



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE RUSSAS
CURSO DE ENGENHARIA DE SOFTWARE
CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE

EMANUEL DE ABREU BARBOSA

**UMA SÍNTESE EMPÍRICA DO MÉTODO USARP: PERCEPÇÃO, APLICAÇÃO E
DESAFIOS EM CONTEXTOS ACADÊMICOS E INDUSTRIAIS**

RUSSAS

2026

EMANUEL DE ABREU BARBOSA

UMA SÍNTESE EMPÍRICA DO MÉTODO USARP: PERCEPÇÃO, APLICAÇÃO E
DESAFIOS EM CONTEXTOS ACADÊMICOS E INDUSTRIAIS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia de Software
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia de Software.

Orientador: Prof. Ms. Hugo Nathan Bar-
bosa Régis.

RUSSAS

2026

EMANUEL DE ABREU BARBOSA

UMA SÍNTESE EMPÍRICA DO MÉTODO USARP: PERCEPÇÃO, APLICAÇÃO E
DESAFIOS EM CONTEXTOS ACADÊMICOS E INDUSTRIAIS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia de Software
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Engenharia de Software.

Aprovado em: 07/07/2026

BANCA EXAMINADORA

Prof. Ms. Hugo Nathan Barbosa Régis (Orientador)
Universidade Federal do Ceará (UFC)

Profa. Dra. Anna Beatriz dos Santos Marques
Universidade Federal do Ceará - UFC

Profa. Ms. Valéria Maria da Silva Pinheiro
Universidade Federal do Ceará - UFC

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me conceder força, saúde e perseverança ao longo desta jornada acadêmica.

À minha família, em especial aos meus pais e meu irmão Rafael, pelo amor, apoio incondicional e por acreditarem no meu potencial mesmo nos momentos mais difíceis.

Aos meus amigos e colegas de curso, por todo companheirismo, pelas conversas, pelo apoio mútuo nas dificuldades e pela leveza que tornaram a caminhada mais prazerosa.

Ao Prof. Ms. Hugo Nathan Barbosa Régis, pela orientação dedicada, pelos conselhos técnicos e acadêmicos, e pelo incentivo constante ao longo da realização deste trabalho.

Às professoras Profa. Dra. Anna Beatriz dos Santos Marques e Profa. Ms. Valéria Maria da Silva Pinheiro, pelo aceite em compor a banca avaliadora, pela disponibilidade e pelas contribuições que certamente enriquecerão este trabalho.

À Universidade Federal do Ceará, pela formação acadêmica e pelo ambiente de aprendizado proporcionado ao longo da graduação.

RESUMO

A elicitação de requisitos de usabilidade permanece um desafio recorrente na Engenharia de Software, especialmente pela dificuldade de tornar explícitos aspectos de usabilidade ainda nas fases iniciais do desenvolvimento. O método USARP (Usability Requirements with Personas and User Stories) foi proposto com o objetivo de apoiar a identificação, discussão e especificação sistemática de requisitos de usabilidade por meio do uso integrado de personas, user stories e mecanismos de usabilidade representados por cartas. Apesar de estudos empíricos indicarem benefícios associados ao método, ainda carece de uma síntese que integre as evidências disponíveis e avalie sua aplicação em diferentes contextos. Este trabalho apresenta uma síntese empírica do método USARP, fundamentada na integração de evidências quantitativas e qualitativas provenientes de quatro estudos primários conduzidos em contextos acadêmicos e industriais. A metodologia combina análise estatística descritiva, redução de dimensionalidade por meio da Análise de Componentes Principais (PCA) e análise temática das respostas abertas, consolidadas no nível do respondente, permitindo investigar dimensões latentes relacionadas ao valor percebido do método e à utilidade percebida de seus mecanismos. Os resultados indicam percepção amplamente positiva e homogênea do método entre os contextos acadêmico e industrial, com forte efeito de teto nos itens avaliados. A Análise de Componentes Principais, de caráter exploratório, evidenciou uma dimensão de percepção positiva geral e contrastes secundários entre os blocos de cartas do método. As comparações entre contextos não revelaram diferenças estatisticamente robustas após a correção para múltiplas comparações, com exceção de um item relativo ao contexto do mecanismo de usabilidade. A análise qualitativa das respostas abertas identificou as cartas e os guias de especificação como o principal elemento de apoio à elicitação, além de tensões relacionadas à curva de aprendizado e ao volume de cartas e da subutilização das cartas de prototipação. Como contribuição, este trabalho fornece uma síntese empírica crítica do USARP, destacando suas potencialidades, limitações e implicações para sua adoção em contextos reais de desenvolvimento de software.

Palavras-chave: USARP; requisitos de usabilidade; síntese empírica; engenharia de software; personas e user stories.

ABSTRACT

The elicitation of usability requirements remains a recurring challenge in Software Engineering, particularly due to the difficulty of making usability aspects explicit during early development phases. The USARP (Usability Requirements with Personas and User Stories) method was proposed to support the systematic identification, discussion, and specification of usability requirements through the integrated use of personas, user stories, and usability mechanisms represented by cards. Although previous empirical studies have reported benefits related to the method, an integrated synthesis of the available evidence across different contexts is still lacking. This work presents an empirical synthesis of the USARP method, grounded in the integration of quantitative and qualitative evidence from four primary studies conducted in academic and industrial contexts. The methodology combines descriptive statistical analysis, dimensionality reduction through Principal Component Analysis (PCA), and thematic analysis of open-ended responses, consolidated at the respondent level, enabling the investigation of latent dimensions associated with the perceived value of the method and the perceived usefulness of its mechanisms. The results indicate a broadly positive and homogeneous perception of the method across academic and industrial contexts, with a strong ceiling effect in the assessed items. The Principal Component Analysis, of an exploratory nature, revealed a dimension of general positive perception and secondary contrasts among the method's card blocks. Comparisons between contexts revealed no statistically robust differences after correction for multiple comparisons, except for one item related to the usability mechanism context. The qualitative analysis of the open-ended responses identified the cards and specification guides as the main element supporting elicitation, along with tensions related to the learning curve and the volume of cards, and the underuse of the prototyping cards. As a contribution, this study provides a critical empirical synthesis of the USARP method, outlining its strengths, limitations, and implications for its adoption in real-world software development settings.

Keywords: USARP; usability requirements; empirical synthesis; software engineering; personas and user stories.

LISTA DE FIGURAS

Figura 1 – Exemplos de cartas de mecanismos de usabilidade do método USARP . . .	17
Figura 2 – Exemplos de cartas de requisitos de usabilidade do método USARP	18
Figura 3 – Exemplos de cartas de prototipação do método USARP	18
Figura 4 – Composição da amostra consolidada por coorte e contexto	43
Figura 5 – Distribuição das respostas por item	44
Figura 6 – Matriz de correlação de Spearman entre as variáveis-núcleo	47
Figura 7 – Escolha do número de componentes: scree plot e variância acumulada . . .	49
Figura 8 – Erro de reconstrução em função do número de componentes	49
Figura 9 – Mapa de calor das cargas fatoriais das cinco primeiras componentes	50
Figura 10 – Biplot das componentes PC1 e PC2, por contexto	52
Figura 11 – Biplot das componentes PC1 e PC3, por contexto	52
Figura 12 – Biplot das componentes PC2 e PC3, por contexto	53
Figura 13 – Projeção dos respondentes nas componentes PC1 e PC2, por contexto e por coorte	53
Figura 14 – Projeção dos respondentes nas componentes PC1 e PC3, por contexto e por coorte	54
Figura 15 – Projeção dos respondentes nas componentes PC2 e PC3, por contexto e por coorte	54
Figura 16 – Projeção tridimensional das três primeiras componentes, por contexto	55
Figura 17 – Matriz de dispersão das três primeiras componentes principais	55
Figura 18 – Detecção de observações atípicas pela distância de Mahalanobis	57
Figura 19 – Seleção do número de grupos: coeficiente de silhueta e índice de Davies- Bouldin	59
Figura 20 – Segmentação k-means ($k = 2$) nos planos das componentes de Kaiser	59
Figura 21 – Composição de contexto por grupo (cluster)	60
Figura 22 – Médias por item entre contextos e diferença ordenada (Indústria menos Acadêmico)	62
Figura 23 – Respostas por pergunta e contexto e extensão dos relatos por contexto . . .	64
Figura 24 – Frequência dos temas por pergunta	66
Figura 25 – Categorias temáticas das características e das adaptações sugeridas	66
Figura 26 – Categorias temáticas dos recursos do site: aspectos positivos e limitações . .	68

Figura 27 – Polaridade das respostas por pergunta e por contexto	68
Figura 28 – Termos de conteúdo mais frequentes nas respostas abertas	69
Figura 29 – Categoria dominante por bloco qualitativo	70
Figura 30 – Síntese integrada: médias do levantamento e frequência dos temas qualitativos	71

LISTA DE TABELAS

Tabela 1 – Comparação entre os trabalhos relacionados e o trabalho proposto	26
Tabela 2 – Trabalhos sobre o USARP e decisão de inclusão na síntese	29
Tabela 3 – Variáveis analíticas derivadas dos questionários	32
Tabela 4 – Distribuição dos respondentes por contexto e coorte	43
Tabela 5 – Estatísticas descritivas das variáveis-núcleo	44
Tabela 6 – Médias das variáveis-núcleo por contexto	45
Tabela 7 – Médias das variáveis-núcleo por coorte	45
Tabela 8 – Teste de normalidade de Shapiro-Wilk e assimetria por item	46
Tabela 9 – Autovalores e variância explicada por componente principal	48
Tabela 10 – Cargas fatoriais (loadings) das cinco primeiras componentes	49
Tabela 11 – Perfil dos grupos identificados pela segmentação k-means	57
Tabela 12 – Composição dos grupos por contexto e coorte	58
Tabela 13 – Kruskal-Wallis entre coortes, por componente principal	60
Tabela 14 – Mann-Whitney entre contextos (Acadêmico e Indústria), por componente principal	60
Tabela 15 – Kruskal-Wallis entre coortes, item a item	61
Tabela 16 – Caracterização do corpus qualitativo por bloco	64
Tabela 17 – Menções por tema, segundo o bloco de pergunta e o contexto	69
Tabela 18 – Síntese qualitativa por bloco	70

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
2.1	Objetivo Geral	14
2.2	Objetivos Específicos	14
3	FUNDAMENTAÇÃO TEÓRICA	15
3.1	Usabilidade na Engenharia de Software	15
3.2	Engenharia de Requisitos e Requisitos de Usabilidade	16
3.3	O Método USARP	16
3.4	O Modelo de Aceitação de Tecnologia (TAM)	19
3.5	Síntese de Evidências Empíricas em Engenharia de Software	20
3.6	Análise Temática e Síntese Qualitativa	20
3.7	Análise de Componentes Principais	21
4	TRABALHOS RELACIONADOS	22
4.1	Métodos para Elicitação de Requisitos de Usabilidade	22
4.2	Abordagens Baseadas em Personas e Prototipação	23
4.3	O Método USARP e Evidências Empíricas	23
4.4	Sínteses Empíricas em Engenharia de Software	24
4.5	Síntese Crítica e Posicionamento do Trabalho	25
5	METODOLOGIA	27
5.1	Caracterização da Pesquisa	27
5.2	Desenho da Síntese Empírica	27
5.2.1	<i>Estudos Primários Incluídos na Síntese</i>	28
5.2.2	<i>Critérios de Seleção e Delimitação do Escopo</i>	29
5.2.3	<i>Naturezas de Evidência e Estratégia de Integração</i>	30
5.3	Instrumento de Coleta de Dados	30
5.4	Escala de Medida e Variáveis	32
5.5	Unidade de Análise e Estrutura dos Dados	33
5.6	Consolidação e Pré-processamento dos Dados	33
5.7	Análise Quantitativa	34
5.7.1	<i>Estatística Descritiva e Teste de Normalidade</i>	34

5.7.2	<i>Estrutura de Correlação entre as Variáveis</i>	35
5.7.3	<i>Adequação Amostral à Análise de Componentes Principais</i>	35
5.7.4	<i>Análise de Componentes Principais</i>	36
5.7.5	<i>Deteção de Observações Atípicas</i>	37
5.7.6	<i>Segmentação Não Supervisionada dos Perfis de Percepção</i>	37
5.7.7	<i>Comparações entre Grupos</i>	37
5.8	Análise Qualitativa	38
5.8.1	<i>Síntese Qualitativa dos Estudos Primários</i>	38
5.8.2	<i>Reanálise Temática das Respostas Abertas Consolidadas</i>	39
5.8.3	<i>Convergência entre Achados Quantitativos e Qualitativos</i>	39
5.9	Ambiente Computacional e Reprodutibilidade	40
5.10	Ameaças à Validade	40
6	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	42
6.1	Composição da Amostra Consolidada	42
6.2	Caracterização Descritiva e Efeito de Teto	44
6.3	Estrutura de Correlação entre as Variáveis	46
6.4	Adequação Amostral à Análise de Componentes Principais	47
6.5	Análise de Componentes Principais	48
6.6	Deteção de Observações Atípicas	56
6.7	Segmentação Não Supervisionada dos Perfis de Percepção	57
6.8	Comparações entre Grupos	60
6.9	Síntese dos Resultados Quantitativos	63
6.10	Análise Qualitativa Temática	63
6.10.1	<i>Caracterização do Corpus Qualitativo</i>	64
6.10.2	<i>Características do Método que Auxiliam a Elicitação</i>	64
6.10.3	<i>Adaptações Sugeridas</i>	66
6.10.4	<i>Recursos do Site de Apoio</i>	67
6.10.5	<i>Polaridade e Tom das Respostas</i>	68
6.10.6	<i>Síntese Qualitativa e Relação com os Estudos Primários</i>	69
6.11	Discussão Integrada dos Resultados	71
7	CONSIDERAÇÕES FINAIS	73
	REFERÊNCIAS	76

1 INTRODUÇÃO

A Engenharia de Software tem enfrentado desafios crescentes relacionados à qualidade de uso dos sistemas desenvolvidos, especialmente em cenários caracterizados por desenvolvimento ágil, equipes distribuídas e ciclos curtos de entrega (Sommerville, 2011). Nesse contexto, a usabilidade assume papel central, uma vez que influencia diretamente a eficácia, a eficiência e a satisfação dos usuários durante a interação com sistemas de software. De acordo com a norma ISO/IEC 25010, a usabilidade é definida como o grau em que um produto pode ser utilizado por usuários específicos para alcançar objetivos determinados com efetividade, eficiência e satisfação em um contexto de uso específico (ISO, 2011).

Apesar de sua reconhecida importância, aspectos de usabilidade ainda são frequentemente tratados de forma tardia no processo de desenvolvimento, sendo abordados apenas durante fases de avaliação ou testes finais. Tal prática tende a gerar retrabalho, aumento de custos e limitações estruturais para correção de problemas já incorporados à arquitetura do sistema (Juristo *et al.*, 2007). Estudos apontam que antecipar a consideração de requisitos de usabilidade nas fases iniciais do desenvolvimento contribui significativamente para a melhoria da qualidade do produto final (Carvajal *et al.*, 2013).

No âmbito da Engenharia de Requisitos, a elicitação de requisitos de usabilidade permanece um desafio recorrente. Diferentemente de requisitos funcionais, esses requisitos envolvem aspectos subjetivos, dependentes do contexto e fortemente associados às características e preferências dos usuários, o que dificulta sua identificação e especificação explícita (Ardito *et al.*, 2014). Em ambientes ágeis, essa dificuldade é potencializada pelo foco em funcionalidades e pela informalidade dos artefatos de requisitos tradicionalmente utilizados.

Com o objetivo de apoiar a elicitação e a especificação sistemática de requisitos de usabilidade em projetos ágeis, Oliveira *et al.* (2020) propuseram o método USARP (*Usability Requirements with Personas and User Stories*). O método integra o uso de *personas*, *user stories* e um conjunto de cartas baseadas em mecanismos de usabilidade, com o intuito de orientar discussões estruturadas e enriquecer requisitos funcionais com aspectos de usabilidade, ainda nas fases iniciais do desenvolvimento (Oliveira *et al.*, 2020).

Embora estudos empíricos tenham indicado benefícios associados à aplicação do USARP, como aumento da efetividade na identificação de aspectos de usabilidade e apoio à tomada de decisão em equipes multidisciplinares (Marques *et al.*, 2023a), ainda se observa uma lacuna na literatura quanto à síntese sistemática das evidências empíricas disponíveis sobre o

método. Em especial, não está claro como os diferentes mecanismos propostos pelo USARP são efetivamente utilizados na prática, quais apresentam maior aderência em contextos reais e como a percepção positiva do método se relaciona com seu uso operacional.

Além disso, os estudos existentes foram conduzidos em contextos variados, incluindo ambientes acadêmicos e industriais, bem como cenários presenciais e remotos, o que sugere possível variabilidade na aplicação do método. No entanto, essa variabilidade ainda não foi analisada de forma integrada, dificultando a compreensão de padrões recorrentes de uso, limitações práticas e implicações para a adoção do USARP em diferentes cenários de desenvolvimento de software.

Diante desse contexto, o objetivo deste trabalho é investigar empiricamente a aplicação do método USARP na eliciação de requisitos de usabilidade, por meio de uma síntese empírica de evidências provenientes de contextos acadêmicos e industriais. Para alcançar esse objetivo, o estudo adota uma abordagem quantitativa e qualitativa, combinando análise estatística descritiva, técnicas de redução de dimensionalidade por meio da Análise de Componentes Principais (PCA) e análise qualitativa interpretativa. As análises são conduzidas em ambiente computacional, com apoio de ferramentas de ciência de dados, visando identificar padrões de avaliação, percepções recorrentes e informações latentes associadas à aplicação do método USARP.

Este trabalho está organizado da seguinte forma: o Capítulo 2 apresenta os objetivos do estudo, detalhando o objetivo geral e os objetivos específicos. O Capítulo 3 apresenta a fundamentação teórica, discutindo conceitos de usabilidade, Engenharia de Requisitos, o Modelo de Aceitação de Tecnologia, a síntese de evidências empíricas e técnicas de redução de dimensionalidade, como a Análise de Componentes Principais (PCA), com foco no método USARP. O Capítulo 4 descreve os trabalhos relacionados, contextualizando este estudo em relação à literatura existente. O Capítulo 5 descreve a metodologia adotada, detalhando o desenho da síntese empírica, os instrumentos de coleta de dados e as técnicas de análise utilizadas. O Capítulo 6 apresenta e discute os resultados obtidos a partir da síntese empírica realizada. Por fim, o Capítulo 7 traz as considerações finais, limitações do estudo e sugestões para trabalhos futuros.

2 OBJETIVOS

2.1 Objetivo Geral

Investigar empiricamente a aplicação do método USARP na elicitação de requisitos de usabilidade, por meio de uma síntese empírica de evidências provenientes de contextos acadêmicos e industriais.

2.2 Objetivos Específicos

- Sintetizar evidências empíricas sobre a percepção de utilidade, produtividade e eficácia associadas ao uso do USARP;
- Caracterizar o uso percebido dos mecanismos de usabilidade propostos pelo método, destacando padrões de recorrência e de subutilização relatados pelos participantes;
- Evidenciar diferenças e semelhanças entre contextos acadêmicos e industriais quanto à aplicação do método USARP;
- Extrair informações latentes associadas à aplicação do método por meio de técnicas de redução de dimensionalidade, como a Análise de Componentes Principais (PCA);
- Discutir implicações práticas e limitações do método com base nos resultados empíricos sintetizados.

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos, métodos e abordagens científicas que embasam a análise do método USARP e a estratégia empírica adotada. Inicialmente, discute-se o papel da usabilidade na Engenharia de Software e sua relação com a qualidade de sistemas interativos. Em seguida, aborda-se a Engenharia de Requisitos com ênfase na elicitação de requisitos de usabilidade. Posteriormente, o método USARP é detalhado de forma aprofundada, incluindo seus fundamentos conceituais, estrutura operacional e evolução. Em seguida, apresenta-se o *Technology Acceptance Model* (TAM), base do instrumento de avaliação empregado nos estudos analisados. Por fim, são discutidos os fundamentos da síntese de evidências empíricas em Engenharia de Software, da análise temática e da Análise de Componentes Principais (PCA), técnicas utilizadas para a síntese e a interpretação dos dados deste estudo.

3.1 Usabilidade na Engenharia de Software

A usabilidade é reconhecida como um atributo central da qualidade de software, especialmente em sistemas interativos. De acordo com a norma ISO/IEC 25010, a usabilidade refere-se ao grau em que um sistema pode ser utilizado por usuários específicos para atingir objetivos definidos com eficácia, eficiência e satisfação, considerando um contexto de uso específico (ISO, 2011). Essa definição evidencia que a usabilidade não se restringe à interface gráfica, mas envolve aspectos relacionados ao comportamento do sistema, às funcionalidades disponibilizadas e à forma como o usuário interage com essas funcionalidades ao longo do tempo.

