvox. files.wordpress.com/2012/11/113617905-scaling-Agile-spotify-11.pdf*, 2012. Citado na página 6.
- KORPIVAARA, I.; TUUNANEN, T.; SEPPÄNEN, V. Performance measurement in scaled agile organizations. In: UNIVERSITY OF HAWAII AT MANOA. *Proceedings of the Annual Hawaii International Conference on System Sciences*. [S.l.], 2021. Citado 10 vezes nas páginas 2, 4, 5, 6, 9, 19, 37, 53, 54 e 75.
- KORSON, T. *Cash-in Comfort*. 2015. <<https://less.works/case-studies/cash-in-comfort>>. (accessed 4 January 2023). Citado 6 vezes nas páginas 34, 35, 36, 40, 71 e 72.
- KUPIAINEN, E.; MÄNTYLÄ, M. V.; ITKONEN, J. Using metrics in agile and lean software development—a systematic literature review of industrial studies. *Information and software technology*, Elsevier, v. 62, p. 143–163, 2015. Citado 2 vezes nas páginas 4 e 5.
- LAANTI, M. Characteristics and principles of scaled agile. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2014. p. 9–20. Citado 8 vezes nas páginas 5, 25, 26, 29, 32, 34, 35 e 36.
- LARMAN, C.; VODDE, B. *Large-scale scrum: More with LeSS*. [S.l.]: Addison-Wesley Professional, 2016. Citado na página 6.
- LEAL, S. et al. How agile organizations use metrics: A systematic literature mapping. In: *Proceedings of the XXI Brazilian Symposium on Software Quality*. [S.l.: s.n.], 2022. p. 1–11. Citado 11 vezes nas páginas 1, 4, 5, 7, 8, 9, 19, 37, 53, 60 e 75.
- LINÅKER, J. et al. Guidelines for conducting surveys in software engineering. Department of Computer Science, Lund University, 2015. Citado 3 vezes nas páginas 19, 20 e 21.
- LIYANAGE, J. R. *Lean Agile Metrics for Scaled Agile Systems*. 2014. <<https://www.methodsandtools.com/archive/leanagilemetrics.php>>. (accessed 4 January 2023). Citado 6 vezes nas páginas 35, 36, 39, 42, 50 e 73.

- LOMIO, F. et al. On the benefits of the accelerate metrics: An industrial survey at vendasta. In: IEEE. *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. [S.l.], 2022. p. 46–50. Citado na página 4.
- MARCONI, M. d. A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. [S.l.]: Atlas, 2003. Citado na página 13.
- MARINHO, M.; CAMARA, R.; SAMPAIO, S. Toward unveiling how safe framework supports agile in global software development. *IEEE Access*, IEEE, v. 9, p. 109671–109692, 2021. Citado na página 1.
- MENEZES, R. *Dados do survey “Métricas para Desenvolvimento Ágil em Larga Escala: Um Levantamento da Literatura e da Indústria de Software Brasileira”*. Harvard Dataverse, 2024. Disponível em: <<https://doi.org/10.7910/DVN/RLFJB5>>. Citado na página 21.
- MENEZES, R.; MARINHO, M.; SAMPAIO, S. *Complete metrics tables for “Goal Question Metrics” modeling*. [S.l.]: Harvard Dataverse, 2024. <<https://doi.org/10.7910/DVN/HTQTEE>>. Citado na página 37.
- MENEZES, R.; MARINHO, M.; SAMPAIO, S. Metrics in large-scale agile software development: A multivocal literature review. In: SBC. *Congresso Ibero-Americano em Engenharia de Software (CIbSE)*. [S.l.], 2024. p. 106–120. Citado 5 vezes nas páginas 11, 48, 49, 50 e 80.
- MENEZES, R.; MARINHO, M. L.; SAMPAIO, S. Metrics in large-scale agile software development: A systematic mapping of literature. In: *International Conference on Information Systems and Technology Management (CONTECSI)*. [s.n.], 2022. Disponível em: <<https://doi.org/10.5748/19CONTECSI/PSE/PRM/7002>>. Citado 2 vezes nas páginas 11 e 80.
- MONTANO, R. R. et al. A countrywide descriptive survey of agile software development in brazil. *Agile Processes in Software Engineering and Extreme Programming*, p. 185, 2022. Citado 2 vezes nas páginas 1 e 2.
- OGAWA, R. T.; MALEN, B. Towards rigor in reviews of multivocal literatures: Applying the exploratory case study method. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 61, n. 3, p. 265–286, 1991. Citado 2 vezes nas páginas 13 e 14.
- PETERSEN, K. An empirical study of lead-times in incremental and agile software development. In: SPRINGER. *International Conference on Software Process*. [S.l.], 2010. p. 345–356. Citado na página 67.
- PETERSEN, K. et al. Systematic mapping studies in software engineering. In: *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*. [S.l.: s.n.], 2008. p. 1–10. Citado 2 vezes nas páginas 13 e 24.
- PETERSEN, K.; WOHLIN, C. Measuring the flow in lean software development. *Software: Practice and experience*, Wiley Online Library, v. 41, n. 9, p. 975–996, 2011. Citado na página 65.
- PHILIPP, P. et al. A method for metric management at a large-scale agile software development organization. 2023. Citado 7 vezes nas páginas 2, 7, 9, 53, 54, 75 e 76.