Na Engenharia de Software, a usabilidade passou a ser compreendida como um fator estratégico para o sucesso de produtos digitais. Estudos clássicos demonstram que sistemas com baixa usabilidade tendem a apresentar maior taxa de erros operacionais, baixa aceitação por parte dos usuários e aumento nos custos de suporte e manutenção (Nielsen, 1993). Além disso, decisões de projeto tomadas nas fases iniciais do desenvolvimento exercem influência direta sobre a experiência do usuário e sobre a sustentabilidade do sistema ao longo de seu ciclo de vida (Preece *et al.*, 2015).

Apesar desse reconhecimento, a usabilidade ainda é frequentemente tratada de forma tardia no processo de desenvolvimento, sendo avaliada apenas em fases finais, geralmente associadas a testes de interface ou validação com usuários. Essa abordagem corretiva limita

a capacidade de resolver problemas estruturais de uso, uma vez que ajustes realizados após a implementação tendem a demandar maior esforço, custo e retrabalho (Juristo *et al.*, 2007).

3.2 Engenharia de Requisitos e Requisitos de Usabilidade

A Engenharia de Requisitos compreende um conjunto de atividades voltadas à identificação, análise, documentação e gerenciamento dos requisitos de um sistema, com o objetivo de garantir que o software atenda às necessidades reais de seus usuários e demais stakeholders (Sommerville, 2011). Tradicionalmente, os requisitos são classificados em funcionais e não funcionais. Os requisitos de usabilidade são, em geral, enquadrados como requisitos não funcionais; contudo, essa classificação tem se mostrado insuficiente para representar adequadamente seu impacto no sistema.

Requisitos de usabilidade apresentam características particulares que dificultam sua elicitação e especificação. Eles são frequentemente subjetivos, dependentes do contexto de uso e fortemente influenciados por fatores humanos, como expectativas, experiência prévia e limitações cognitivas dos usuários (Ardito *et al.*, 2014). Além disso, muitos aspectos de usabilidade manifestam-se por meio de comportamentos funcionais do sistema, afetando diretamente o design e a arquitetura do software.

Nesse contexto, Juristo, Moreno e Sánchez-Segura propuseram o conceito de *Functional Usability Features* (FUFs), defendendo que determinados aspectos de usabilidade devem ser tratados como funcionalidades explícitas do sistema (Juristo *et al.*, 2007). As FUFs são compostas em mecanismos de usabilidade, como feedback, prevenção de erros, alertas, desfazer ações e ajuda multinível, que podem ser especificados como requisitos verificáveis e rastreáveis. Essa abordagem reforça a necessidade de considerar a usabilidade desde a fase de elicitação de requisitos, integrando-a às decisões técnicas e funcionais do sistema.

3.3 O Método USARP

O método USARP (*Usability Requirements with Personas and User Stories*) foi proposto com o objetivo de apoiar a elicitação e a especificação sistemática de requisitos de usabilidade em fases iniciais do desenvolvimento de software, especialmente em contextos ágeis (Oliveira *et al.*, 2020). O método surge da constatação de que, embora existam diretrizes e recomendações de usabilidade bem estabelecidas, há uma lacuna prática quanto à sua incorporação

efetiva durante a elicitación de requisitos, sobretudo por equipes não especialistas em usabilidade.

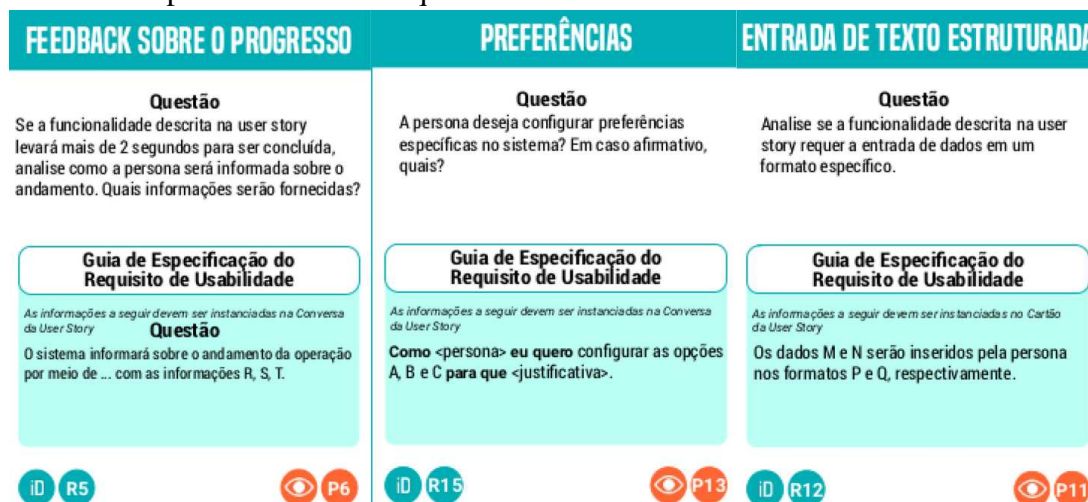
O USARP fundamenta-se na integração de três elementos principais: *personas*, *user stories* e mecanismos de usabilidade derivados das *Functional Usability Features*. As *personas* são utilizadas para representar perfis de usuários reais, considerando suas características, necessidades, objetivos e limitações. As *user stories* oferecem um formato simples e amplamente adotado no desenvolvimento ágil para a especificação de requisitos sob a perspectiva do usuário. Os mecanismos de usabilidade atuam como guias estruturados que direcionam a reflexão sobre aspectos funcionais relacionados à qualidade de uso. Na Figura 1, são apresentadas as cartas de mecanismos de usabilidade do USARP, que orientam a reflexão sobre funcionalidades relacionadas à qualidade de uso. Já as cartas de requisitos de usabilidade, ilustradas na Figura 2, auxiliam na especificação desses requisitos. Por fim, a Figura 3 traz as cartas de prototipação, que apoiam a materialização das soluções propostas.

Figura 1 – Exemplos de cartas de mecanismos de usabilidade do método USARP



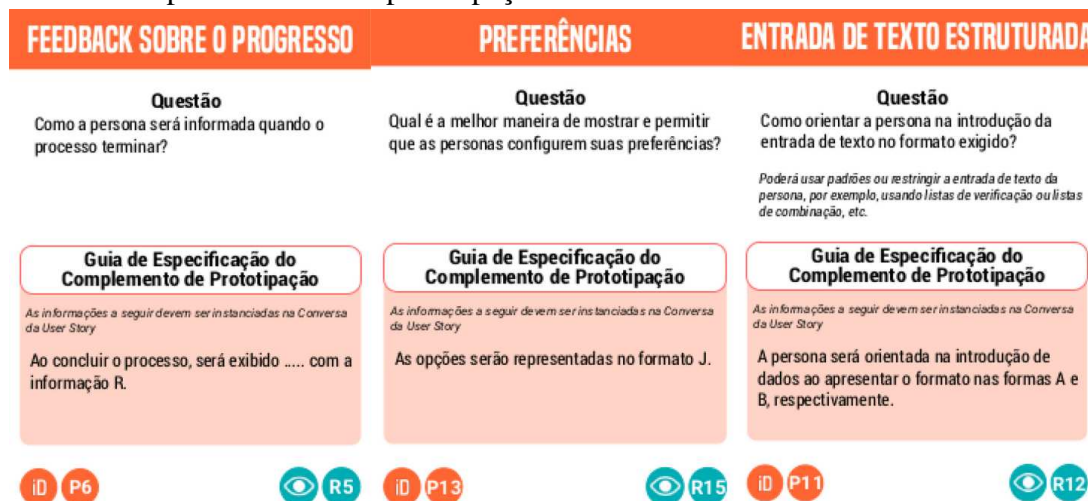
Fonte: Elaborado pelo autor (2025), com base em Oliveira, Ferreira e Marques (2020).

Figura 2 – Exemplos de cartas de requisitos de usabilidade do método USARP



Fonte: Elaborado pelo autor (2025), com base em Oliveira, Ferreira e Marques (2020).

Figura 3 – Exemplos de cartas de prototipação do método USARP



Fonte: Elaborado pelo autor (2025), com base em Oliveira, Ferreira e Marques (2020).

O método propõe um processo estruturado que se inicia com a criação de *personas*, segue pela extração de requisitos potenciais e culmina na escrita e no enriquecimento de *user stories* com mecanismos de usabilidade explícitos. Essa abordagem permite que aspectos de usabilidade sejam discutidos, negociados e documentados de forma sistemática, reduzindo a dependência de especialistas e promovendo maior conscientização da equipe de desenvolvimento.

Estudos empíricos indicam que o USARP é eficaz na obtenção de requisitos de usabilidade, embora também evidenciem desafios relacionados à carga cognitiva imposta por alguns mecanismos e à necessidade de maior direcionamento no uso dos artefatos (Marques *et al.*, 2023a). Esses resultados motivaram a evolução do método, incluindo a transformação dos mecanismos de usabilidade em cartas, com o objetivo de tornar o processo mais dinâmico, acessível e colaborativo.

Mais recentemente, Marques *et al.* (2023b) consolidaram uma evolução do método a partir de um estudo exploratório conduzido em uma turma de Engenharia de Software, no qual foram observadas dificuldades na seleção das cartas durante as sessões de *brainstorming*. Com base nessa análise, os autores revisaram as cartas, removendo conteúdos redundantes e combinando cartas de conteúdo semelhante, com a eliminação de oito cartas. Além disso, propuseram dois novos artefatos: um *checklist* que organiza os mecanismos de usabilidade em quatro categorias (personalização do sistema, feedback do sistema, controle e apoio ao usuário e entrada de dados) e um quadro para dispor as cartas selecionadas conforme as características dos usuários e do sistema. O processo de adoção do método foi atualizado para incorporar duas novas etapas, de seleção e de organização das cartas, anteriores ao *brainstorming*. É essa versão evoluída, baseada em cartas e em uma etapa de filtragem por *checklist*, que fundamenta os estudos de aplicação analisados neste trabalho.

3.4 O Modelo de Aceitação de Tecnologia (TAM)

O instrumento quantitativo utilizado nos estudos primários analisados fundamenta-se no *Technology Acceptance Model* (TAM), modelo proposto por Davis (1989) para explicar a aceitação e o uso de tecnologias da informação. O TAM postula que a intenção de uso de uma tecnologia é determinada, sobretudo, por dois construtos: a utilidade percebida (*perceived usefulness*), definida como o grau em que uma pessoa acredita que o uso de determinado sistema melhoraria seu desempenho, e a facilidade de uso percebida (*perceived ease of use*), entendida como o grau em que a pessoa acredita que esse uso seria livre de esforço (Davis, 1989). Esses construtos têm sido amplamente empregados para avaliar a aceitação de métodos, técnicas e ferramentas no campo da Engenharia de Software.

Em sua evolução, o modelo foi estendido pelo TAM3, que detalha os determinantes da utilidade e da facilidade de uso percebidas e organiza um conjunto de intervenções capazes de influenciar a aceitação de uma tecnologia ao longo de sua adoção (Venkatesh; Bala, 2008). No contexto deste trabalho, o TAM oferece o arcabouço conceitual para o bloco de percepção geral do instrumento, cujos itens relativos a utilidade, produtividade, eficácia e facilidade de aplicação do método derivam diretamente dos construtos de utilidade e de facilidade de uso percebidas.

3.5 Síntese de Evidências Empíricas em Engenharia de Software

A consolidação de evidências provenientes de múltiplos estudos é uma estratégia central da Engenharia de Software Empírica, sobretudo em cenários caracterizados por heterogeneidade de contextos e de métodos (Kitchenham *et al.*, 2009). Revisões sistemáticas e estudos de síntese permitem identificar, avaliar criticamente e integrar resultados de estudos primários por meio de procedimentos explícitos e reprodutíveis, reduzindo vieses e aumentando a confiabilidade das conclusões (Kitchenham *et al.*, 2009; Kitchenham; Charters, 2007).

A literatura distingue, contudo, diferentes graus e formas de síntese. A meta-análise, em sentido estrito, consiste na agregação estatística de tamanhos de efeito padronizados provenientes de estudos comparáveis, usualmente com grupos de comparação e com a estimação da heterogeneidade entre estudos (Rodrigues; Ziegelmann, 2010). Em grande parte da pesquisa em Engenharia de Software, entretanto, essa agregação não é viável, seja pela ausência de grupos de controle, seja pela diversidade de instrumentos e de desenhos de estudo. Por essa razão, Cruzes e Dybå (2010) observam que muitas revisões da área não realizam síntese estatística propriamente dita e defendem a adoção de métodos de síntese mais amplos, como as sínteses narrativas e temáticas, capazes de integrar evidências qualitativas e quantitativas de naturezas distintas.

É nesse espectro que se situa o presente trabalho. Não se trata de uma meta-análise, pois não há agregação ponderada de tamanhos de efeito nem grupo de comparação controlado; trata-se de uma síntese empírica que consolida, no nível do respondente, os dados de levantamento de diferentes aplicações de um mesmo método e os integra a uma síntese qualitativa das evidências reportadas. Essa abordagem é adequada para investigar padrões de percepção, uso e avaliação de um método de Engenharia de Software aplicado em diferentes contextos acadêmicos e industriais, em linha com a discussão de Runeson *et al.* (2012) sobre estudos de caso conduzidos em múltiplos ambientes, nos quais a comparação entre cenários distintos fortalece a validade externa dos resultados e aprofunda a compreensão das condições de uso de um método.

3.6 Análise Temática e Síntese Qualitativa

A interpretação das evidências qualitativas deste trabalho apoia-se na análise temática, abordagem voltada à identificação, à organização e à interpretação de padrões de significado,

denominados temas, em dados textuais (Braun; Clarke, 2006). A análise temática é particularmente adequada para sintetizar respostas abertas e relatos de experiência, por admitir tanto a codificação indutiva, derivada dos próprios dados, quanto o diálogo com categorias previamente reportadas na literatura (Saldaña, 2013). No campo da Engenharia de Software, a síntese qualitativa e temática é reconhecida como um recurso essencial para integrar evidências provenientes de estudos conduzidos em contextos distintos (Cruzes; Dybå, 2010). Neste estudo, a análise temática é aplicada às respostas abertas consolidadas e articulada às codificações originais dos estudos primários, conforme detalhado na metodologia.

3.7 Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é uma técnica estatística multivariada de redução de dimensionalidade que transforma um conjunto de variáveis correlacionadas em um número menor de variáveis não correlacionadas, denominadas componentes principais, que retêm a maior parte da variância original dos dados (Jolliffe, 2002). Cada componente é uma combinação linear das variáveis observadas, e a análise das cargas associadas a cada componente permite interpretar as estruturas latentes que organizam as respostas (Hair *et al.*, 2019). Na presente síntese, a PCA é empregada de forma exploratória e descritiva, com o objetivo de resumir a variabilidade das avaliações e identificar agrupamentos de itens, e não como análise fatorial confirmatória, ressalva detalhada na metodologia em razão das características dos dados, em especial o tamanho reduzido da amostra e o acentuado efeito de teto.

4 TRABALHOS RELACIONADOS

Este capítulo apresenta e discute trabalhos relacionados à elicitación e especificação de requisitos de usabilidade no contexto da Engenharia de Software, com ênfase em métodos, técnicas e abordagens aplicadas em ambientes ágeis. O objetivo desta seção é posicionar o presente trabalho em relação à literatura existente, identificando convergências, limitações e lacunas que justificam a realização de uma síntese empírica do método USARP.

Os estudos analisados foram identificados a partir de buscas em bases reconhecidas, como Google Scholar, ACM Digital Library, IEEE Xplore e repositório de trabalhos acadêmicos da Universidade Federal do Ceará, priorizando trabalhos que abordam usabilidade, requisitos não funcionais, *personas*, *user stories* e evidências empíricas de aplicação em contextos acadêmicos ou industriais.

4.1 Métodos para Elicitación de Requisitos de Usabilidade

Juristo *et al.* (2007) propuseram um conjunto de diretrizes para elicitación de funcionalidades de usabilidade, introduzindo o conceito de *Functional Usability Features* (FUFs), que defende a incorporação de aspectos de usabilidade como funcionalidades explícitas do sistema. Embora o trabalho represente um avanço conceitual importante ao tratar a usabilidade como parte do comportamento funcional do software, sua abordagem é fortemente orientada a documentação tradicional de requisitos, o que dificulta sua adoção em ambientes ágeis baseados em artefatos leves.

Moreno e Yagüe (2012) exploraram a integração das FUFs com *user stories*, investigando como mecanismos de usabilidade podem influenciar a criação, modificação ou refinamento de histórias de usuário e critérios de aceitação. Apesar dos resultados positivos, o estudo concentra-se predominantemente na fase de especificação, não oferecendo um processo estruturado para a elicitación sistemática dos requisitos de usabilidade a partir das características dos usuários.

Choma *et al.* (2016) propuseram a abordagem *UserX Story*, que incorpora aspectos de Experiência do Usuário (UX) à escrita de *user stories* por meio do uso de *personas* e das heurísticas de Nielsen. Nessa abordagem, as heurísticas são utilizadas principalmente como base para critérios de aceitação. Embora a proposta contribua para aproximar UX e métodos ágeis, ela trata a usabilidade como requisito não funcional e foca majoritariamente em aspectos de

interface, não explorando seu impacto funcional ou arquitetural.

4.2 Abordagens Baseadas em Personas e Prototipação

Ferreira *et al.* (2018) propuseram a técnica PATHY, voltada à criação de *personas* com foco na identificação de requisitos potenciais de software. A técnica se mostrou eficaz para compreender o perfil dos usuários e suas necessidades, porém não é direcionada especificamente à elicitacão de requisitos de usabilidade, exigindo sua combinaçao com outras abordagens para esse fim.

Outros estudos investigaram processos que combinam elicitacão de requisitos e prototipacão. Teixeira *et al.* (2014) apresentaram uma metodologia aplicada ao domínio da saúde, na qual protótipos são utilizados como mediadores para validacão iterativa de requisitos. Apesar de evidenciar benefícios práticos, o foco do estudo não está na usabilidade como requisito sistematicamente elicitado, mas na aceitacão da soluçao pelos usuários finais.

Gonçalves e Rocha (2019) propuseram um processo metodológico baseado em normas internacionais para o desenvolvimento de interfaces inteligentes, integrando práticas de IHC ao ciclo de vida do software. O trabalho destaca a importância da usabilidade, porém carece de evidências empíricas de aplicacão contínua em contextos reais de desenvolvimento.

4.3 O Método USARP e Evidências Empíricas

O método USARP foi proposto por Oliveira *et al.* (2020) com o objetivo de apoiar a elicitacão e especificacão de requisitos de usabilidade em fases iniciais do desenvolvimento, integrando *personas*, *user stories* e mecanismos de usabilidade derivados das FUFs. Diferentemente de abordagens anteriores, o USARP oferece um processo estruturado que conecta explicitamente características dos usuários aos requisitos de usabilidade, promovendo discussões colaborativas entre membros da equipe.

Estudos subsequentes avaliaram o método em diferentes contextos, incluindo ambientes distribuídos, indicando benefícios relacionados à conscientizacão sobre usabilidade e à melhoria da qualidade das especificacões produzidas (Marques *et al.*, 2023a). A partir de um estudo exploratório conduzido em contexto acadêmico, no qual identificaram dificuldades na seleçao das cartas, Marques *et al.* (2023b) consolidaram uma evoluçao do método, com a revisao das cartas e a proposta de um *checklist* e de um quadro de seleçao que orientam a

escolha das cartas mais adequadas a cada conjunto de histórias de usuário. Ainda assim, esses estudos evidenciam variações na forma como os mecanismos do método são utilizados, bem como diferenças entre a percepção positiva e o uso relatado dos artefatos propostos.

4.4 Sínteses Empíricas em Engenharia de Software

Uma vez que este trabalho se caracteriza como uma síntese empírica, cabe examinar os trabalhos que adotam essa mesma natureza metodológica, e não apenas aqueles que compartilham o objeto de estudo. Essa comparação é o que permite situar a contribuição do presente trabalho na última coluna da Tabela 1.

No campo da Engenharia de Software, a síntese de evidências por meio de revisões sistemáticas e estudos de mapeamento tem ganhado espaço como forma de organizar e interpretar o conhecimento produzido na área. Kitchenham *et al.* (2009) destacam que, embora haja um crescimento na produção de revisões sistemáticas em Engenharia de Software, ainda existem desafios relacionados ao rigor metodológico e, especialmente, à forma como os resultados são sintetizados. Cruzes e Dybå (2010) complementam essa discussão ao mostrarem que uma parcela significativa das revisões analisadas não apresenta síntese propriamente dita, e que, entre as que apresentam, predominam abordagens narrativas ou temáticas. Esses autores defendem a ampliação do repertório de métodos de síntese para tornar as revisões mais úteis para a pesquisa e para a prática.

Especificamente no domínio da usabilidade de software, Almazroi (2021) realizou um mapeamento sistemático com o objetivo de classificar os estudos existentes e identificar tendências e lacunas na literatura. Os resultados indicam a predominância de estudos de proposta e de avaliação, enquanto estudos de validação empírica ainda aparecem de forma mais limitada. Em conjunto, esses trabalhos mostram que a literatura da área ainda é marcada, em grande parte, por sínteses descritivas ou narrativas, voltadas a panoramas amplos de técnicas, temas e tendências.

Cabe explicitar, contudo, que as sínteses existentes situam-se em dois extremos. De um lado, estudos de natureza metodológica ou terciária, como os de Kitchenham *et al.* (2009) e Cruzes e Dybå (2010), que tomam como objeto o próprio conjunto de revisões da área. De outro, mapeamentos sistemáticos amplos, como o de Almazroi (2021), que classificam a literatura de usabilidade por temas e tendências. Em nenhum dos casos há agregação de evidências no nível do respondente sobre a aplicação de um método determinado.

Essa ausência tem explicação metodológica e não decorre de omissão da literatura. Uma meta-análise, em sentido estrito, requer a combinação de tamanhos de efeito provenientes de estudos independentes e comparáveis, usualmente dotados de grupo de comparação (Kitchenham *et al.*, 2009). Os estudos sobre métodos de elicitação, incluindo os do USARP, avaliam a técnica de forma isolada, sem comparador e em número reduzido, o que impossibilita a estimação e a agregação de efeitos. Some-se a isso a heterogeneidade de instrumentos e desfechos entre estudos, que inviabiliza a comparabilidade exigida pelo procedimento. É precisamente por essa razão que este trabalho se define como uma síntese empírica, e não como meta-análise: em vez de agregar medidas resumidas por estudo, ele reúne os dados individuais dos respondentes em uma base única, estratégia recomendada quando tais dados estão disponíveis, por preservar a variabilidade intragrupo e permitir a estratificação por contexto e coorte (Cruzes; Dybå, 2010; Hair *et al.*, 2019).

4.5 Síntese Crítica e Posicionamento do Trabalho

A análise dos trabalhos relacionados evidencia que, embora existam diversas propostas para incorporar usabilidade ao processo de Engenharia de Requisitos, ainda são limitadas as abordagens que investigam simultaneamente a elicitação e a especificação de requisitos de usabilidade em contextos ágeis, com base em evidências empíricas sistematizadas. Além disso, observa-se que a maioria dos estudos apresenta avaliações pontuais ou contextuais, sem uma síntese integrada das evidências disponíveis.

Diferentemente dessas abordagens, o presente trabalho adota uma síntese híbrida, integrando análise estatística descritiva, redução de dimensionalidade por meio da Análise de Componentes Principais (PCA) e interpretação qualitativa dos resultados obtidos com o método USARP. Assim, enquanto os trabalhos relacionados oferecem um escopo mais amplo ou sínteses não quantitativas, este TCC focaliza evidências empíricas específicas do USARP de forma estatisticamente informada, aprofundando a análise de um método específico em contextos acadêmicos e industriais e destacando padrões de percepção e de avaliação ainda pouco explorados na literatura.

A Tabela 1 apresenta uma comparação entre os trabalhos relacionados e o trabalho proposto, destacando dimensões relevantes para o posicionamento deste estudo.

Tabela 1 – Comparação entre os trabalhos relacionados e o trabalho proposto

Trabalho	Eng. de Requisitos	Elicitação e Especificação	Requisitos de Usabilidade	Personas / User Stories	Evidência Empírica	Síntese Sistematizada
Juristo et al. (2007)	X	X	X		X	
Moreno e Yagüe (2012)	X		X	X	X	
Choma et al. (2016)	X		X	X	X	
Ferreira et al. (2018)	X	X		X	X	
Oliveira et al. (2020)	X	X	X	X	X	
Marques et al. (2023)	X		X	X	X	
Kitchenham et al. (2009)					X	X
Cruzes e Dybå (2010)					X	X
Almazroi (2021)			X		X	X
Este trabalho	X	X	X	X	X	X

Fonte: Elaborada pelo autor (2025).

5 METODOLOGIA

Este capítulo apresenta os procedimentos metodológicos adotados no trabalho. Foi utilizada uma abordagem empírica, exploratória e de natureza mista, combinando métodos quantitativos e qualitativos. A metodologia foi estruturada de modo a assegurar rigor científico, transparência nos procedimentos e a possibilidade de replicação do estudo (Sommerville, 2011; Runeson *et al.*, 2012). O capítulo está organizado em torno de três blocos: a caracterização da pesquisa e o desenho da síntese empírica (Seções 5.1 e 5.2); a descrição do instrumento, das variáveis e do tratamento dos dados (Seções 5.3 a 5.6); e os procedimentos analíticos quantitativos e qualitativos, seguidos do ambiente computacional e das ameaças à validade (Seções 5.7 a 5.10).

5.1 Caracterização da Pesquisa

Do ponto de vista de sua finalidade, esta pesquisa é classificada como aplicada, pois investiga a utilização de um método específico, o USARP, em contextos reais e educacionais, visando produzir conhecimento com potencial impacto prático. Quanto aos objetivos, o estudo possui caráter exploratório e descritivo, uma vez que busca compreender padrões de uso, percepção e efetividade do método, bem como descrever empiricamente os resultados observados (Sommerville, 2011; Runeson *et al.*, 2012).

No que se refere à abordagem, o trabalho adota uma estratégia mista, combinando análise quantitativa, baseada em estatística descritiva, redução de dimensionalidade e testes não-paramétricos, com análise qualitativa, fundamentada na interpretação de evidências reportadas e produzidas a partir dos estudos primários (Runeson *et al.*, 2012). As duas vertentes são tratadas como leituras complementares de um mesmo fenômeno: a análise quantitativa resume padrões de percepção, enquanto a qualitativa contextualiza e aprofunda esses padrões, conforme detalhado na Seção 5.8.

5.2 Desenho da Síntese Empírica

Este estudo foi conduzido como uma síntese empírica de evidências quantitativas e qualitativas provenientes de diferentes estudos sobre o método USARP, integradas de forma articulada. Diferentemente de uma meta-análise estatística comparativa clássica, o trabalho não estabelece comparação entre grupos de controle e experimental, nem realiza agregação ponderada de tamanhos de efeito; o objetivo é identificar padrões recorrentes, tendências de percepção

e estruturas latentes associadas à aplicação do método (Kitchenham *et al.*, 2009; Rodrigues; Ziegelmann, 2010). Essa escolha é adequada quando não há grupos de comparação direta e o interesse recai sobre a síntese de evidências de um mesmo método aplicado em contextos distintos (Rodrigues; Ziegelmann, 2010; Cruzes; Dybå, 2010), situação típica da avaliação de um método de Engenharia de Software em diferentes cenários (Runeson *et al.*, 2012).

5.2.1 Estudos Primários Incluídos na Síntese

Os dados analisados foram extraídos de estudos empíricos realizados em contextos acadêmicos e industriais, envolvendo estudantes de Engenharia de Software e equipes de desenvolvimento atuando de forma remota. Foram considerados quatro estudos primários sobre a aplicação do USARP, apresentados a seguir em ordem cronológica de publicação.

- **Oliveira *et al.* (2020)** introduzem o método USARP e relatam um estudo de viabilidade com cinco estudantes de Engenharia de Software, em ambiente acadêmico e de forma remota. A condução seguiu a metodologia de *Design Science Research* (Wieringa, 2009), e a avaliação combinou a inspeção das *user stories* produzidas com a análise de um grupo focal, codificado por procedimentos de *Grounded Theory*. É importante registrar que esse estudo antecede a versão do método baseada em cartas: as cartas e a posterior etapa de filtragem por *checklist* resultaram de uma evolução do método consolidada por Marques *et al.* (2023b), motivada pela percepção de que os padrões originais eram extensos e cansativos. Por essa razão, o estudo integra a presente síntese sobretudo no plano qualitativo e conceitual, e não compõe a matriz quantitativa.
- **Fiori (2022)** descrevem um estudo de caso no contexto acadêmico, realizado de forma remota, no qual o método foi aplicado por uma turma de Engenharia de Software organizada em nove equipes, no semestre 2021-2. O estudo reutilizou o instrumento de levantamento originado no estudo de indústria descrito adiante e combinou análise quantitativa com análise qualitativa por *Grounded Theory*. Essa mesma aplicação foi posteriormente publicada, em versão revisada e estendida, por Marques *et al.* (2023b), que, a partir das dificuldades observadas na seleção das cartas, propuseram a evolução do método (revisão das cartas, *checklist* e quadro de seleção), detalhada na fundamentação teórica. Por se referirem à mesma aplicação, os dados de levantamento dessa coorte são considerados uma única vez na matriz quantitativa, evitando dupla contagem.
- **Simões (2022)** descrevem um estudo de caso na indústria, com equipes trabalhando

remotamente em um cenário de Desenvolvimento Distribuído de Software. O estudo reaproveitou o instrumento adotado por Fiori (2022), acrescido de um bloco específico sobre a aplicação do método na indústria, e empregou análise quantitativa e qualitativa por *Grounded Theory*.

- **Marques *et al.* (2023a)** relatam um estudo de caso participativo conduzido por uma equipe ágil remota na indústria, aplicando o método em um projeto real. O estudo adotou análise documental, observação e um levantamento (*survey*) com os participantes. Embora publicado nos anais do simpósio somente em 2023, o estudo foi conduzido em 2021, e seu levantamento é a origem do instrumento quantitativo reaproveitado pelos demais estudos e pelas aplicações posteriores.

Dessa forma, três dos quatro estudos (Fiori, Simões e Marques *et al.*) aplicaram um instrumento de levantamento comum, cujas respostas, no nível do respondente, compõem a base quantitativa desta síntese. O estudo pioneiro de Oliveira *et al.* (2020) contribui com evidência qualitativa e com a fundamentação do próprio instrumento e do método em sua forma atual.

5.2.2 Critérios de Seleção e Delimitação do Escopo

Cabe esclarecer que a literatura sobre o USARP não se restringe aos quatro estudos acima. O conjunto de publicações identificadas sobre o método é mais amplo, e os demais trabalhos foram deliberadamente excluídos da síntese por não atenderem ao critério de inclusão adotado. A Tabela 2 sistematiza essa decisão.

Tabela 2 – Trabalhos sobre o USARP e decisão de inclusão na síntese

Trabalho	Natureza	Decisão
Oliveira <i>et al.</i> (2020)	Proposta do método e estudo de viabilidade (grupo focal)	Incluído (evidência qualitativa)
Fiori (2022)	Estudo de caso acadêmico com levantamento	Incluído
Simões (2022)	Estudo de caso na indústria com levantamento	Incluído
Marques <i>et al.</i> (2023a)	Estudo de caso na indústria com levantamento	Incluído
Marques <i>et al.</i> (2023b)	Publicação revisada da mesma aplicação de Fiori e proposta de evolução do método	Não contabilizado como estudo independente, para evitar dupla contagem dos mesmos respondentes
Saraiva (2023)	Mineração de padrões de requisitos a partir dos artefatos produzidos	Excluído da matriz: a unidade de análise é a história de usuário, e não a percepção do respondente
Silva <i>et al.</i> (2024)	Proposta da ferramenta USARP Tool, ainda em desenvolvimento	Excluído: não apresenta avaliação empírica com participantes

Fonte: Elaborada pelo autor (2026).

Os trabalhos excluídos não são, contudo, irrelevantes para a interpretação dos achados: Marques *et al.* (2023b) fundamentam a versão do método efetivamente avaliada, Saraiva (2023) oferecem evidência complementar sobre os artefatos gerados pela adoção do método e Silva *et al.* (2024) apontam a direção de automação discutida nas considerações finais. Todos são retomados no Capítulo 4.

5.2.3 *Naturezas de Evidência e Estratégia de Integração*

Foram consideradas duas naturezas de evidência. A primeira corresponde aos dados quantitativos, obtidos por meio de escalas ordinais (Likert e similar), voltados à percepção de desempenho, produtividade, eficácia e utilidade do método, bem como à utilidade percebida do conteúdo de suas cartas. A segunda corresponde aos registros qualitativos, como respostas abertas, comentários e observações dos participantes, que permitem interpretar o uso percebido do método e a percepção sobre seus artefatos. A integração das duas naturezas segue uma estratégia sequencial e complementar: as evidências quantitativas, consolidadas no nível do respondente, são submetidas a análises exploratórias e inferenciais (Seção 5.7); em paralelo, as evidências qualitativas são sintetizadas tematicamente (Seção 5.8); por fim, os dois conjuntos são confrontados para evidenciar convergências e tensões entre percepção declarada e uso relatado.

5.3 **Instrumento de Coleta de Dados**

Os dados quantitativos analisados são provenientes de estudos empíricos primários que aplicaram o método USARP em contextos acadêmicos e industriais. Assim, este trabalho realiza uma análise secundária baseada nas respostas já coletadas por meio de questionários estruturados nesses estudos, o que permite uma síntese comparativa entre diferentes cenários de aplicação do método.

O instrumento comum, originalmente proposto por Marques *et al.* (2023a) e reutilizado por Fiori (2022) e Simões (2022), organiza-se em dois blocos principais. O primeiro fundamenta-se no *Technology Acceptance Model* (TAM) (Davis, 1989), modelo amplamente consolidado para avaliar a aceitação de tecnologias a partir dos construtos de utilidade percebida e facilidade de uso percebida, em sua formulação estendida por Venkatesh e Bala (2008), com foco em captar percepções relacionadas à aceitação e ao uso percebido do método. O segundo avalia a utilidade percebida do conteúdo das cartas do método, considerando suas funcionalidades

específicas no apoio à eliciação de requisitos de usabilidade e à prototipação.

1. **Bloco 1 – Percepção Geral do Método USARP (baseado no TAM):** contempla quatro dimensões associadas à aceitação do método:

- **Utilidade:** grau em que o USARP foi considerado útil para melhorar o desempenho na eliciação de requisitos de usabilidade, alinhada ao construto *perceived usefulness* do TAM;
- **Eficiência:** facilidade percebida na aplicação do método, relacionada ao construto *perceived ease of use*;
- **Produtividade:** percepção de ganho de tempo e de eficácia prática durante a utilização do método;
- **Compleitude:** percepção de que o método contribui para uma identificação mais abrangente dos requisitos de usabilidade.

Cada dimensão foi mensurada por itens em escala Likert de 5 pontos, variando de “Discordo totalmente” (1) a “Concordo totalmente” (5).

2. **Bloco 2 – Percepção Específica sobre o Conteúdo das Cartas do USARP:** subdividido conforme os três tipos de cartas do método:

- **Cartas Amarelas (Mecanismos de Usabilidade);**
- **Cartas Azuis (Requisitos);**
- **Cartas Laranjas (Prototipação).**

Para cada tipo de carta, os participantes avaliaram aspectos como clareza da descrição, contextualização, apoio à eliciação de requisitos e suporte à prototipação. As respostas foram registradas em uma escala ordinal de 6 pontos, variando de “Não foi utilizado” (1) a “Ajudou muito” (6).

Além dos itens fechados, o instrumento contém questões abertas, voltadas às características do método consideradas úteis (ou não) para a eliciação e para a prototipação, e a sugestões de adaptação. Essas questões não foram convertidas em variáveis numéricas, servindo de base para a análise qualitativa (Seção 5.8). Dessa forma, o instrumento reúne tanto percepções gerais sobre a aceitação do método quanto avaliações específicas sobre os artefatos que o compõem, permitindo compreender não apenas se o método foi considerado útil, mas também quais de seus elementos foram percebidos como mais eficazes.

5.4 Escala de Medida e Variáveis

As questões fechadas utilizaram escalas ordinais do tipo Likert, com cinco ou seis níveis de resposta conforme o bloco. Para permitir a análise estatística, as respostas textuais foram convertidas em valores numéricos inteiros, procedimento usual em análises exploratórias com dados perceptivos (Hair *et al.*, 2019). As respostas do bloco de percepção geral foram codificadas na escala de 1 a 5, e as respostas sobre o conteúdo das cartas, na escala de 1 a 6.

A partir dessas respostas, definiram-se as variáveis quantitativas diretamente relacionadas aos construtos avaliados nos estudos primários. Essas variáveis representam, de um lado, a percepção geral de aceitação do método e, de outro, a utilidade percebida de seus artefatos específicos. A Tabela 3 apresenta as variáveis analíticas derivadas dos questionários e suas descrições.

Tabela 3 – Variáveis analíticas derivadas dos questionários

Variável	Descrição
Produtividade	Mede a percepção de ganho de tempo e de desempenho na elicitação de requisitos de UI/UX ao utilizar o USARP.
Eficiência	Indica o grau de facilidade percebida para aplicar o método durante o processo de elicitação.
Completeness	Representa a percepção de que o USARP permite identificar um conjunto mais abrangente de requisitos de usabilidade.
Utilidade	Expressa o quanto o método é percebido como útil para apoiar a elicitação de requisitos de UI/UX.
Descrição do mecanismo de usabilidade	Avalia se a descrição apresentada nas cartas amarelas foi considerada clara e compreensível.
Contexto do mecanismo de usabilidade	Verifica se o contexto associado às cartas amarelas auxiliou na compreensão e no uso do mecanismo.
Questão a ser explorada sobre requisitos	Mede o quanto as cartas azuis contribuíram para orientar a identificação de requisitos de usabilidade.
Guia de especificação do requisito	Avalia se as cartas azuis auxiliaram na formulação e na especificação dos requisitos identificados.
Questão sobre prototipação	Verifica se as cartas laranjas apoiaram a reflexão sobre prototipação durante a elicitação.
Guia de especificação da prototipação	Indica se as cartas laranjas contribuíram para orientar a especificação de soluções de prototipação.

Fonte: Elaborada pelo autor (2026).

Além das dez variáveis-núcleo descritas na Tabela 3, o conjunto de dados inclui um grupo auxiliar de variáveis sobre o site de apoio ao método. Essas variáveis foram tratadas apenas de forma descritiva e fora das análises multivariadas, conforme justificado na Seção 5.6.

5.5 Unidade de Análise e Estrutura dos Dados

A unidade de análise adotada é o respondente individual. Cada observação da base consolidada corresponde a um único participante que avaliou o método em um dos estudos primários, e não a uma média agregada por grupo. Essa decisão é metodologicamente necessária, pois as técnicas multivariadas empregadas (correlação, análise de componentes principais e segmentação por agrupamento) exigem variabilidade entre observações individuais e um número de observações superior ao de variáveis, condição indispensável para que estimativas de associação e de estrutura latente sejam interpretáveis (Hair *et al.*, 2019). A agregação prévia das respostas em médias por coorte colapsaria a variabilidade intragrupo e reduziria drasticamente o número de observações, inviabilizando a estimação da matriz de correlação e, por consequência, a própria PCA.

A base consolidada reúne as respostas individuais dos participantes dos estudos que adotaram o instrumento comum, totalizando aproximadamente sessenta e oito respondentes distribuídos entre os contextos acadêmico e industrial. A matriz analítica foi organizada em duas camadas. A primeira reúne as dez variáveis-núcleo, de natureza ordinal, listadas na Tabela 3. A segunda compreende as variáveis auxiliares sobre o site de apoio. Além das variáveis-núcleo, preservaram-se dois atributos categóricos de estratificação: o contexto de aplicação, recodificado de forma binária (Acadêmico *versus* Indústria), e a coorte, que identifica o estudo de origem e o ano de aplicação, mantida para viabilizar as comparações entre grupos descritas na Seção 5.7.7.

5.6 Consolidação e Pré-processamento dos Dados

Os dados provenientes dos estudos primários foram consolidados em uma única base no nível do respondente, preservando-se os identificadores originais de cada participante para fins de rastreabilidade. As respostas abertas foram organizadas em um arquivo complementar em formato longo, com uma linha por par respondente e pergunta, dedicado à análise qualitativa.

Quanto a dados ausentes, identificou-se um único valor faltante entre as dez variáveis-núcleo, no item de guia de prototipação, referente a um respondente da coorte industrial de 2022. Dada a magnitude desprezível da ausência e o fato de as técnicas multivariadas empregadas exigirem observações completas, optou-se pela exclusão desse respondente do conjunto analítico (*listwise deletion*), procedimento adequado quando a ausência é pontual e não sistemática (Hair *et al.*, 2019). Esse é o único motivo pelo qual a base bruta, de 68 registros, resulta em 67

respondentes nas análises multivariadas. A decisão e seu impacto foram documentados de forma transparente.

Antes da aplicação das técnicas multivariadas, as variáveis-núcleo foram padronizadas por *escore-z*, com média zero e desvio-padrão unitário. A padronização é indispensável neste caso porque os dois blocos de itens utilizam amplitudes distintas (escala de 1 a 5 e de 1 a 6); sem ela, os itens de maior amplitude dominariam artificialmente a variância total e enviesariam a extração dos componentes (Hair *et al.*, 2019). As variáveis referentes ao site de apoio foram tratadas em separado, apenas de forma descritiva e fora da PCA, porque foram respondidas somente pelos participantes que declararam ter utilizado o recurso, configurando uma ausência estrutural, e não aleatória; incluí-las na matriz multivariada introduziria viés de seleção e padrões espúrios de correlação.