- PRODANOV, C. C.; FREITAS, E. C. D. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. [S.l.]: Editora Feevale, 2013. Citado 2 vezes nas páginas 11 e 12.
- RAZZAK, M. A. et al. Transition from plan driven to safe@: periodic team self-assessment. In: SPRINGER. *International Conference on Product-Focused Software Process Improvement*. [S.l.], 2017. p. 573–585. Citado 11 vezes nas páginas 5, 6, 25, 26, 30, 32, 34, 35, 36, 39 e 42.
- REINERTSEN, D. G. *The principles of product development flow: second generation lean product development*. [S.l.]: Celeritas Redondo Beach, 2009. v. 62. Citado na página 64.
- ROGER, S. P.; BRUCE, R. M. *Software engineering: a practitioner's approach*. [S.l.]: McGraw-Hill Education, 2015. Citado na página 37.
- SAFE. *Measure and Grow*. 2022. <<https://scaledagileframework.com/measure-and-grow/>>. (accessed 4 January 2023). Citado 9 vezes nas páginas 6, 35, 36, 39, 49, 65, 67, 68 e 69.
- SCHWABER, K. Nexus guide. the definitive guide to nexus: The exoskeleton of scaled scrum development. Available from: *scrum.org*, 2015. Citado na página 6.
- SCRUM@SCALE. *The Scrum At Scale® Guide. The Definitive Guide to the Scrum@Scale Framework*. 2022. <<https://www.scrumatscale.com/scrum-at-scale-guide-online/>>. (accessed 4 January 2023). Citado 6 vezes nas páginas 6, 35, 36, 37, 39 e 43.
- SHAMEEM, M. et al. Systematic review of success factors for scaling agile methods in global software development environment: A client-vendor perspective. In: IEEE. *2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*. [S.l.], 2017. p. 17–24. Citado 2 vezes nas páginas 5 e 6.
- STARON, M.; MEDING, W.; CAIMAN, M. Improving completeness of measurement systems for monitoring software development workflows. In: SPRINGER. *International Conference on Software Quality*. [S.l.], 2013. p. 230–243. Citado 8 vezes nas páginas 25, 28, 32, 34, 35, 36, 38 e 42.
- STARON, M.; MEDING, W.; PALM, K. Release readiness indicator for mature agile and lean software development projects. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2012. p. 93–107. Citado 8 vezes nas páginas 25, 28, 32, 34, 35, 36, 39 e 42.
- STETTINA, C. J.; SCHOEMAKER, L. Reporting in agile portfolio management: Routines, metrics and artefacts to maintain an effective oversight. In: SPRINGER, CHAM. *International Conference on Agile Software Development*. [S.l.], 2018. p. 199–215. Citado 11 vezes nas páginas 25, 26, 31, 32, 33, 34, 35, 36, 39, 67 e 69.
- TABIB, R. Need 4 speed: leverage new metrics to boost your velocity without compromising on quality. In: IEEE. *2013 Agile Conference*. [S.l.], 2013. p. 117–120. Citado 12 vezes nas páginas 24, 25, 27, 32, 34, 35, 36, 38, 42, 50, 72 e 76.
- TESSAROLO, F. et al. Developing ambient assisted living technologies exploiting potential of user-centred co-creation and agile methodology: the captain project experience. *Journal of Ambient Intelligence and Humanized Computing*, Springer, p. 1–16, 2022. Citado 9 vezes nas páginas 25, 27, 28, 32, 34, 35, 36, 38 e 42.