5.7 Análise Quantitativa

A análise quantitativa foi conduzida sobre a matriz consolidada no nível do respondente. Ela compreende a caracterização descritiva e o teste de normalidade, o exame da estrutura de correlação, a verificação da adequação amostral, a análise de componentes principais, a detecção de observações atípicas, a segmentação não supervisionada e as comparações entre grupos, descritas nas subseções a seguir.

5.7.1 Estatística Descritiva e Teste de Normalidade

A primeira etapa consistiu em estatística descritiva das dez variáveis-núcleo, com o objetivo de caracterizar o comportamento das respostas em cada item. Foram calculadas medidas de tendência central (média e mediana) e de dispersão (desvio-padrão e amplitude interquartílica), além de distribuições de frequência representadas por histogramas, de modo a evidenciar a concentração das respostas ao longo da escala.

Em itens de percepção avaliados por escalas ordinais positivas, é comum observar um efeito de teto, isto é, a concentração de respostas nos níveis superiores da escala, o que produz assimetria negativa na distribuição. Para verificar formalmente em que medida esse efeito comprometeria a suposição de normalidade, aplicou-se o teste de Shapiro-Wilk a cada item (Shapiro; Wilk, 1965), considerado adequado e sensível para amostras de tamanho reduzido (Field, 2018). Esse diagnóstico é o critério objetivo que orienta a escolha entre procedimentos

paramétricos e não-paramétricos nas etapas inferenciais: a presença de assimetria acentuada e de desvios de normalidade recomenda o emprego de testes baseados em postos, descritos na Seção 5.7.7.

5.7.2 Estrutura de Correlação entre as Variáveis

Como passo preparatório à redução de dimensionalidade, examinou-se a estrutura de associação entre as variáveis-núcleo por meio da matriz de correlação de Spearman. A escolha do coeficiente de postos de Spearman, em vez do coeficiente de Pearson, justifica-se pela natureza ordinal dos itens e pela ausência de normalidade verificada na Seção 5.7.1: a correlação por postos não pressupõe linearidade nem normalidade e é mais apropriada para escalas do tipo Likert (Field, 2018). A inspeção dessa matriz permite antecipar quais conjuntos de itens tendem a variar em conjunto e fornece subsídio interpretativo para os componentes extraídos pela PCA.

A matriz de correlação foi estimada sobre a amostra completa, sem estratificação por contexto ou por coorte. Essa decisão é deliberada e apoia-se em três razões. Primeiro, o propósito desta etapa é diagnóstico: examinar a estrutura de associação entre os itens que alimentará a PCA, a qual também é ajustada sobre a matriz completa; estimar correlações em subamostras produziria um diagnóstico que não corresponderia ao objeto efetivamente analisado. Segundo, a estimação de uma matriz de dez variáveis exige um número de observações substancialmente superior ao de variáveis (Hair *et al.*, 2019); a subamostra industrial dispõe de onze respondentes, de modo que as correlações estimadas nesse subgrupo seriam instáveis e com intervalos de confiança demasiado amplos para sustentar qualquer comparação entre matrizes. Terceiro, a investigação de diferenças entre contextos e coortes é conduzida na Seção 5.7.7 por procedimentos apropriados ao tamanho da amostra, isto é, testes não-paramétricos com correção para múltiplas comparações, e não pelo contraste informal de matrizes de correlação estimadas em subgrupos pequenos.

5.7.3 Adequação Amostral à Análise de Componentes Principais

Antes de extrair os componentes, avaliou-se a adequação dos dados à PCA por dois diagnósticos complementares. O teste de esfericidade de Bartlett verifica se a matriz de correlação difere significativamente de uma matriz identidade; sua rejeição indica que existem correlações suficientes entre as variáveis para justificar a extração de estrutura latente (Bartlett, 1950). A medida de adequação amostral de Kaiser-Meyer-Olkin (KMO) quantifica a proporção de variância potencialmente comum entre as variáveis, sendo usualmente interpretada por faixas

em que valores abaixo de 0,5 são considerados inaceitáveis e a faixa de 0,5 a 0,6, medíocre (Kaiser, 1974; Field, 2018).

Reconhece-se que, em amostras pequenas e com forte efeito de teto, situação esperada neste estudo, o KMO tende a se situar em faixa modesta, uma vez que diversos itens saturam no valor máximo e reduzem a variância compartilhável. Por essa razão, e em consonância com Hair *et al.* (2019), a PCA é empregada aqui como técnica exploratória e descritiva, voltada a resumir a variabilidade e a identificar agrupamentos de itens, e não como análise fatorial confirmatória. Essa ressalva é explicitada na interpretação dos resultados, de modo a não superestimar o alcance das estruturas identificadas.

5.7.4 *Análise de Componentes Principais*

A PCA foi ajustada sobre a matriz padronizada das dez variáveis-núcleo. O número de componentes a reter não foi fixado a priori; reportaram-se múltiplos critérios complementares e a decisão resultou de sua leitura conjunta, conforme recomendado pela literatura de análise multivariada (Hair *et al.*, 2019). Foram considerados: o critério de Kaiser, que retém os componentes com autovalor superior a um (Kaiser, 1960); o *scree plot*, que examina o ponto de inflexão na curva de autovalores (Cattell, 1966); a variância acumulada, observando-se os limiares de 80%, 90% e 95%; e o erro de reconstrução, isto é, o erro quadrático médio entre a matriz original e a reconstruída a partir de k componentes, como recurso adicional para avaliar a perda de informação associada a cada nível de redução.

A partir desses critérios, definiram-se dois conjuntos de retenção com propósitos distintos. O conjunto mais parcimonioso, indicado pela regra de Kaiser, foi adotado para a interpretação substantiva dos componentes. Um conjunto mais amplo, suficiente para atingir 80% da variância acumulada, foi empregado nas análises multidimensionais que demandam mais dimensões, como a segmentação por agrupamento e a inspeção da matriz de dispersão. Para cada componente retido, analisaram-se as cargas fatoriais (*loadings*), a fim de identificar quais variáveis mais contribuem para cada dimensão, e projetaram-se os escores dos respondentes, estratificados por contexto e por coorte, em *bipLOTS* e em gráficos de dispersão bidimensionais e tridimensionais, permitindo comparar visualmente a proximidade entre os diferentes cenários de aplicação.

5.7.5 *Detecção de Observações Atípicas*

A detecção de observações atípicas foi realizada no espaço dos componentes retidos por meio da distância de Mahalanobis, comparada a um limiar qui-quadrado ao nível de 97,5%, com graus de liberdade iguais ao número de componentes retidos (Hair *et al.*, 2019). Optou-se deliberadamente por não remover os respondentes sinalizados: em uma amostra reduzida, a exclusão de casos extremos diminuiria ainda mais o poder analítico e poderia eliminar justamente a variação de interesse entre contextos. Tais observações foram apenas registradas, servindo como ressalva de robustez e como insumo para a discussão qualitativa.

5.7.6 *Segmentação Não Supervisionada dos Perfis de Percepção*

Para investigar a existência de perfis homogêneos de avaliação do método, aplicou-se o algoritmo de agrupamento *k-means* sobre os escores dos componentes retidos (MacQueen, 1967; Jain, 2010). A opção por agrupar no espaço dos componentes, e não sobre as variáveis originais, apoia-se no fato de que esse espaço é descorrelacionado e concentra a variância sistemática, reduzindo o ruído e a redundância entre itens (Hair *et al.*, 2019).

O número de grupos não foi fixado previamente. Avaliaram-se soluções de dois a oito grupos por meio de dois índices internos de validação: o coeficiente de silhueta, que mede a coesão interna e a separação entre grupos (Rousseeuw, 1987), e o índice de Davies-Bouldin, em que valores menores indicam melhor separação (Davies; Bouldin, 1979). Adotou-se a solução que maximiza a silhueta média, confirmada pela leitura conjunta do índice de Davies-Bouldin. Os grupos resultantes foram caracterizados pelas médias das variáveis originais, para permitir interpretação substantiva, e cruzados com o contexto de aplicação, a fim de verificar se a segmentação reflete a distinção entre ambientes acadêmicos e industriais. Para garantir reprodutibilidade, fixou-se a semente aleatória e adotaram-se múltiplas inicializações do algoritmo.

5.7.7 *Comparações entre Grupos*

As comparações entre cenários de aplicação foram conduzidas por testes não-paramétricos, escolha justificada pela natureza ordinal das variáveis, pelo tamanho reduzido da amostra e pelos desvios de normalidade evidenciados na Seção 5.7.1 (Field, 2018). Empregou-se o teste de Kruskal-Wallis para investigar diferenças entre as coortes (Kruskal; Wallis, 1952) e o

teste U de Mann-Whitney para contrastar os contextos acadêmico e industrial (Mann; Whitney, 1947). As comparações foram aplicadas tanto sobre os escores dos componentes principais quanto, de forma item a item, sobre as variáveis originais.

Para o contraste entre contextos, reportou-se o tamanho de efeito $r = Z/\sqrt{N}$, que permite avaliar a magnitude prática das diferenças além de sua significância estatística (Field, 2018). Considerando a multiplicidade de testes realizados, aplicou-se a correção de Holm sobre os valores- p , procedimento sequencial que controla a taxa de erro do tipo I por família de comparações com maior poder do que a correção de Bonferroni (Holm, 1979). Para os itens que permaneceram significativos após a correção, conduziu-se análise *post-hoc* por comparações par a par de Mann-Whitney, igualmente ajustadas por Holm, a fim de localizar entre quais coortes residem as diferenças.

5.8 Análise Qualitativa

A análise qualitativa opera em dois níveis articulados: a síntese das evidências qualitativas reportadas nos quatro estudos primários e a reanálise temática das respostas abertas consolidadas no nível do respondente. Ao final, os achados qualitativos são confrontados com os quantitativos.

5.8.1 Síntese Qualitativa dos Estudos Primários

A síntese das evidências qualitativas reportadas nos estudos primários seguiu a lógica de síntese de estudos empíricos em Engenharia de Software (Kitchenham *et al.*, 2009; Cruzes; Dybå, 2010). Cada estudo foi examinado quanto às categorias temáticas que seus autores reportaram, observando-se que todos adotaram codificação qualitativa de respostas abertas ou de transcrições. Oliveira *et al.* (2020) codificaram, por procedimentos de *Grounded Theory*, as percepções de um grupo focal, identificando categorias relacionadas à experiência de uso, à utilidade do método, à percepção sobre o uso dos padrões e a sugestões de melhoria, entre as quais a recomendação de simplificar os padrões e convertê-los em cartas. Fiori (2022) e Simões (2022) aplicaram codificação aberta de *Grounded Theory* às respostas abertas de seus questionários, organizando os achados em redes de códigos por categoria. Marques *et al.* (2023a) analisaram qualitativamente as respostas às questões abertas de seu levantamento para interpretar os resultados quantitativos.

Da leitura comparada emergem padrões recorrentes entre contextos. No eixo positivo, destacam-se a percepção de que as cartas e os guias conferem estrutura e ampliam a visão do projeto, o apoio à classificação e à identificação mais ágil de requisitos, a utilidade do uso conjunto de personas e *user stories*, e a contribuição do método para a discussão colaborativa em equipe. No eixo das dificuldades, recorrem a curva de aprendizado inicial, a dificuldade de compreensão das cartas e, sobretudo, a percepção de que o elevado volume de cartas torna o processo cansativo, observação presente desde o estudo pioneiro e reiterada nos contextos acadêmico e industrial. Esses padrões fundamentam o esquema de codificação adotado na reanálise consolidada.

5.8.2 Reanálise Temática das Respostas Abertas Consolidadas

As respostas abertas dos estudos que adotaram o instrumento comum foram reunidas em um corpus consolidado e submetidas a uma análise temática (Braun; Clarke, 2006; Saldaña, 2013). A codificação seguiu orientação indutiva: a partir da leitura repetida das respostas, e em diálogo com as categorias reportadas nos estudos primários (Seção 5.8.1), construiu-se um esquema temático no qual cada tema é definido por um conjunto explícito de termos-chave, normalizados e sem acentuação. A explicitação desse dicionário de termos tem por finalidade tornar o processo auditável e replicável, permitindo que terceiros reproduzam ou ajustem a categorização. Adotou-se codificação de rótulos múltiplos, de modo que uma mesma resposta pode ser associada a mais de um tema quando aborda aspectos distintos. Em paralelo, atribuiu-se a cada resposta uma polaridade (positiva, negativa, mista ou neutra) com base em léxicos de sentimento construídos para o domínio, viabilizando uma leitura agregada do tom dos comentários por pergunta e por contexto.

Os resultados qualitativos foram organizados pela frequência dos temas por pergunta e por contexto, pela identificação dos termos mais recorrentes e pela seleção de trechos representativos de cada categoria. Essa reanálise é compatível com as codificações originais por *Grounded Theory* dos estudos primários, das quais difere por consolidar os relatos de diferentes coortes em um único quadro temático comparável.

5.8.3 Convergência entre Achados Quantitativos e Qualitativos

Por fim, os achados qualitativos foram confrontados com os quantitativos, comparando-se os itens de maior e menor avaliação no levantamento com os temas mais recorrentes nas

respostas abertas. O objetivo é evidenciar convergências, por exemplo entre a utilidade percebida do método e os relatos de ampliação da visão do projeto, e tensões, por exemplo entre a percepção positiva geral e os relatos sobre o volume de cartas e a subutilização das cartas de prototipação. Trata-se de uma interpretação complementar de natureza exploratória, na qual o conteúdo qualitativo ajuda a explicar os padrões observados nos dados numéricos, em coerência com a abordagem mista adotada no estudo (Runeson *et al.*, 2012).

5.9 Ambiente Computacional e Reprodutibilidade

Todas as análises foram conduzidas na linguagem Python, organizadas em um único *notebook* reprodutível que executa, de ponta a ponta, desde o carregamento dos dados até a geração das figuras. Para a manipulação e o pré-processamento dos dados utilizaram-se as bibliotecas *pandas* (McKinney, 2010) e *NumPy* (Harris *et al.*, 2020); para os testes estatísticos (Shapiro-Wilk, Spearman, Kruskal-Wallis e Mann-Whitney), o módulo *stats* da biblioteca *SciPy* (Virtanen *et al.*, 2020); para a PCA, o agrupamento *k-means* e os índices de validação, a biblioteca *scikit-learn* (Pedregosa *et al.*, 2011); e para a correção de múltiplas comparações e a conferência dos diagnósticos de adequação amostral, as bibliotecas *statsmodels* e *factor_analyzer*. As visualizações foram produzidas com *Matplotlib* (Hunter, 2007) e *seaborn*. Para assegurar a reprodutibilidade e a transparência metodológica, princípios centrais da pesquisa empírica em Engenharia de Software (Kitchenham *et al.*, 2009), todas as etapas estocásticas foram conduzidas com semente aleatória fixa, garantindo a obtenção de resultados idênticos em execuções sucessivas. O *notebook*, os arquivos de dados e o dicionário de variáveis foram organizados de modo a permitir a replicação integral do estudo.

5.10 Ameaças à Validade

Em coerência com as boas práticas de estudos empíricos em Engenharia de Software, discutem-se a seguir as principais ameaças à validade (Wohlin *et al.*, 2012; Runeson *et al.*, 2012). Quanto à validade de conclusão, o tamanho reduzido da amostra, o efeito de teto e o valor modesto do KMO limitam o alcance das inferências; por isso a PCA é tratada como exploratória e as comparações empregam testes não-paramétricos com correção de Holm, que reduzem o risco de conclusões espúrias. Quanto à validade de construto, o uso de dados secundários, coletados por instrumentos não idênticos entre todos os estudos, exige harmonização

das variáveis, procedimento documentado para permitir uma análise detalhada; cabe registrar que o estudo pioneiro não aplicou o instrumento quantitativo, contribuindo apenas no plano qualitativo. Quanto à validade interna, por se tratar de uma síntese de evidências, sem grupo de comparação controlado, não se formulam afirmações causais; as análises descrevem padrões de percepção e de avaliação do método. Por fim, quanto à validade externa, a inclusão de coortes em contextos acadêmicos e industriais, presenciais e remotos, amplia a base de evidências, mas a generalização permanece circunscrita a cenários semelhantes aos estudados.

6 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Este capítulo apresenta e discute os resultados quantitativos da síntese empírica do método USARP, obtidos a partir da reanálise da base consolidada no nível do respondente e interpretados em conjunto com as evidências reportadas nos quatro estudos primários. Conforme detalhado na metodologia, a matriz quantitativa reúne as respostas dos estudos que aplicaram o instrumento comum de levantamento, a saber, o estudo de caso na indústria de Marques *et al.* (2023a), o estudo de caso acadêmico de Fiori (2022) e o estudo de caso na indústria de Simões (2022), além das demais coortes que aplicaram o mesmo instrumento. O estudo pioneiro de Oliveira *et al.* (2020), por anteceder a versão do método em cartas e adotar grupo focal com *Grounded Theory*, não compõe a matriz quantitativa e é retomado na análise qualitativa. A organização das seções acompanha a sequência analítica da metodologia: caracterização da amostra, estatística descritiva e normalidade, estrutura de correlação, adequação amostral, análise de componentes principais, detecção de observações atípicas, segmentação não supervisionada e comparações entre grupos. A análise qualitativa temática e a discussão integrada são apresentadas nas seções finais deste capítulo (a partir da Seção 6.10).

6.1 Composição da Amostra Consolidada

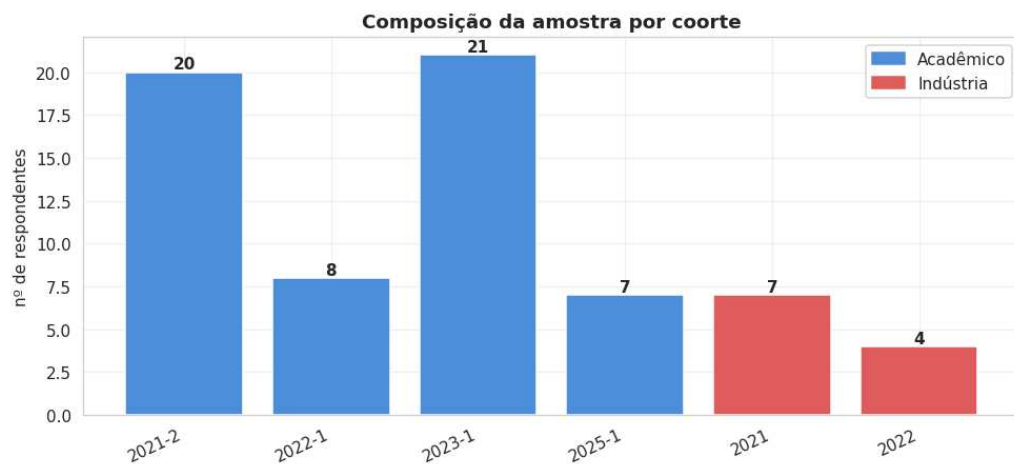
A base consolidada reuniu 68 registros provenientes das aplicações do instrumento comum. Identificou-se um único valor ausente entre as dez variáveis-núcleo (item de guia de prototipação, em respondente do contexto industrial), tratado conforme descrito na metodologia, de impacto desprezível (um valor em 680 células). Após as verificações de consistência e a classificação completa por contexto e coorte, a amostra analítica empregada nas etapas multivariadas compreendeu 67 respondentes. Desses, 56 pertencem ao contexto acadêmico e 11 ao contexto industrial, distribuídos em seis coortes, conforme a Figura 4 e a Tabela 4.

A predominância de respondentes acadêmicos reflete a natureza das aplicações disponíveis: enquanto os estudos industriais envolveram equipes de desenvolvimento de porte reduzido, com sete respondentes em Marques *et al.* (2023a) e poucos respondentes em Simões (2022), as aplicações acadêmicas reuniram turmas inteiras de Engenharia de Software organizadas em equipes, como em Fiori (2022). Essa assimetria amostral, com subamostra industrial pequena, é considerada explicitamente na interpretação das comparações entre contextos (Seção 6.8).

Cabe esclarecer a procedência das coortes posteriores à última publicação sobre o

método. O instrumento não foi aplicado apenas nas ocasiões reportadas nos estudos primários. Em especial, a coorte de 2025-1 é composta por integrantes do LUDI (Laboratório de Pesquisa e Desenvolvimento para Usabilidade, Diversidade e Inclusão), laboratório ao qual o método está vinculado, que utilizaram o USARP em projetos desenvolvidos no próprio laboratório. Esses participantes atuaram em diferentes funções, como analista de requisitos, agilista, desenvolvedor, projetista de UI/UX e profissional de garantia da qualidade. As respostas coletadas foram incorporadas a um registro unificado. Por se tratar de um ambiente universitário de pesquisa e desenvolvimento, essa coorte foi classificada como pertencente ao contexto acadêmico, enquanto a classificação de indústria foi reservada às equipes de empresas parceiras.

Figura 4 – Composição da amostra consolidada por coorte e contexto



Fonte: Elaborada pelo autor (2026).

Tabela 4 – Distribuição dos respondentes por contexto e coorte

Contexto	Coorte	n
Acadêmico	2021-2	20
Acadêmico	2022-1	8
Acadêmico	2023-1	21
Acadêmico	2025-1	7
Indústria	2021	7
Indústria	2022	4
Total		67

Fonte: Elaborada pelo autor (2026).

6.2 Caracterização Descritiva e Efeito de Teto

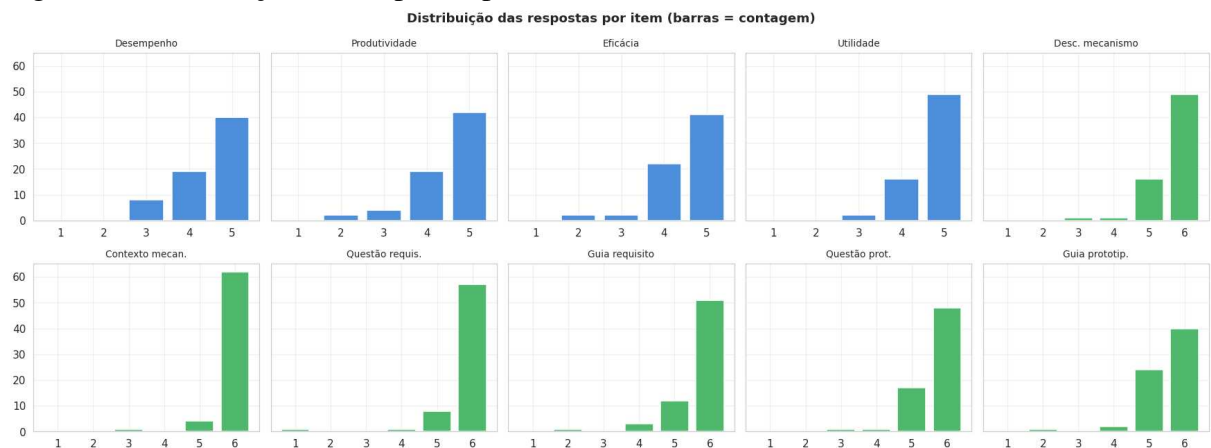
A Tabela 5 resume as estatísticas descritivas das dez variáveis-núcleo e a Figura 5 apresenta a distribuição das respostas por item. Os quatro itens de percepção geral, avaliados em escala de 1 a 5, concentram-se em valores elevados, com médias entre 4,48 e 4,70. Os seis itens referentes ao conteúdo das cartas, avaliados em escala de 1 a 6, situam-se ainda mais próximos do extremo superior, com médias entre 5,52 e 5,90. O item de contexto do mecanismo de usabilidade (cartas amarelas) apresenta a maior concentração, com média 5,90 e mediana 6.

Tabela 5 – Estatísticas descritivas das variáveis-núcleo

Item	Média	DP	Mín	Q1	Mediana	Q3	Máx
Desempenho	4,48	0,70	3	4	5	5	5
Produtividade	4,51	0,75	2	4	5	5	5
Eficácia	4,52	0,70	2	4	5	5	5
Utilidade	4,70	0,52	3	4	5	5	5
Desc. mecanismo	5,69	0,58	3	5	6	6	6
Contexto mecan.	5,90	0,43	3	6	6	6	6
Questão requis.	5,78	0,71	1	6	6	6	6
Guia requisito	5,67	0,70	2	6	6	6	6
Questão prot.	5,67	0,59	3	5	6	6	6
Guia prototip.	5,52	0,70	2	5	6	6	6

Fonte: Elaborada pelo autor (2026).

Figura 5 – Distribuição das respostas por item



Fonte: Elaborada pelo autor (2026).

Esse padrão evidencia um acentuado efeito de teto, isto é, a saturação das respostas nos níveis superiores das escalas. O fenômeno é coerente com os resultados individuais dos estudos primários, nos quais a percepção sobre o método foi consistentemente positiva: em Marques *et al.* (2023a), todos os respondentes manifestaram concordância quanto à utilidade e à

efetividade do método; em Fiori (2022), a maioria dos estudantes concordou parcial ou totalmente com as afirmações de desempenho, produtividade, eficácia e utilidade; e em Simões (2022), os participantes igualmente concordaram quanto ao potencial do método para a elicitação de requisitos de usabilidade. A base consolidada, portanto, expressa de forma agregada a avaliação favorável já observada isoladamente em cada cenário.

As Tabelas 6 e 7 detalham as médias por contexto e por coorte. As diferenças de média entre os contextos são pequenas em todos os itens, e a variação entre coortes também é modesta, antecipando o quadro de convergência confirmado nas análises inferenciais (Seção 6.8).

Tabela 6 – Médias das variáveis-núcleo por contexto

Contexto	Desemp.	Produt.	Eficácia	Utilid.	Desc. mec.	Cont. mec.	Q. requis.	Guia req.	Q. prot.	Guia prot.
Acadêmico	4,48	4,50	4,46	4,70	5,71	5,96	5,77	5,64	5,64	5,50
Indústria	4,45	4,55	4,82	4,73	5,55	5,55	5,82	5,82	5,82	5,64

Fonte: Elaborada pelo autor (2026).

Tabela 7 – Médias das variáveis-núcleo por coorte

Coorte	Desemp.	Produt.	Eficácia	Utilid.	Desc. mec.	Cont. mec.	Q. requis.	Guia req.	Q. prot.	Guia prot.
2021	4,71	4,86	4,86	4,86	5,86	5,86	5,86	5,71	5,86	5,71
2021-2	4,35	4,30	4,40	4,65	5,50	6,00	5,80	5,50	5,60	5,20
2022	4,00	4,00	4,75	4,50	5,00	5,00	5,75	6,00	5,75	5,50
2022-1	4,50	5,00	4,75	4,88	5,88	6,00	5,88	5,75	5,88	5,75
2023-1	4,52	4,38	4,33	4,62	5,81	6,00	5,86	5,90	5,57	5,67
2025-1	4,71	4,86	4,71	4,86	5,86	5,71	5,29	5,14	5,71	5,57

Fonte: Elaborada pelo autor (2026).

Quanto ao site de apoio, dos 67 respondentes, 15 declararam ter utilizado o recurso. Entre esses, as médias dos cinco aspectos avaliados foram 4,73 para aprendizagem, 4,60 para relembrar, 4,53 para interpretar, 4,40 para aplicação e 4,33 para *brainstorming*, em escala de 1 a 5. Como discutido na metodologia, essas variáveis foram tratadas apenas de forma descritiva, dada a ausência estrutural de respostas para quem não utilizou o recurso.

A Tabela 8 apresenta o teste de normalidade de Shapiro-Wilk e a assimetria de cada item. Todos os dez itens rejeitam a hipótese de normalidade ($p < 0,001$), e a assimetria é negativa em todos eles, o que confirma quantitativamente o efeito de teto. Esse diagnóstico fundamenta as escolhas inferenciais adotadas nas seções seguintes: a correlação de postos de Spearman e os testes não-paramétricos de Kruskal-Wallis e Mann-Whitney nas comparações entre grupos (Field, 2018).

Tabela 8 – Teste de normalidade de Shapiro-Wilk e assimetria por item

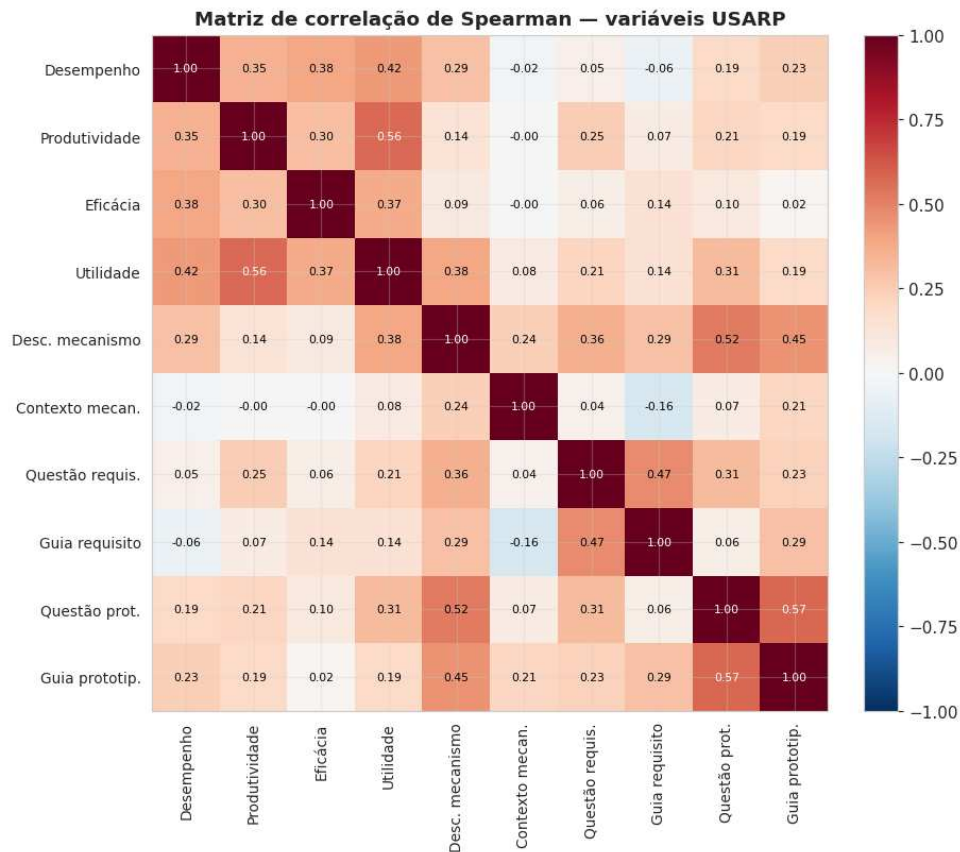
Item	Assimetria	Estatística W	Conclusão
Desempenho	-0,97	0,705	não-normal
Produtividade	-1,57	0,677	não-normal
Eficácia	-1,65	0,667	não-normal
Utilidade	-1,50	0,590	não-normal
Desc. mecanismo	-2,15	0,571	não-normal
Contexto mecan.	-5,16	0,261	não-normal
Questão requis.	-4,94	0,346	não-normal
Guia requisito	-2,85	0,524	não-normal
Questão prot.	-2,05	0,585	não-normal
Guia prototip.	-2,18	0,641	não-normal

Fonte: Elaborada pelo autor (2026). Nota: todos os itens apresentaram $p < 0,001$.

6.3 Estrutura de Correlação entre as Variáveis

A Figura 6 apresenta a matriz de correlação de Spearman entre as dez variáveis-núcleo. A matriz foi estimada sobre a amostra completa, isto é, sobre os 67 respondentes, sem separação por contexto ou por coorte, conforme justificado na Seção 5.7.2: esta etapa é diagnóstica e prepara a PCA, que também é ajustada sobre a matriz completa, e a estimação de uma matriz de dez variáveis em subgrupos pequenos, como a subamostra industrial de onze respondentes, produziria coeficientes instáveis. As diferenças entre contextos e coortes são investigadas na Seção 6.8, por testes apropriados ao tamanho da amostra. A correlação média, em módulo, fora da diagonal, é de 0,225, e a correlação máxima observada é de 0,572. As associações mais fortes ocorrem dentro de cada bloco de itens. No bloco de percepção geral, destacam-se as correlações entre utilidade e produtividade ($\rho = 0,56$) e entre utilidade e desempenho ($\rho = 0,42$). No bloco das cartas, sobressaem as correlações entre questão de prototipação e guia de prototipação ($\rho = 0,57$), entre descrição do mecanismo e questão de prototipação ($\rho = 0,52$) e entre questão e guia de requisitos ($\rho = 0,47$). As correlações entre os dois blocos são, em geral, fracas, e o item de contexto do mecanismo mostra-se praticamente independente dos demais, inclusive com associação ligeiramente negativa com o guia de requisitos ($\rho = -0,16$). Esse padrão de blocos relativamente coesos e fracamente associados entre si antecipa a estrutura latente revelada pela análise de componentes principais.

Figura 6 – Matriz de correlação de Spearman entre as variáveis-núcleo



Fonte: Elaborada pelo autor (2026).

6.4 Adequação Amostral à Análise de Componentes Principais

Os diagnósticos de adequação amostral apresentaram resultados divergentes, o que orienta a interpretação subsequente. O teste de esfericidade de Bartlett rejeitou fortemente a hipótese de ausência de correlação ($\chi^2 = 238,6$; $gl = 45$; $p = 3,54 \times 10^{-28}$), indicando que existe estrutura a ser extraída (Bartlett, 1950). A medida de Kaiser-Meyer-Olkin, por sua vez, resultou em $KMO = 0,461$, valor classificado como inaceitável segundo os limiares usuais (Kaiser, 1974; Field, 2018). Esse contraste é coerente com a natureza dos dados: em amostras pequenas e com forte efeito de teto, diversas variáveis saturam no valor máximo e reduzem a variância compartilhável, deprimindo o KMO. Por essa razão, e em consonância com Hair *et al.* (2019), a análise de componentes principais é empregada neste trabalho como técnica exploratória e descritiva, voltada a resumir a variabilidade e a identificar agrupamentos de itens, e não como análise fatorial confirmatória. Essa ressalva é mantida em toda a interpretação dos componentes.

6.5 Análise de Componentes Principais

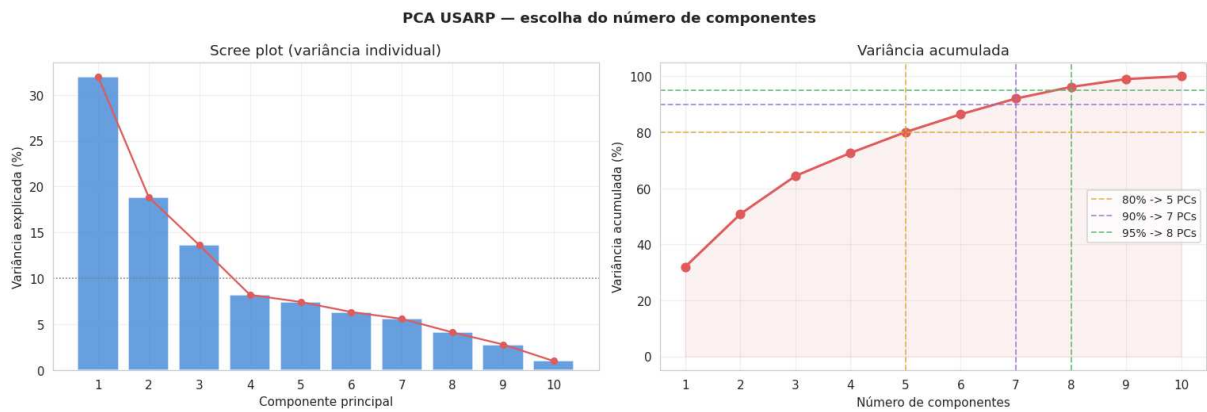
A Tabela 9 apresenta os autovalores e a variância explicada por componente, e a Figura 7 exibe o *scree plot* e a variância acumulada. Pelo critério de Kaiser (autovalor superior a um), retêm-se três componentes, que acumulam 64,4% da variância (Kaiser, 1960). Os limiares de variância acumulada de 80%, 90% e 95% são atingidos, respectivamente, com cinco, sete e oito componentes. O erro de reconstrução (Figura 8) decresce de forma suave, sem um cotovelo abrupto: três componentes recompõem os dados com erro quadrático médio de aproximadamente 0,356, e cinco componentes reduzem esse erro para cerca de 0,199. A ausência de um ponto de inflexão nítido é típica de dados com efeito de teto e correlações moderadas, nos quais a informação se distribui por várias dimensões, o que reforça o caráter exploratório da análise. Conforme a metodologia, adotaram-se três componentes para a interpretação substantiva e cinco componentes para as análises que dependem de mais dimensões, como a segmentação e a matriz de dispersão.

Tabela 9 – Autovalores e variância explicada por componente principal

Componente	Autovalor	Variância (%)	Acumulada (%)
PC1	3,242	31,94	31,94
PC2	1,913	18,85	50,78
PC3	1,383	13,62	64,40
PC4	0,835	8,22	72,63
PC5	0,756	7,45	80,07
PC6	0,645	6,35	86,43
PC7	0,569	5,61	92,03
PC8	0,420	4,14	96,17
PC9	0,286	2,82	98,99
PC10	0,103	1,01	100,00

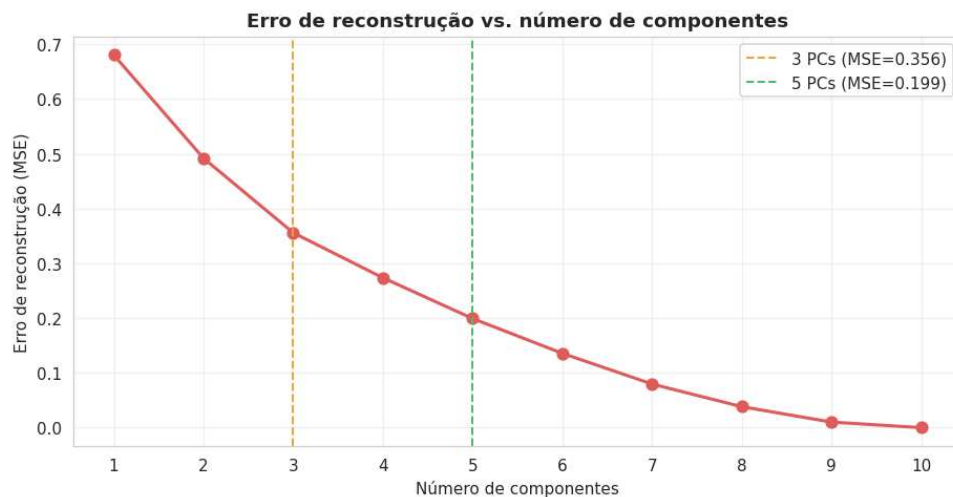
Fonte: Elaborada pelo autor (2026). Nota: três componentes retidos pelo critério de Kaiser; cinco componentes atingem 80% da variância.

Figura 7 – Escolha do número de componentes: scree plot e variância acumulada



Fonte: Elaborada pelo autor (2026).

Figura 8 – Erro de reconstrução em função do número de componentes



Fonte: Elaborada pelo autor (2026).

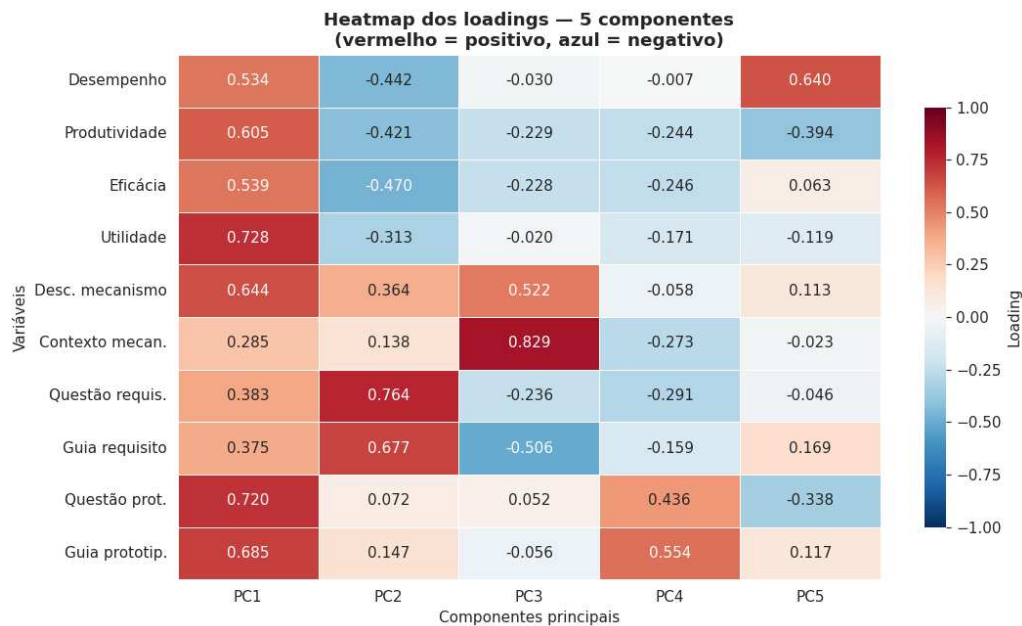
A Tabela 10 e a Figura 9 apresentam as cargas fatoriais das cinco primeiras componentes, que orientam a interpretação substantiva.