- THAWABA, A. A. et al. A mechanism to support agile frameworks enhancing reliability assessment for scs development: a case study of medical surgery departments. In: SPRINGER. *International Conference on Soft Computing and Data Mining*. [S.l.], 2020. p. 66–76. Citado 14 vezes nas páginas [25](#), [26](#), [30](#), [32](#), [33](#), [34](#), [35](#), [36](#), [39](#), [42](#), [63](#), [64](#), [66](#) e [67](#).
- TRIPATHI, N. et al. Scaling kanban for software development in a multisite organization: challenges and potential solutions. In: SPRINGER. *International Conference on Agile Software Development*. [S.l.], 2015. p. 178–190. Citado 16 vezes nas páginas [1](#), [2](#), [25](#), [26](#), [31](#), [32](#), [33](#), [34](#), [35](#), [36](#), [39](#), [43](#), [50](#), [62](#), [71](#) e [72](#).
- WAGNER, S. et al. Challenges in survey research. *Contemporary Empirical Methods in Software Engineering*, Springer, p. 93–125, 2020. Citado na página [13](#).
- WIERINGA, R. et al. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, Springer, v. 11, p. 102–107, 2006. Citado na página [24](#).
- WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer Science & Business Media, 2012. Citado 2 vezes nas páginas [20](#) e [77](#).

A Apêndice - Formulário de Pesquisa.

Métricas no Desenvolvimento de Software Ágil em Larga Escala

B *I* U ↻ ✕

À medida que os projetos de grande porte adotam práticas ágeis, surge a necessidade de métricas para coordenar equipes, promover melhorias contínuas e monitorar o progresso. Neste trabalho de conclusão do mestrado em Engenharia de Software, busco compreender e oferecer uma visão abrangente das métricas no desenvolvimento ágil em ambientes escaláveis, proporcionando insights valiosos para a comunidade.

Se você faz parte de um time ágil, gostaria de pedir a disponibilidade de 10 minutos do seu tempo para nos ajudar com essa pesquisa. Durante a leitura das perguntas e afirmações tenha sempre em mente suas experiências dentro do seu respectivo projeto.

O objetivo dessa pesquisa é acadêmico, sua contribuição será muito valiosa e utilizada exclusivamente para a análise dos resultados e futuros estudos na área. Ao responder as questões e submeter suas respostas você permite que os pesquisadores abaixo nomeados façam uso de suas respostas na análise do estudo.