Tabela 10 – Cargas fatoriais (loadings) das cinco primeiras componentes

Variável	PC1	PC2	PC3	PC4	PC5
Desempenho	0,534	-0,442	-0,030	-0,007	0,640
Produtividade	0,605	-0,421	-0,229	-0,244	-0,394
Eficácia	0,539	-0,470	-0,228	-0,246	0,063
Utilidade	0,728	-0,313	-0,020	-0,171	-0,119
Desc. mecanismo	0,644	0,364	0,522	-0,058	0,113
Contexto mecan.	0,285	0,138	0,829	-0,273	-0,023
Questão requis.	0,383	0,764	-0,236	-0,291	-0,046
Guia requisito	0,375	0,677	-0,506	-0,159	0,169
Questão prot.	0,720	0,072	0,052	0,436	-0,338
Guia prototip.	0,685	0,147	-0,056	0,554	0,117

Fonte: Elaborada pelo autor (2026).

Figura 9 – Mapa de calor das cargas fatoriais das cinco primeiras componentes



Fonte: Elaborada pelo autor (2026).

A primeira componente (31,9% da variância) apresenta cargas positivas em todas as variáveis, com maior peso em utilidade (0,728), questão de prototipação (0,720), guia de prototipação (0,685) e descrição do mecanismo (0,644). Trata-se, portanto, de um eixo de percepção positiva geral do método: respondentes que avaliam bem um aspecto tendem a avaliar bem os demais. A segunda componente (18,8%) opõe os itens das cartas de requisitos, questão de requisitos (0,764) e guia de requisitos (0,677), aos itens de percepção geral, eficácia (-0,470), desempenho (-0,442) e produtividade (-0,421), configurando um contraste entre a utilidade percebida das cartas de requisitos e a percepção de eficiência operacional do método. A terceira componente (13,6%) é dominada pelo contexto do mecanismo (0,829) e pela descrição do mecanismo (0,522), em oposição ao guia de requisitos (-0,506), capturando a dimensão das cartas de mecanismo, ligada à contextualização. As componentes seguintes têm peso reduzido: a quarta associa-se à prototipação (guia 0,554 e questão 0,436) e a quinta a um contraste interno do bloco de percepção, liderado pelo desempenho (0,640).

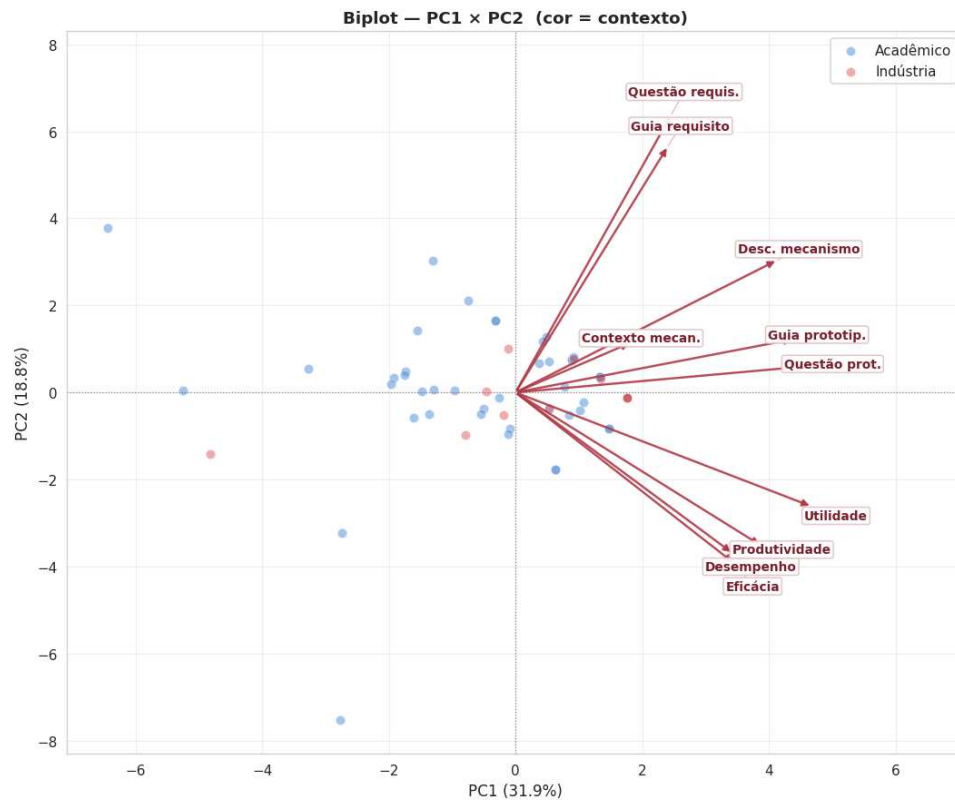
As Figuras 10, 11 e 12 apresentam os *biplots* das combinações das três primeiras componentes, e as Figuras 13, 14 e 15 exibem as projeções dos respondentes por contexto e por coorte. A Figura 16 apresenta a projeção tridimensional e a Figura 17 a matriz de dispersão das três primeiras componentes.

Como esses gráficos concentram boa parte da interpretação da PCA, convém explicitar sua leitura. Um *biplot* sobrepõe, em um mesmo plano formado por duas componentes, dois

elementos de naturezas distintas. Os **pontos** representam os respondentes, posicionados segundo seus escores nas duas componentes do plano e coloridos conforme o contexto de aplicação; dois pontos próximos indicam participantes com perfis de resposta semelhantes. Os **vetores** representam as dez variáveis originais, projetadas no mesmo plano a partir de suas cargas fatoriais. A leitura dos vetores segue três convenções: a **direção** indica o sentido em que a variável cresce, de modo que respondentes situados ao longo dela tendem a atribuir notas mais altas naquele item; o **comprimento** indica o quanto a variável é bem representada naquele plano, sendo os vetores mais longos os que mais contribuem para as componentes exibidas; e o **ângulo** entre dois vetores aproxima a associação entre as variáveis, pois ângulos pequenos sugerem itens que variam em conjunto, ângulos próximos de noventa graus sugerem itens praticamente independentes e ângulos próximos de cento e oitenta graus sugerem itens que variam em sentidos opostos.

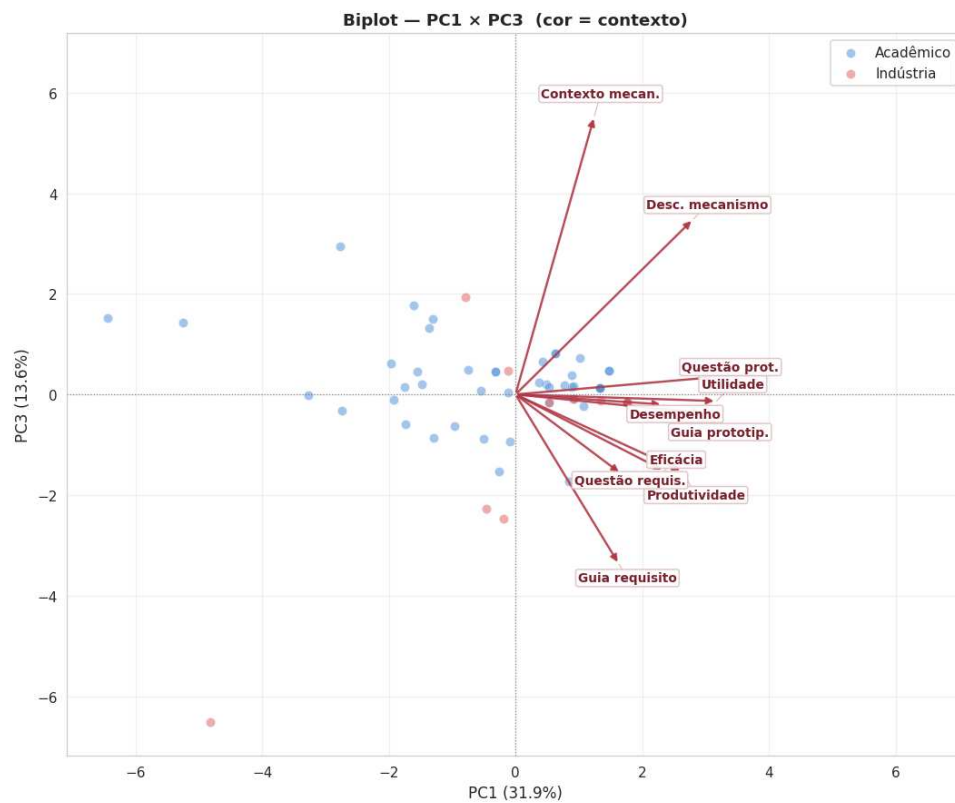
Nos três biplots, todos os vetores das variáveis projetam-se sobre o lado positivo de PC1, o que confirma visualmente o eixo de percepção positiva geral: independentemente do plano observado, as setas apontam para a direita, de modo que escores altos em PC1 correspondem a avaliações favoráveis em todos os itens. A nuvem de respondentes concentra-se majoritariamente nesse lado positivo, com poucos casos deslocados para a esquerda, padrão coerente com o efeito de teto e com a predominância de percepções favoráveis. No plano PC1 e PC2 (Figura 10), os vetores de questão e guia de requisitos apontam para cima e formam entre si um ângulo pequeno, enquanto utilidade, produtividade, desempenho e eficácia apontam para baixo, materializando o contraste identificado nas cargas: a valorização das cartas de requisitos varia em sentido oposto à percepção de eficiência operacional. Nos planos que incluem PC3 (Figuras 11 e 12), o contexto e a descrição do mecanismo destacam-se na direção vertical, isolando a dimensão das cartas de mecanismo das demais variáveis. Quanto ao contexto de aplicação, os pontos acadêmicos e de indústria sobrepõem-se amplamente nos três planos, sem formar agrupamentos distintos, com os respondentes de indústria situados dentro da nuvem acadêmica. Não há, assim, indício visual de que o contexto module o perfil de resposta.

Figura 10 – Biplot das componentes PC1 e PC2, por contexto



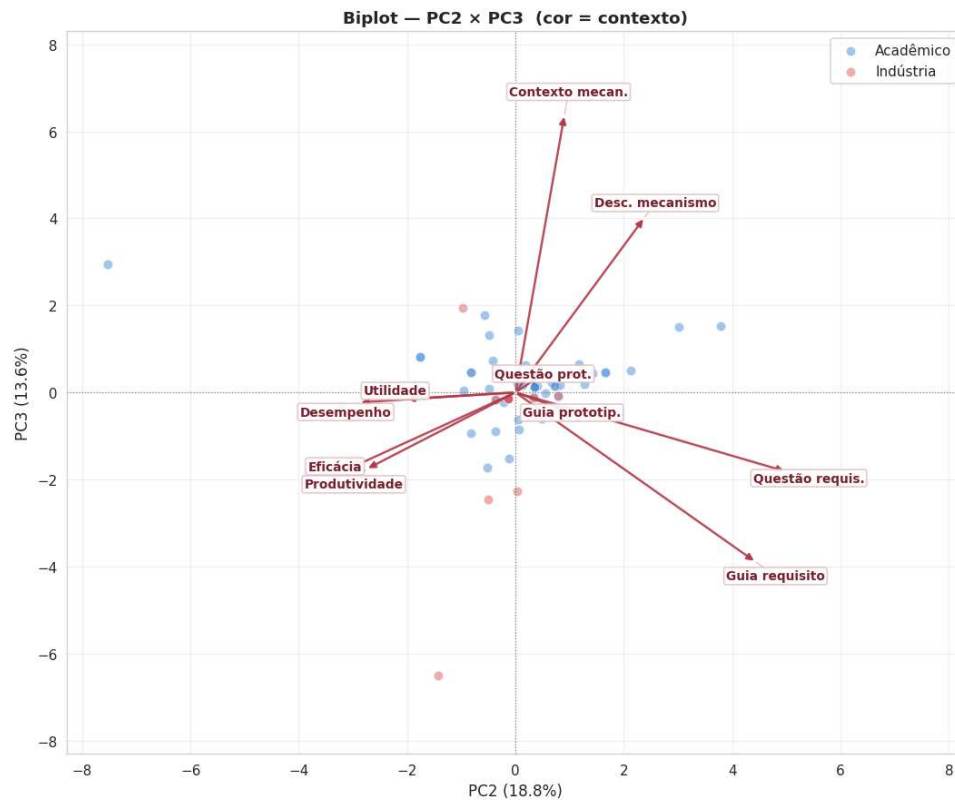
Fonte: Elaborada pelo autor (2026).

Figura 11 – Biplot das componentes PC1 e PC3, por contexto



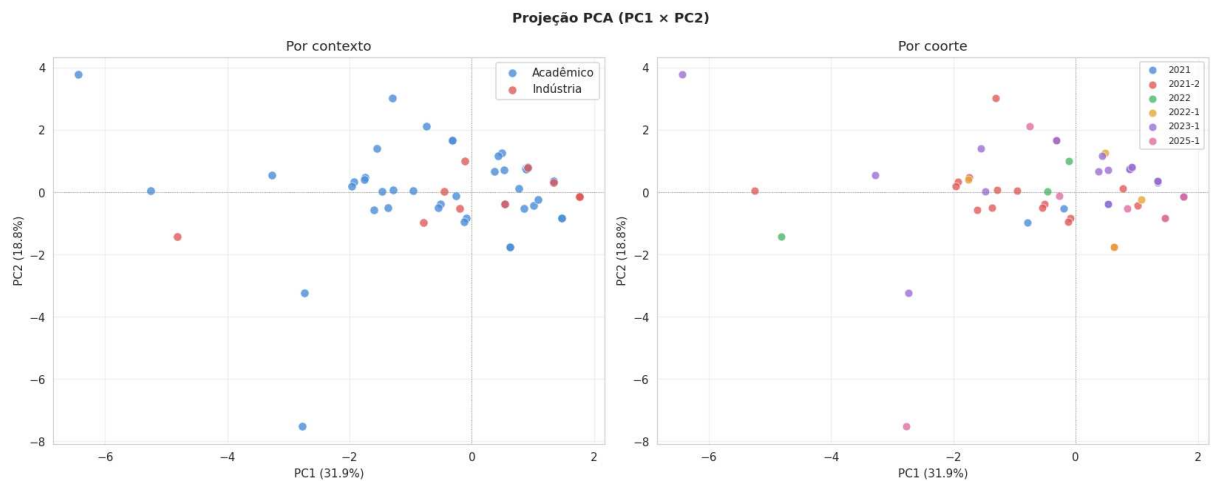
Fonte: Elaborada pelo autor (2026).

Figura 12 – Biplot das componentes PC2 e PC3, por contexto



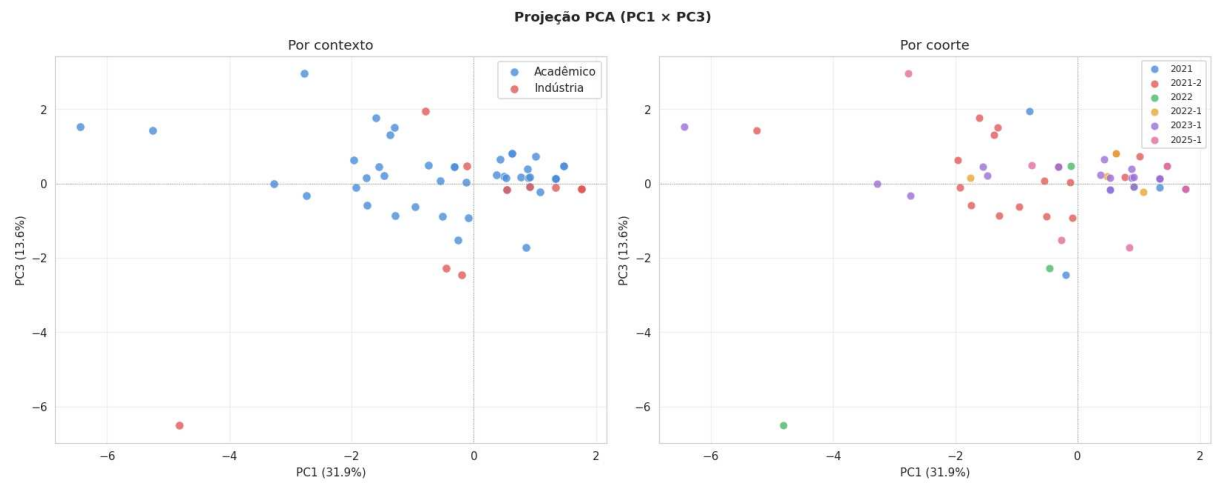
Fonte: Elaborada pelo autor (2026).

Figura 13 – Projeção dos respondentes nas componentes PC1 e PC2, por contexto e por coorte



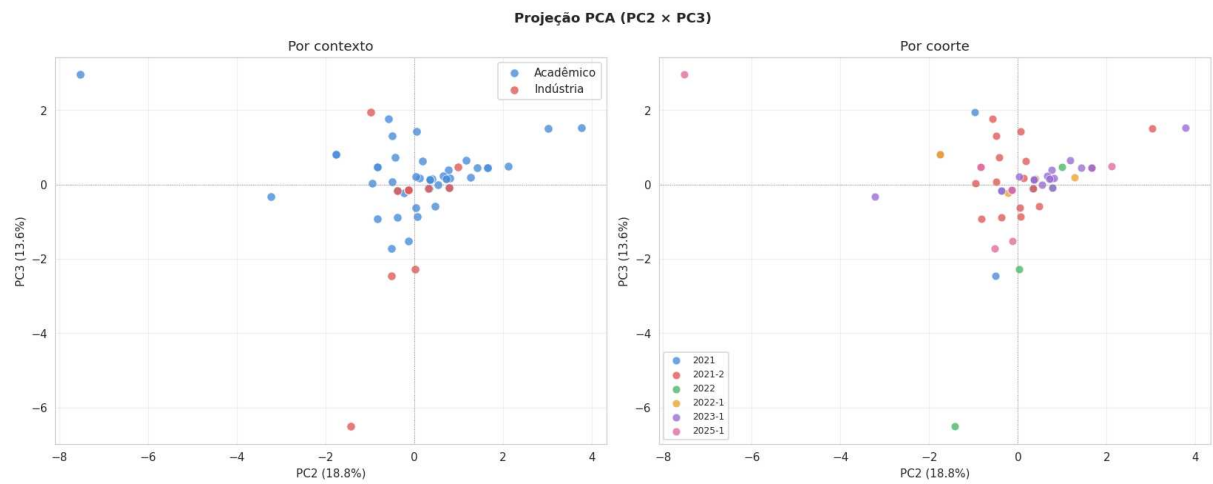
Fonte: Elaborada pelo autor (2026).

Figura 14 – Projeção dos respondentes nas componentes PC1 e PC3, por contexto e por coorte



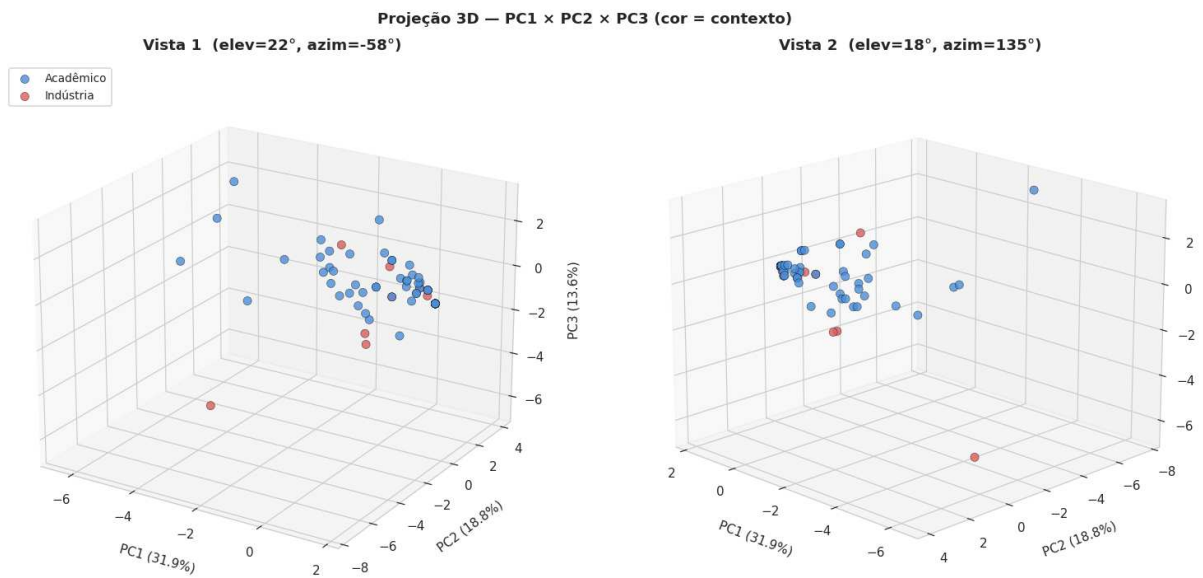
Fonte: Elaborada pelo autor (2026).