Caso tenha alguma dúvida entre em contato pelo e-mail: renato.vmenezes@ufrpe.br

Pesquisadores:

Renato Vieira de Menezes (Mestrando em Engenharia de Software - UFRPE)

Marcelo Luiz Monteiro Marinho (Departamento de Computação - UFRPE)

Suzana Cândido de Barros Sampaio (Departamento de Computação - UFRPE)

E-mail *

E-mail válido

Este formulário está coletando e-mails. [Alterar configurações](#)

Declaro que concordo em participar dessa pesquisa e autorizo o uso dos resultados deste estudo para atividades acadêmicas e científicas. *

Sim

Não

Aplicação de Métricas

Nesta seção apresentaremos algumas questões relacionadas às aplicações das métricas utilizadas no desenvolvimento de software ágil em larga escala. Por favor, responda sobre.

Como as métricas ágeis são selecionadas no seu projeto ou time? *

- Literatura
- Por meio da experiência dos membros da equipe
- Definidas pela organização/processo?
- Outros...

Para quais finalidades sua empresa utiliza métricas no desenvolvimento de software ágil? *

- Apoiar o escalonamento ágil para além das equipes de desenvolvimento de software
- Apoiar equipes de desenvolvimento situadas em diferentes localizações geográficas
- Apoiar transformações organizacionais
- Apoiar o monitoramento do fluxo de trabalho
- Apoiar as metas de desempenho da organização
- Apoiar o monitoramento de recursos e custos no processo
- Apoiar a sustentabilidade de práticas
- Apoiar a qualidade do produto
- Outros...

Quem faz uso dos resultados da medição? *

- A equipe ágil
- A gestão
- O cliente
- Outros...

Métricas Ágeis



Nesta seção apresentaremos uma variedade de métricas e instrumentos com métricas incorporadas existentes no desenvolvimento de software ágil em larga escala. Por favor, informe a frequência de utilização no seu projeto ou time. Reflita sobre o seu dia-a-dia de trabalho e classifique-as de 1 a 5, com 1 para "não uso no meu projeto" e 5 sendo "sempre uso no meu projeto". [Clique aqui \(Guia de Métricas\)](#) caso possua dúvidas e deseje consultar a fórmula de cálculo das métricas.

Com que frequência você utiliza as seguintes métricas relacionadas à responsividade aos clientes? *

	Não uso	Uso raramente	As vezes uso	Uso frequente...	Sempre uso
Tempo gasto d...	<input type="radio"/>				
Tempo gasto d...	<input type="radio"/>				
True Sprint Len...	<input type="radio"/>				
Horizonte de pr...	<input type="radio"/>				
Lead Time	<input type="radio"/>				
Tempo de resp...	<input type="radio"/>				
Tempo de ciclo...	<input type="radio"/>				
Distribuição de...	<input type="radio"/>				
Indicador de pr...	<input type="radio"/>				
Tempo por fun...	<input type="radio"/>				
Tempo restante	<input type="radio"/>				
Tempo gasto (t...	<input type="radio"/>				
Tempo gasto (t...	<input type="radio"/>				
Tempo disponí...	<input type="radio"/>				
Tempo corrido	<input type="radio"/>				
Tempo de espe...	<input type="radio"/>				
Preparação par...	<input type="radio"/>				
Índice de previ...	<input type="radio"/>				

Com que frequência você utiliza as seguintes métricas relacionadas à geração de valor para os clientes? *

	Não uso	Uso raramente	As vezes uso	Uso frequente...	Sempre uso
Valor presente ...	<input type="radio"/>				
Valor comercia...	<input type="radio"/>				
Net Promoter S...	<input type="radio"/>				
Eficiência de fl...	<input type="radio"/>				
Previsibilidade ...	<input type="radio"/>				

Com que frequência você utiliza as seguintes métricas relacionadas à produtividade? *

	Não uso	Uso raramente	As vezes uso	Uso frequente...	Sempre uso
Velocidade	<input type="radio"/>				
Tarefas bloque...	<input type="radio"/>				
Pontos de funç...	<input type="radio"/>				
Gráfico burndo...	<input type="radio"/>				
Gráfico de burn...	<input type="radio"/>				
Desvio de velo...	<input type="radio"/>				
Contagem de d...	<input type="radio"/>				
Eficiência	<input type="radio"/>				
Impedimentos	<input type="radio"/>				
Pulso de com...	<input type="radio"/>				
Velocidade pla...	<input type="radio"/>				
Custo para cad...	<input type="radio"/>				
Custo restante	<input type="radio"/>				
Custo das des...	<input type="radio"/>				

Funções restan...	<input type="radio"/>				
Funções concl...	<input type="radio"/>				
Custo disponível	<input type="radio"/>				
Frequência de i...	<input type="radio"/>				
Aceleração	<input type="radio"/>				

Com que frequência você utiliza as seguintes métricas relacionadas à sustentabilidade das práticas ágeis? *

	Não uso	Uso raramente	As vezes uso	Uso frequente...	Sempre uso
Pulso de equip...	<input type="radio"/>				
Delta de precis...	<input type="radio"/>				
Eficácia	<input type="radio"/>				
Custo por hora ...	<input type="radio"/>				
Custo real virtu...	<input type="radio"/>				
Desperdício	<input type="radio"/>				
Custo delta	<input type="radio"/>				
Indicador de in...	<input type="radio"/>				
Eficácia perceb...	<input type="radio"/>				
Envolvimento d...	<input type="radio"/>				
Trabalho em pr...	<input type="radio"/>				
Carga de fluxo	<input type="radio"/>				

Com que frequência você utiliza as seguintes métricas relacionadas à qualidade do produto? *

	Não uso	Uso raramente	As vezes uso	Uso frequente...	Sempre uso
Taxa de defeito...	<input type="radio"/>				
Números de pr...	<input type="radio"/>				
Dias abertos c...	<input type="radio"/>				
Linhas de códi...	<input type="radio"/>				
Número de arq...	<input type="radio"/>				
Linhas de códi...	<input type="radio"/>				
Número de des...	<input type="radio"/>				
Cobertura de t...	<input type="radio"/>				
Taxa de aprova...	<input type="radio"/>				
Número de def...	<input type="radio"/>				
Taxa de falhas ...	<input type="radio"/>				
Tempo de rest...	<input type="radio"/>				
Taxa de aprova...	<input type="radio"/>				
Taxa de defeito...	<input type="radio"/>				
Taxa de defeito	<input type="radio"/>				
Tempo de inati...	<input type="radio"/>				
Dívida técnica	<input type="radio"/>				

Alguma métrica que você usa no dia-a-dia não foi apresentada?

Texto de resposta longa

.....

Dados Demográficos

Nesta seção você responderá algumas perguntas gerais sobre seu projeto, sua empresa e você que nos ajudarão a identificar e categorizar melhor os dados extraídos deste formulário.

Quais métodos/frameworks ágeis em larga escala você utiliza em seu projeto na sua empresa? *

- Scrum
- Kanban
- XP
- Agile
- Lean
- SAFe
- Scrum-of-scrum
- Scrum-at-scale
- Spotify
- Nexus
- DA
- Outros...

Qual é o tamanho da sua empresa? *

- Micro-empresa (até 9 colaboradores)
- Pequena empresa (entre 10 e 49 colaboradores)
- Média empresa (entre 50 e 99 colaboradores)
- Grande empresa (100 ou mais colaboradores)

Quantas equipes no seu projeto ou time? *

- Pequena escala (1 equipe)
- Larga Escala (entre 2 e 9 equipes)
- + Larga escala (maior que 9 equipes)

Em qual dos cenários de distribuição se enquadra seu projeto ou time? *

- Localmente distribuído (pessoas trabalhando remotamente da mesma cidade)
- Nacionalmente distribuído (pessoas trabalhando remotamente de cidades diferentes de um mesmo país)
- Continentalmente distribuído (pessoas trabalhando remotamente de países diferentes que estão em um...)
- Globalmente distribuído (pessoas trabalhando remotamente de países diferentes que estão em contine...)