Figura 15 – Projeção dos respondentes nas componentes PC2 e PC3, por contexto e por coorte



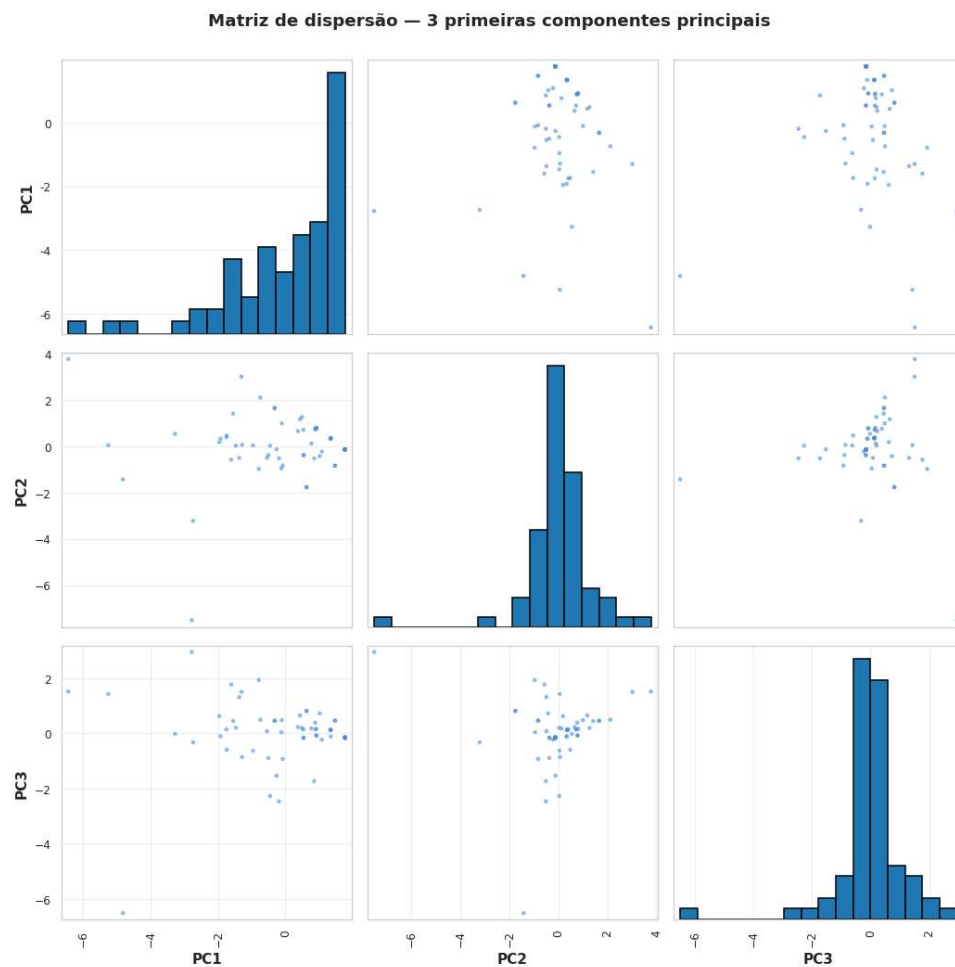
Fonte: Elaborada pelo autor (2026).

Figura 16 – Projeção tridimensional das três primeiras componentes, por contexto



Fonte: Elaborada pelo autor (2026).

Figura 17 – Matriz de dispersão das três primeiras componentes principais



Fonte: Elaborada pelo autor (2026).

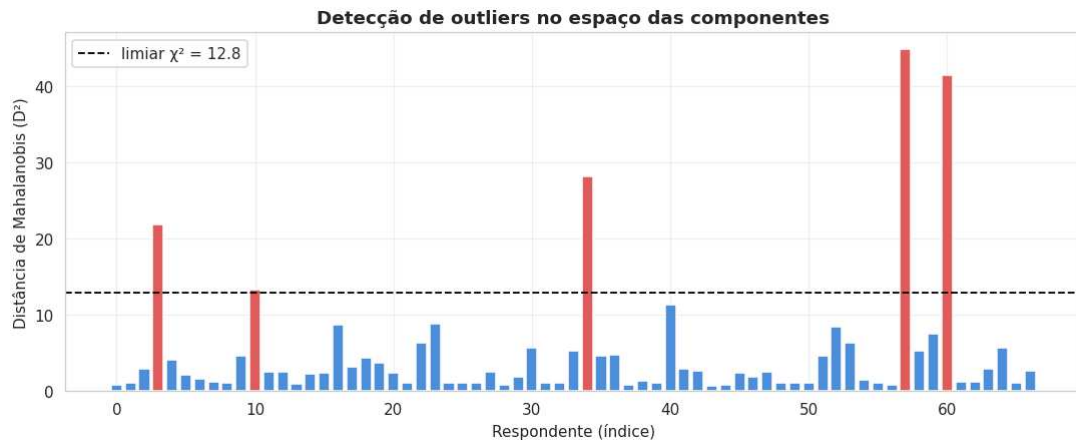
As projeções não revelam separação nítida entre os contextos acadêmico e industrial, nem entre as coortes: os respondentes dos dois contextos ocupam, em larga medida, a mesma região do espaço reduzido, o que é compatível com a percepção convergente sugerida pelas estatísticas descritivas. A matriz de dispersão evidencia, na diagonal, o formato das distribuições dos escores: a primeira componente concentra a maioria dos respondentes em uma faixa estreita à direita, com uma cauda à esquerda formada por poucos respondentes de avaliação geral destoante, que reaparecem como candidatos a observações atípicas na Seção 6.6. As componentes seguintes são mais simétricas e centradas em zero. A leitura geométrica dos *biplots* corrobora a estrutura de blocos: os vetores dos itens de percepção geral (desempenho, produtividade, eficácia e utilidade) agrupam-se em uma direção, os das cartas de requisitos em outra, e o vetor do contexto do mecanismo aparece destacado na terceira componente.

Esses resultados dialogam com os achados dos estudos primários. O eixo de percepção positiva geral (PC1) corresponde à avaliação uniformemente favorável reportada por Marques *et al.* (2023a), Fiori (2022) e Simões (2022). Por sua vez, os contrastes capturados por PC2 e PC3, que distinguem o bloco das cartas de requisitos, o bloco de mecanismos e os itens de percepção operacional, são consistentes com as observações qualitativas desses estudos sobre o uso diferenciado dos tipos de carta, em particular a menor utilização das cartas de prototipação e a sensibilidade da carta de contexto do mecanismo, retomadas na análise qualitativa.

6.6 Detecção de Observações Atípicas

A Figura 18 apresenta a distância de Mahalanobis de cada respondente no espaço das cinco componentes retidas. Adotando o limiar $\chi^2_{0,975}(gl = 5) \approx 12,83$, cinco dos 67 respondentes (cerca de 7%) foram sinalizados como atípicos: os respondentes de índice 3 (Acadêmico 2021-2; $D^2 = 21,87$), 10 (Acadêmico 2021-2; $D^2 = 13,22$), 34 (Acadêmico 2023-1; $D^2 = 28,14$), 57 (Indústria 2022; $D^2 = 44,77$) e 60 (Acadêmico 2025-1; $D^2 = 41,35$). Os casos mais extremos combinam avaliação geral mais baixa, situada na cauda à esquerda da primeira componente, com um perfil incomum de avaliação das cartas. Coerentemente com a natureza diagnóstica da análise, esses pontos não foram removidos: em uma amostra pequena, descartá-los reduziria ainda mais o poder estatístico e poderia eliminar justamente parte da variação industrial de interesse. As observações foram registradas como ressalva de robustez e como insumo para a análise qualitativa.

Figura 18 – Detecção de observações atípicas pela distância de Mahalanobis



Fonte: Elaborada pelo autor (2026).

6.7 Segmentação Não Supervisionada dos Perfis de Percepção

A Figura 19 apresenta os índices internos de validação para soluções de dois a oito grupos. O coeficiente de silhueta é máximo em $k = 2$ (0,403), e o índice de Davies-Bouldin corrobora essa leitura, de modo que se adotou a solução com dois grupos (Rousseeuw, 1987; Davies; Bouldin, 1979). A Figura 20 ilustra a segmentação nos planos das componentes de Kaiser, com um primeiro grupo de 17 respondentes e um segundo grupo de 50 respondentes. Para permitir a interpretação substantiva desses grupos, e não apenas sua visualização no espaço reduzido, a Tabela 11 apresenta as médias das dez variáveis originais em cada um deles, acompanhadas do contraste de Mann-Whitney item a item.

Tabela 11 – Perfil dos grupos identificados pela segmentação k-means

Item	Grupo 1 (n = 17)	Grupo 2 (n = 50)	Dif.	p
Desempenho	4,06	4,62	+0,56	0,003
Produtividade	3,88	4,72	+0,84	<0,001
Eficácia	4,00	4,70	+0,70	0,003
Utilidade	4,12	4,90	+0,78	<0,001
Desc. mecanismo	5,18	5,86	+0,68	<0,001
Contexto mecan.	5,82	5,92	+0,10	0,837
Questão requis.	5,24	5,96	+0,72	<0,001
Guia requisito	5,12	5,86	+0,74	0,001
Questão prot.	5,12	5,86	+0,74	<0,001
Guia prototip.	5,00	5,70	+0,70	0,001

Fonte: Elaborada pelo autor (2026). Nota: Dif. = Grupo 2 menos Grupo 1; p refere-se ao teste U de Mann-Whitney entre os dois grupos.

O perfil dos grupos é inequívoco e tem uma leitura direta. O **Grupo 2**, majoritário, reúne 50 respondentes cuja avaliação é sistematicamente mais elevada, com médias entre 4,62 e 5,96, próximas do extremo superior das escalas; pode ser caracterizado como o perfil de **avaliação**

entusiástica do método. O **Grupo 1**, minoritário, reúne 17 respondentes cuja avaliação, embora ainda positiva, é consistentemente mais contida, com médias entre 3,88 e 5,82; corresponde ao perfil de **avaliação moderada**. A diferença entre os grupos é transversal: ela se manifesta em nove dos dez itens, todos com diferença significativa, e não se concentra em um bloco específico de cartas ou em uma dimensão isolada do TAM. Em outras palavras, os dois perfis não expressam avaliações qualitativamente distintas do método, mas dois níveis de entusiasmo ao longo de um mesmo eixo de percepção positiva, exatamente o eixo capturado pela primeira componente principal.

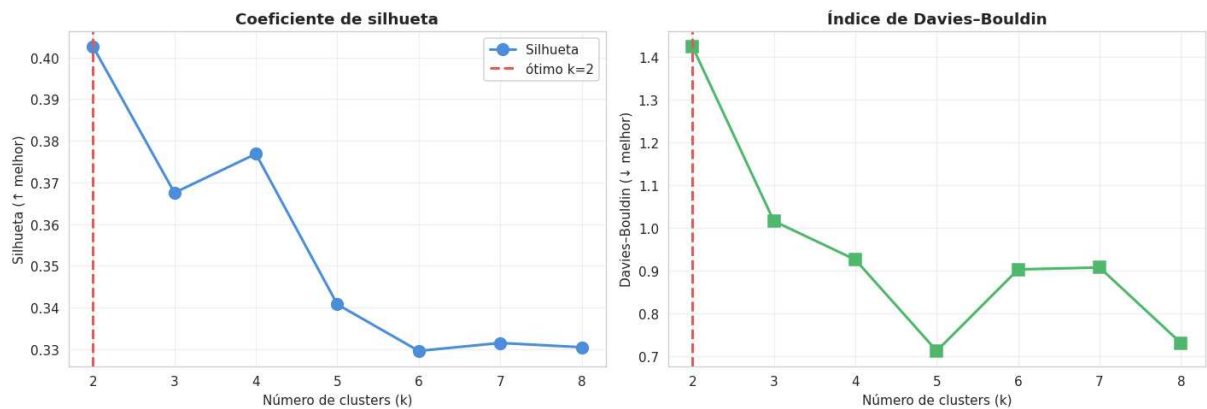
Duas observações reforçam essa leitura. A primeira é que o único item que não distingue os grupos é justamente o contexto do mecanismo de usabilidade (5,82 contra 5,92; $p = 0,837$), item que satura no teto em ambos e que, por isso, não discrimina perfis de avaliação. Trata-se, do mesmo item que constitui a única diferença robusta entre coortes (Seção 6.8), o que sugere que ele varia com a familiaridade com o método, e não com o nível geral de entusiasmo do respondente. A segunda é que ambos os grupos são compostos por respondentes de todas as seis coortes e dos dois contextos, conforme a Tabela 12, o que confirma, por uma via não supervisionada, que a segmentação reflete o nível de avaliação individual e não a origem do respondente.

Tabela 12 – Composição dos grupos por contexto e coorte

Contexto	Coorte	Grupo 1	Grupo 2
Acadêmico	2021-2	8	12
Acadêmico	2022-1	1	7
Acadêmico	2023-1	5	16
Acadêmico	2025-1	1	6
Indústria	2021	1	6
Indústria	2022	1	3
Total		17	50

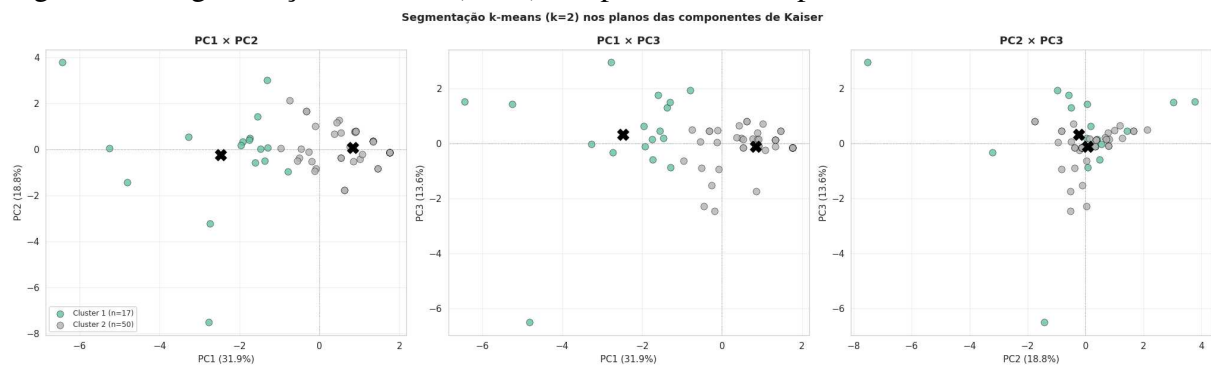
Fonte: Elaborada pelo autor (2026).

Figura 19 – Seleção do número de grupos: coeficiente de silhueta e índice de Davies-Bouldin



Fonte: Elaborada pelo autor (2026).

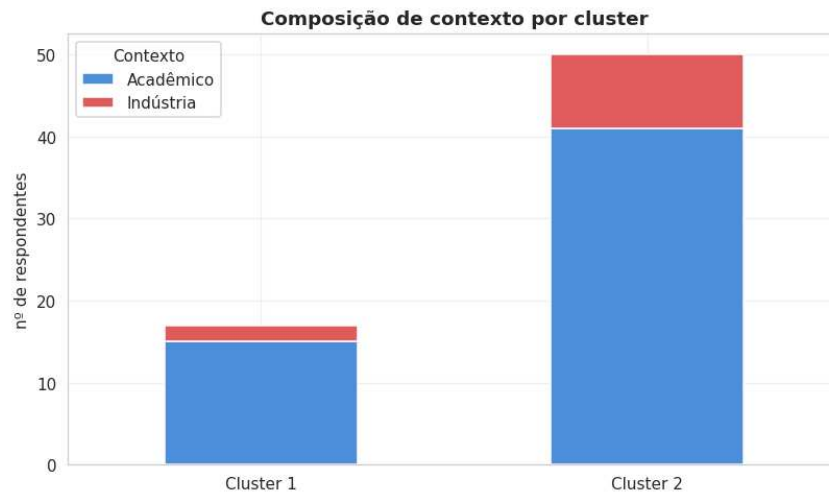
Figura 20 – Segmentação k-means (k = 2) nos planos das componentes de Kaiser



Fonte: Elaborada pelo autor (2026).

A Figura 21 apresenta a composição de contexto por grupo. O primeiro grupo reúne 15 respondentes acadêmicos e 2 industriais (88,2% e 11,8%), e o segundo, 41 acadêmicos e 9 industriais (82,0% e 18,0%). O teste qui-quadrado de associação entre grupo e contexto não foi significativo ($\chi^2 = 0,05$; $gl = 1$; $p = 0,825$), resultado a ser interpretado com cautela em razão das contagens esperadas reduzidas. A segmentação, portanto, não é definida pelo contexto de aplicação: os dois perfis refletem essencialmente o nível geral de avaliação ao longo da primeira componente, e não a distinção entre ambientes acadêmicos e industriais. Esse achado reforça, por uma via não supervisionada, a convergência de percepção entre os cenários.

Figura 21 – Composição de contexto por grupo (cluster)



Fonte: Elaborada pelo autor (2026).

6.8 Comparações entre Grupos

As comparações foram conduzidas sobre os escores das componentes principais e, em seguida, item a item, sempre com testes não-paramétricos e correção de Holm para múltiplas comparações (Holm, 1979). A Tabela 13 apresenta o teste de Kruskal-Wallis entre as coortes, por componente, e a Tabela 14 apresenta o teste de Mann-Whitney entre os contextos acadêmico e industrial, com o respectivo tamanho de efeito.

Tabela 13 – Kruskal-Wallis entre coortes, por componente principal

Componente	H	p	p _{Holm}	Signif.
PC1	10,022	0,0746	0,1492	não
PC2	12,028	0,0344	0,1032	não
PC3	4,270	0,5112	0,5112	não

Fonte: Elaborada pelo autor (2026).

Tabela 14 – Mann-Whitney entre contextos (Acadêmico e Indústria), por componente principal

Comp.	Med. Acad.	Med. Ind.	U	p	r	p _{Holm}	Signif.
PC1	0,53	0,53	274	0,5747	0,069	0,6993	não
PC2	-0,04	-0,13	363	0,3496	0,115	0,6993	não
PC3	0,13	-0,15	417	0,0638	0,226	0,1914	não

Fonte: Elaborada pelo autor (2026).

A leitura componente a componente é consistente. A primeira componente, eixo de percepção positiva geral, não difere entre coortes (Kruskal-Wallis $H = 10,02$; $p = 0,075$) nem entre contextos (Mann-Whitney $U = 274$; $p = 0,575$), com tamanho de efeito desprezível ($r =$

0,07), o que indica um valor percebido alto e homogêneo entre os cenários, coerente com o efeito de teto. A segunda componente apresenta diferença bruta entre coortes ($H = 12,03$; $p = 0,034$), que deixa de ser significativa após a correção de Holm ($p_{\text{Holm}} = 0,103$). A terceira componente exibe o maior efeito da tabela na comparação entre contextos ($r = 0,23$, pequeno a moderado) e fica próxima do limiar ($p = 0,064$), mas também não resiste à correção ($p_{\text{Holm}} = 0,191$). Em síntese, após a correção de Holm, nenhuma componente difere de forma robusta entre os grupos, e os únicos sinais de contraste mostram-se frágeis frente ao controle de múltiplas comparações.

A análise item a item (Tabela 15) confirma esse quadro. Nove dos dez itens são estatisticamente homogêneos entre as coortes após a correção de Holm. A produtividade chega a apresentar diferença bruta ($H = 13,04$; $p = 0,023$), mas não sobrevive à correção ($p_{\text{Holm}} = 0,207$). O único item que permanece significativo é o contexto do mecanismo de usabilidade (carta amarela: $H = 19,46$; $p_{\text{Holm}} = 0,016$). A análise *post-hoc* de Mann-Whitney par a par, ajustada por Holm, isola as coortes responsáveis pela diferença: Indústria 2022 contra Acadêmico 2023-1 ($p_{\text{Holm}} = 0,019$) e Indústria 2022 contra Acadêmico 2021-2 ($p_{\text{Holm}} = 0,023$).