A qual setor pertence a sua empresa? *

- Consultoria
- Bancário
- Saúde
- Telecom
- Público/Governamental
- Financeiro
- Varejo
- Instituição acadêmica
- Fábrica de software
- Outros... *

Sua empresa possui fins lucrativos?

- Sim
- Não

Qual é o papel principal que você desempenha no seu projeto ou time? *

- Analista / Engenheiro de Requisito
- Arquiteto
- Desenvolvedor
- Product Manager/Owner
- Gerente do Projeto
- Gerente de Qualidade
- Scrum Master/Agile Coach
- Líder de projeto
- Testador
- UX/UI
- Outros...

*

Qual seu gênero?

- Masculino
- Feminino
- Prefiro não responder
- Outros...

*

Qual a sua idade?

Texto de resposta curta

.....

Qual é o seu nível de escolaridade (concluído)? *

Ensino Médio

Ensino Técnico

Graduação

Mestrado

Doutorado *

Quantos anos de atuação no mercado?

Menos de 1 ano

Entre 1 e 5 anos

Entre 6 e 10 anos

Entre 10 e 15 anos

Mais de 15 anos

Você só precisa enviar suas respostas agora!



Não se esqueça de clicar no botão Enviar no final desta página para nos enviar suas respostas!

Obrigado por contribuir com esta pesquisa. Suas respostas foram extremamente importantes para nós e serão analisadas cuidadosamente.

Desejamos a você tudo de melhor em sua vida e trabalho.

B Apêndice - Guia de Métricas.

Pesquise no documento o nome da métrica e obtenha a respectiva fórmula.

Métrica	Fórmula
Tempo gasto desde o início do projeto até a entrega do primeiro incremento	Data de entrega do primeiro incremento - Data de início do projeto
Tempo gasto desde o início do projeto até o encerramento do projeto	Data de encerramento do projeto - Data de início do projeto
True Sprint Length	Data real de entrega do incremento quando maior do que a data de término do sprint definida anteriormente - Data de início da sprint
Horizonte de previsão (Forecast Horizon)	Se conhecermos a velocidade estimada da equipe $\mu(V)$ e o desvio padrão $\sigma(V)$, podemos expressar o Forecast Horizon em sprints como $hl\mu(V) \pm \sigma(V) * hl\mu(V)$ para 68% de certeza ou $hl\mu(V) \pm 2\sigma(V) * hl\mu(V)$ para 95% certeza. Se conhecermos o comprimento da sprint l , podemos expressar o Forecast Horizon em unidades de tempo, $hl\mu(V) \pm l\sigma(V) * hl\mu(V)$
Lead Time	Data de conclusão de um processo - Data de início de um processo
Tempo de resposta da solicitação de atendimento ao cliente	Data de solicitação de atendimento resolvida - data de solicitação do serviço
Tempo de ciclo por recurso	Data de conclusão do recurso - Data em que o recurso foi adicionado ao backlog
Distribuição de fluxo	Número de cada tipo de item de trabalho em um determinado momento
Indicador de prontidão para liberação	#Defeitos / Taxa de remoção de defeitos - (Taxa de execução de teste - Taxa de aprovação no teste)
Tempo por função	Tempo de duração da tarefa/Quantidade da função da tarefa
Tempo restante	Tempo de duração da tarefa / (Data do teste da tarefa - Data de início da tarefa)

Tempo gasto (todas as tarefas)	$((\text{Data do teste da tarefa} - \text{Data de início da tarefa}) / \text{Tempo de duração da tarefa}) * 100$
Tempo gasto (tarefas principais usando a conclusão de subtarefas)	Soma(Tempo gasto em subtarefas) / Número de subtarefas
Tempo disponível	Tempo gasto em subtarefas - (Data de teste da tarefa - Data de início da tarefa)
Tempo corrido	período médio de tempo que leva para concluir um determinado tipo de item de trabalho
Tempo de espera pelas alterações	Quantidade de tempo de espera para fazer uma alteração
Preparação para liberação	Quantidade de tags de correção obrigatórias
Índice de previsibilidade	Combine as seguintes métricas de variação: (i) Variação de escopo: Quantidade de story points entregues/Quantidade de story points comprometidos; (ii) Variação da velocidade de liberação: Velocidade atual / Velocidade média; (iii) Quantidade de defeitos entregues; e (iv) variação do valor do negócio. Calculando a média das métricas de variância, um índice geral pode ser criado e então a previsibilidade de cada lançamento pode ser plotado como uma tendência
Valor presente líquido por esforço	Valor Presente Líquido / Esforço
Valor comercial por esforço	Número de lançamentos importantes em um ano / Horas por pessoa
Net Promoter Score	Calculado ao perguntar aos clientes se eles recomendariam o produto a colegas em uma escala de 0 a 10. As respostas são categorizadas como detratores (0-6), <u>neutros</u> (7-8) e promotores (9-10). Por fim, o percentual total de detratores é subtraído do percentual de promotores para determinar o NPS (Net

	Promoter Score).
Eficiência de fluxo	Tempo ativo total / tempo de fluxo
Previsibilidade de fluxo	Razão entre o valor planejado do negócio alcançado e valor comercial real entregue em um Program Increment (PI)
Velocidade	Quantidade de pontos de história concluídos para itens de trabalho de um tipo durante um período de tempo
Tarefas bloqueadas	Número de tarefas bloqueadas durante um período específico
Pontos de função por homem-ano	Quantidade de trabalho realizado por um indivíduo ao longo do ano
Gráfico burndown	Mostra a quantidade de trabalho restante ao longo do tempo. Normalmente, o eixo horizontal representa o tempo (geralmente em iterações ou dias), enquanto o eixo vertical representa a quantidade de trabalho restante (por exemplo, em pontos de história ou tarefas pendentes)
Gráfico de burnup	Mostra o progresso do trabalho ao longo do tempo, mas diferencia-se do burndown na medida em que também mostra o aumento do escopo. Assim como o burndown, o eixo horizontal representa o tempo, mas o eixo vertical no gráfico de burnup representa tanto o trabalho concluído quanto o trabalho total planejado (escopo).
Desvio de velocidade (Velocity Deviation)	$\sigma(V) / \mu(V)$. Onde $\mu(V)$ é a velocidade esperada (média velocidade baseada em um certo número de sprints anteriores) e $\sigma(V)$ é o desvio padrão da velocidade
Contagem de dependências (Dependency Count)	Número de equipes dependentes imediatas onde há bugs e atrasos
Eficiência	AH / C . Onde AH representa as horas reais gastas na produção resultados tangíveis e C é o número ideal de horas uma equipe pode entregar
Impedimentos	Número de horas que não produzem resultados tangíveis
Pulso de commit (Commit Pulse)	Número de dias entre commits
Velocidade planejada	Quantidade de trabalho que uma equipe espera concluir durante uma sprint

Custo para cada função	Custo da tarefa / Número de funções da tarefa
Custo restante	Custo da tarefa - Custo da despesa
Custo das despesas (todas as tarefas)	(Custo das despesas / Custo da tarefa) * 100
Funções restantes	Funções de tarefa - Funções de tarefa concluídas
Funções concluídas (todas as tarefas)	(Funções concluídas da tarefa / Total de funções da tarefa) * 100
Custo disponível	Custo da tarefa - Custo da despesa
Frequência de implantação	Número de implantações em um determinado período
Aceleração	$(\text{VelocidadeSprint}_X - \text{VelocidadeSprint}_Y) / \text{VelocidadeSprint}_Y$
Pulso de equipe ágil	Avaliações regulares com as equipes para entender suas opiniões sobre a adoção de práticas de desenvolvimento iterativas. Por meio de pesquisas informais
Delta de precisão de estimativa	$(PH - AH) / PH$. Onde PH são as horas estimadas para cada sprint para entregar o proposto e AH são as horas reais gasto produzindo resultados tangíveis
Eficácia	AH / PH . Onde PH são as horas estimadas para cada sprint para entregar o proposto e AH são as horas reais gastas produzindo resultados tangíveis
Custo por hora virtual planejado	CM / PH . Onde CM é o modelo de custo aplicado pelo fornecedor para a empresa relacionada a economias de escala e PH são as horas estimadas para cada sprint entregar o proposto
Custo real virtual por hora	CM / AH . Onde CM é o modelo de custo aplicado pelo fornecedor para a empresa relacionada a economias de escala e AH são as horas reais gastas na produção de resultados tangíveis
Desperdício	$(\text{Delta de precisão da estimativa} * \text{Custo horário virtual real}) / 100$
Custo delta	$(\text{Custo horário virtual real} - \text{Custo horário virtual planejado}) / 100$
Indicador de integridade do sistema de medição	$(\#Atividades \text{ com medidas} / \#Atividades \text{ no total}) * 100\%$

Eficácia percebida	Compilada por meio de pesquisas realizadas com a equipe e stakeholders, onde as respostas são avaliadas em uma escala para medir aspectos como trabalho em equipe, requisitos coleta, planejamento, qualidade da metodologia, cultura, compartilhamento de conhecimento, percepção geral do processo, moral da equipe, engajamento e satisfação dos participantes
Envolvimento dos funcionários	Alguns usam uma pesquisa anual, enquanto outros contam com uma Employee Net Promoter Score (eNPS), perguntando aos funcionários qual a probabilidade de eles devem recomendar seu empregador em uma escala de 10 pontos
Trabalho em progresso (Work In Progress)	Quantidade máxima de trabalho que pode existir em cada status de um fluxo de trabalho
Carga de fluxo	Um Diagrama de Fluxo Cumulativo (CFD) é uma ferramenta amplamente utilizada para representar visualmente o fluxo de trabalho ao longo do tempo. Ilustra o trabalho em um determinado estado, a velocidade com que novos itens entram na fila de trabalho (taxa de chegada) e a velocidade com que são concluídos (taxa de saída)
Taxa de defeitos por gravidade	Número de defeitos (gravidade 1 ou 2...) na produção / 100
Números de problemas externos relatados	Número de problemas externos relatados de uma determinado versão
Dias abertos com problemas externos relatados	Data de problemas externos relatados resolvidos - Data de problemas externos relatados
Linhas de código por história de usuário	Linhas de código / história de usuário
Número de arquivos por história de usuário	Número de arquivos / história de usuário
Linhas de código para refatoração	Quantidade de linhas de código para refatoração
Número de desenvolvedores por recurso	Número de desenvolvedores / recurso
Cobertura de testes unitários por história de usuário	Cobertura de testes unitários / história de usuário
Taxa de aprovação em testes unitários por história de usuário	Taxa de aprovação em testes unitários / história de usuário
Número de defeitos por história de usuário	Número de defeitos / história de usuário
Taxa de falhas de mudanças	Porcentagem de alterações que exigem

	correção após a conclusão em produção
Tempo de restauração do serviço	Quantidade de tempo de espera para restauração do serviço
Taxa de aprovação de testes	Porcentagem de taxa de aprovação para conjunto de testes automatizados
Taxa de defeito de entrada para taxa de defeito de saída	Defeitos de entrada / defeitos de saída
Taxa de defeito	Número de defeitos / 100
Tempo de inatividade do serviço	Quantidade de tempo que um serviço específico ficou inativo
Dívida técnica	Duas maneiras: (i) Abordagens baseadas em ferramentas como SonarQube e outros; ou (ii) a equipe deve discutir e selecionar um conjunto dos resultados de dívida técnica mais dolorosos e começar a rastreá-los como parte de suas reuniões retrospectivas