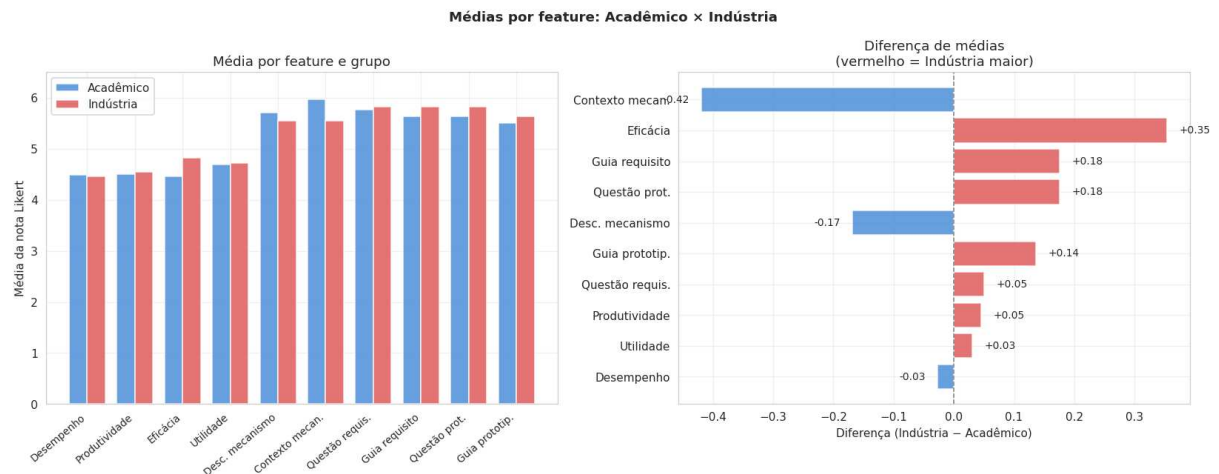
Tabela 15 – Kruskal-Wallis entre coortes, item a item

Item	H	p	p _{Holm}	Signif.
Desempenho	5,34	0,3755	1,0000	não
Produtividade	13,04	0,0230	0,2068	não
Eficácia	5,47	0,3615	1,0000	não
Utilidade	3,59	0,6099	1,0000	não
Desc. mecanismo	9,78	0,0817	0,6536	não
Contexto mecan.	19,46	0,0016	0,0158	sim
Questão requis.	1,06	0,9580	1,0000	não
Guia requisito	8,21	0,1452	1,0000	não
Questão prot.	2,84	0,7245	1,0000	não
Guia prototip.	6,02	0,3042	1,0000	não

Fonte: Elaborada pelo autor (2026).

A Figura 22 compara as médias por item entre os contextos e ordena as diferenças. A maior diferença ocorre justamente no contexto do mecanismo, mais bem avaliado no ambiente acadêmico (5,96) do que no industrial (5,55), uma diferença de 0,42. A eficácia é o item em que a indústria supera a academia (4,82 contra 4,46, diferença de 0,35), e o desempenho é o item mais similar entre os contextos (diferença de 0,03). É importante notar que, à exceção do contexto do mecanismo, nenhuma dessas diferenças resiste ao controle de múltiplas comparações.

Figura 22 – Médias por item entre contextos e diferença ordenada (Indústria menos Acadêmico)



Fonte: Elaborada pelo autor (2026).

A diferença pontual no contexto do mecanismo de usabilidade é plausível à luz da evolução do próprio método ao longo do tempo. Como descrito por Oliveira *et al.* (2020), as cartas e o mecanismo de contextualização foram introduzidos e refinados após o estudo de viabilidade inicial, de modo que o grau de familiaridade e o preparo das diferentes turmas e equipes variaram entre as coortes. A coorte industrial de 2022 avalia esse item de maneira distinta das turmas acadêmicas, o que sugere que a dimensão em que os grupos mais divergem é precisamente a mais sensível à maturidade de uso do método. Esse achado dialoga diretamente com o tema qualitativo relativo ao contexto de uso e aos mecanismos, retomado na análise qualitativa.

Cabe registrar a ressalva metodológica que perpassa estas comparações. A subamostra industrial é pequena ($n = 11$), o que reduz o poder estatístico: a ausência de significância não comprova equivalência entre os grupos, mas apenas indica ausência de evidência de diferença (Field, 2018). Os tamanhos de efeito foram reportados justamente para qualificar essa leitura (Hair *et al.*, 2019), e a correção de Holm foi adotada por ser uniformemente mais poderosa do que a de Bonferroni, mantendo o mesmo controle do erro do tipo I por família de comparações (Holm, 1979). Além disso, com seis coortes de tamanhos desiguais, o número de comparações par a par é elevado e o poder é limitado, de modo que os resultados da etapa *post-hoc* devem ser tratados como indicativos, e não conclusivos.

6.9 Síntese dos Resultados Quantitativos

Os resultados quantitativos convergem para um quadro consistente. A percepção sobre o USARP é elevada e homogênea entre contextos e coortes, com forte efeito de teto em todos os itens, o que reproduz, de forma agregada, as avaliações favoráveis reportadas individualmente por Marques *et al.* (2023a), Fiori (2022) e Simões (2022). A redução de dimensionalidade revela uma primeira dimensão de percepção positiva geral, acompanhada de contrastes secundários entre os blocos de cartas (requisitos, mecanismos e prototipação) e os itens de percepção operacional. Nem a segmentação não supervisionada, que separa os respondentes por nível geral de avaliação e não por contexto, nem as comparações inferenciais, que não acusam diferenças robustas após a correção de Holm, sustentam uma distinção marcante entre ambientes acadêmicos e industriais. A única diferença que resiste ao controle estatístico, no item de contexto do mecanismo de usabilidade, aponta para a sensibilidade desse mecanismo à maturidade de uso do método, e não para uma divergência ampla de percepção.

Esses achados respondem parcialmente aos objetivos do trabalho: confirmam a percepção favorável de utilidade, produtividade e eficácia (primeiro objetivo específico) e evidenciam mais semelhanças do que diferenças entre os contextos acadêmico e industrial (terceiro objetivo específico), ao mesmo tempo em que a redução de dimensionalidade explicita as estruturas latentes associadas à aplicação do método (quarto objetivo específico). A caracterização do uso percebido dos mecanismos e a discussão das implicações práticas e limitações dependem, contudo, da leitura conjunta com as evidências qualitativas, apresentada nas seções seguintes (a partir da Seção 6.10), que aprofundam os padrões de recorrência e de subutilização das cartas e contextualizam os contrastes aqui identificados.

6.10 Análise Qualitativa Temática

Esta seção apresenta a análise qualitativa das respostas abertas, conduzida conforme a metodologia: codificação temática indutiva, com dicionário de termos explícito, atribuição de rótulos múltiplos e classificação de polaridade. Os trechos transcritos ao longo da seção são identificados pelo código do respondente no registro consolidado, na forma P_n , seguido do contexto e da coorte de origem, o que permite rastrear cada citação até a resposta correspondente, preservando o anonimato dos participantes. O corpus reúne as respostas abertas dos estudos que adotaram o instrumento comum, e a leitura é articulada com as evidências qualitativas reportadas

nos quatro estudos primários, em especial as categorias identificadas por Oliveira *et al.* (2020) a partir de grupo focal e as codificações por *Grounded Theory* de Fiori (2022) e Simões (2022).

6.10.1 Caracterização do Corpus Qualitativo

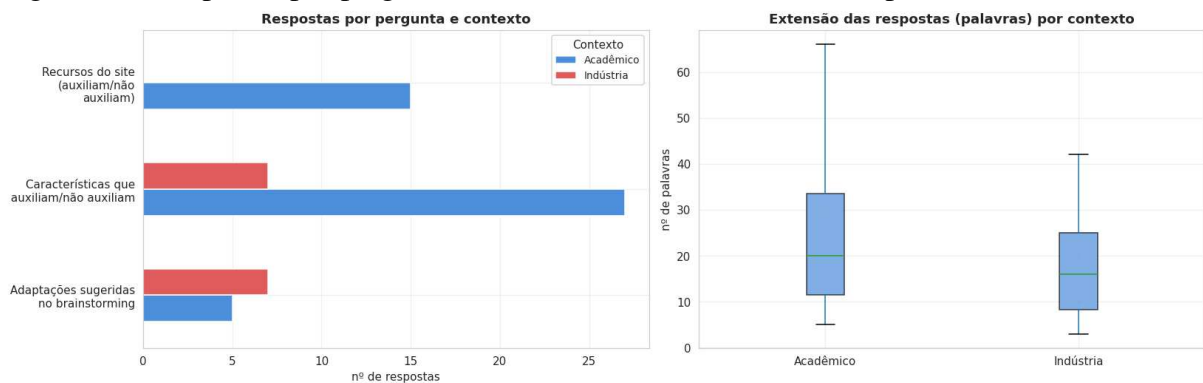
O corpus consolidado reúne 61 respostas abertas, distribuídas em três blocos: 34 respostas sobre características que auxiliam ou não a elicitação, 15 sobre os recursos do site de apoio e 12 sobre adaptações sugeridas para as sessões de *brainstorming*. Quanto ao contexto, 47 respostas provêm do ambiente acadêmico e 14 do industrial (Tabela 16). A Figura 23 apresenta a distribuição das respostas e a extensão dos relatos por contexto. As respostas do bloco de características são as mais extensas, com média de 25,3 palavras, e os relatos acadêmicos tendem a ser ligeiramente mais longos do que os industriais.

Tabela 16 – Caracterização do corpus qualitativo por bloco

Bloco	n	Caracteres (média)	Palavras (média)
Características que auxiliam/não auxiliam	34	154,3	25,3
Recursos do site (auxiliam/não auxiliam)	15	108,0	18,5
Adaptações sugeridas no brainstorming	12	131,6	20,8

Fonte: Elaborada pelo autor (2026).

Figura 23 – Respostas por pergunta e contexto e extensão dos relatos por contexto



Fonte: Elaborada pelo autor (2026).

6.10.2 Características do Método que Auxiliam a Elicitação

No bloco de características (Figuras 24 e 25), o tema dominante é, de forma destacada, o das cartas e guias como estrutura do método, com 16 menções, seguido por organização e padronização, curva de aprendizado e dificuldades, personas e *user stories* e colaboração em

equipe, todos com 7 menções, e por agilidade e produtividade (5), completude da elicitação (4) e contexto de uso e mecanismos (3).

A centralidade das cartas e guias confirma que os artefatos centrais do USARP, isto é, as cartas, o checklist e os guias de especificação, são percebidos como o principal motor da elicitação. Um participante da coorte acadêmica de 2025-1 sintetiza o papel de filtragem da etapa inicial e, ao mesmo tempo, a tensão de aprendizado:

A primeira etapa da USARP é ótima para “filtrar” quais cartas devemos avaliar se fazem sentido com aquele requisito e ajuda muito em relação à curva de aprendizado. Algumas cartas trazem uma linguagem um pouco mais técnica de UI/UX, que por não ser da área acaba que me perco um pouco, mas se no momento do *brainstorming* tiver uma pessoa do time de UI/UX fica tranquilo. (P96, Acadêmico 2025-1)

Esse relato é significativo do ponto de vista da evolução do método. A etapa de filtragem em formato de checklist resulta da evolução do método consolidada por Marques *et al.* (2023b), que respondeu às dificuldades observadas por Oliveira *et al.* (2020) no estudo de viabilidade, no qual os padrões originais (USEPs) foram considerados extensos e cansativos, motivando sua conversão em cartas. O tema da curva de aprendizado e da linguagem técnica de UI/UX, recorrente também em Fiori (2022), mostra que essa tensão não foi eliminada, mas mitigada pela presença de perfis distintos na sessão.

A colaboração em equipe e a completude da elicitação aparecem com força nos relatos industriais, em consonância com Marques *et al.* (2023a). Um participante da coorte industrial de 2021 destaca a cobertura sistemática dos requisitos:

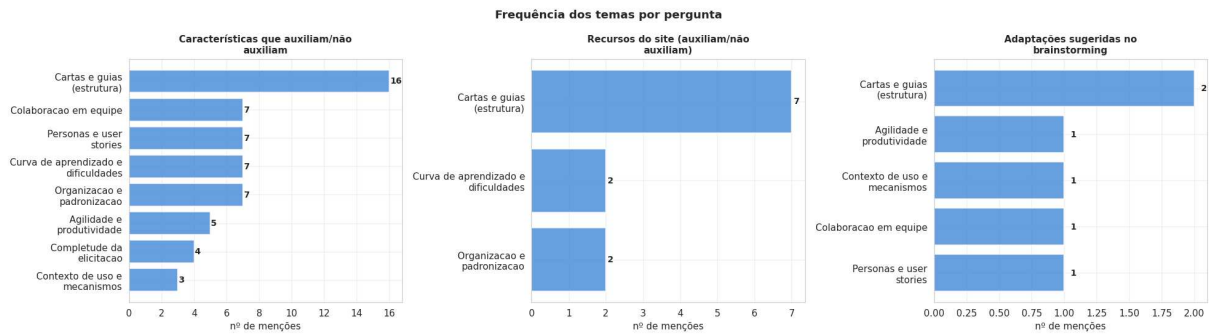
É interessante que todos os requisitos do sistema são lembrados de forma [organizada]; a gente tem uma ordem e um padrão a seguir para todas as *user stories*. (P84, Indústria 2021)

Outro relato da mesma coorte industrial associa o detalhamento das *user stories* ao debate entre profissionais de diferentes áreas, reforçando a dimensão colaborativa do método:

A descrição de como cada *feature* de uma *user story* se comportará de acordo com cada ação e status do sistema. Outra característica importante é o debate e as considerações entre profissionais de áreas diferentes sobre uma determinada *feature*. (P80, Indústria 2021)

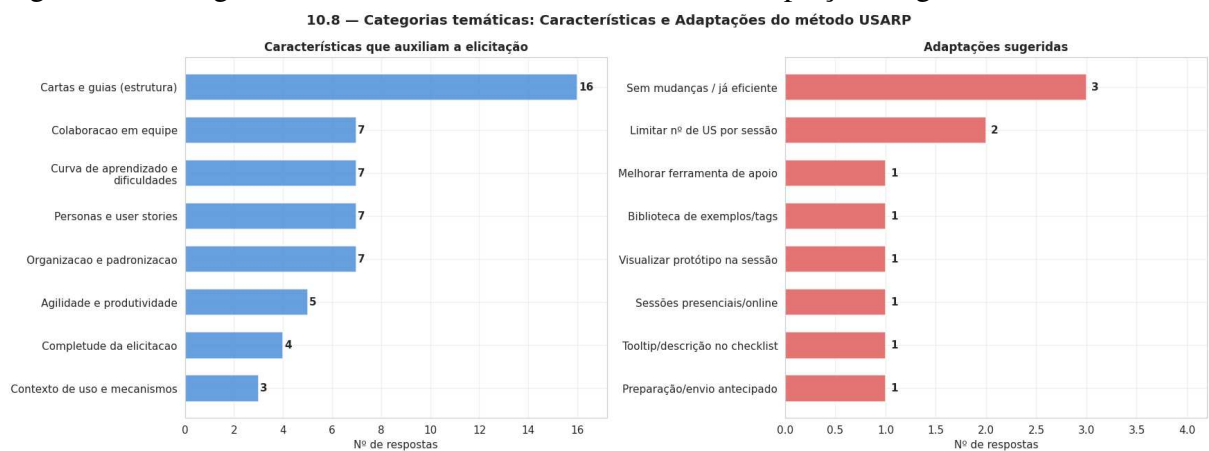
A agilidade e a produtividade, por fim, aparecem associadas ao fato de as cartas oferecerem um exemplo prévio que orienta a escrita do requisito, otimizando o tempo do analista de requisitos.

Figura 24 – Frequência dos temas por pergunta



Fonte: Elaborada pelo autor (2026).

Figura 25 – Categorias temáticas das características e das adaptações sugeridas



Fonte: Elaborada pelo autor (2026).

6.10.3 Adaptações Sugeridas

As adaptações sugeridas (Figura 25, à direita) são majoritariamente incrementais. A categoria mais frequente reúne participantes que não propõem mudanças, considerando o método já eficiente (3 menções), seguida pela sugestão de limitar o número de *user stories* por sessão (2 menções). As demais sugestões, com uma menção cada, concentram-se em ajustes de processo e de ferramenta: tooltips com a descrição dos mecanismos no checklist, preparação e envio prévio do material, visualização do protótipo durante a sessão, realização presencial ou online e ampliação da biblioteca de exemplos. Não há propostas de revisão estrutural do método, padrão coerente com Marques *et al.* (2023a), cujos participantes igualmente não apontaram necessidade de melhorias substantivas. Um relato da coorte acadêmica de 2025-1 ilustra a percepção de eficiência:

Creio que são bem eficientes, já que em uma sessão de cerca de 15 minutos conseguimos tirar todos os aspectos de UI/UX relacionados a um requisito.

(P96, Acadêmico 2025-1)

A sugestão de limitar o número de histórias por sessão e a de incluir tooltips no checklist são especialmente informativas, pois apontam para o mesmo ponto sensível identificado na análise quantitativa: o contexto e a compreensão dos mecanismos.

6.10.4 Recursos do Site de Apoio

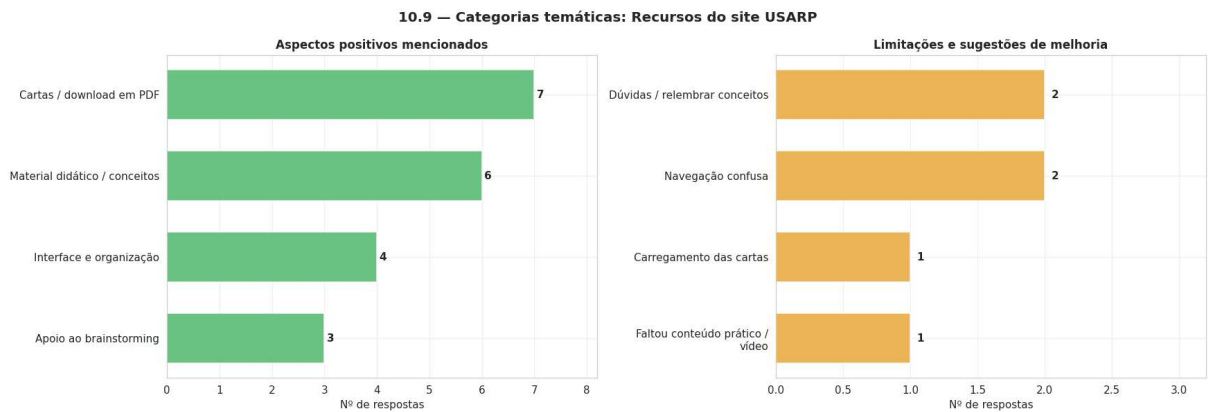
Dos 61 relatos, 15 referem-se ao site de apoio (Figura 26). Os aspectos positivos predominam: o material para download das cartas em PDF (7 menções), o material didático e a explicação dos conceitos (6), a interface e a organização (4) e o apoio ao *brainstorming* (3). As limitações são pontuais e dizem respeito à usabilidade da ferramenta: navegação por vezes confusa (2), dúvidas não totalmente esclarecidas (2), ausência de conteúdo prático em vídeo (1) e carregamento das cartas (1). Um relato da coorte acadêmica de 2021-2 sintetiza o valor didático do recurso e, ao mesmo tempo, uma limitação de navegação:

Gostei bastante do site; no começo fiquei muito confusa e perdida quanto ao método e às cartas, mas as explicações presentes no site ajudaram, e muito, a conseguir organizar as reuniões de maneira mais efetiva. (P25, Acadêmico 2021-2)

Outro participante da mesma coorte sugere conteúdo prático adicional:

Todo o conteúdo do site ajudou o grupo a saber como fazer a elicitação de requisitos usando o método USARP, mas eu achei que faltou informação prática; seria bom algum link para ver um vídeo com pessoas fazendo a elicitação com esse método, mostrando o objetivo do tipo de cada carta. (P6, Acadêmico 2021-2)

Figura 26 – Categorias temáticas dos recursos do site: aspectos positivos e limitações

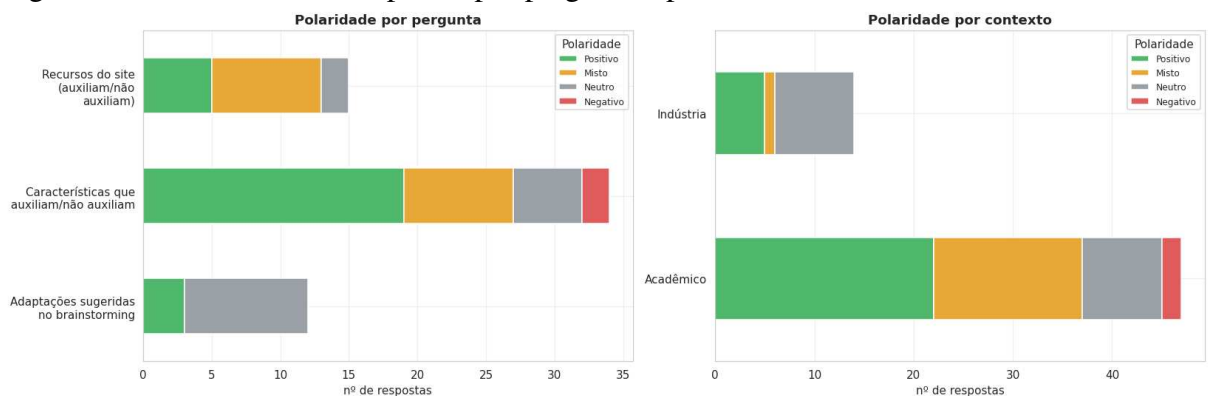


Fonte: Elaborada pelo autor (2026).

6.10.5 Polaridade e Tom das Respostas

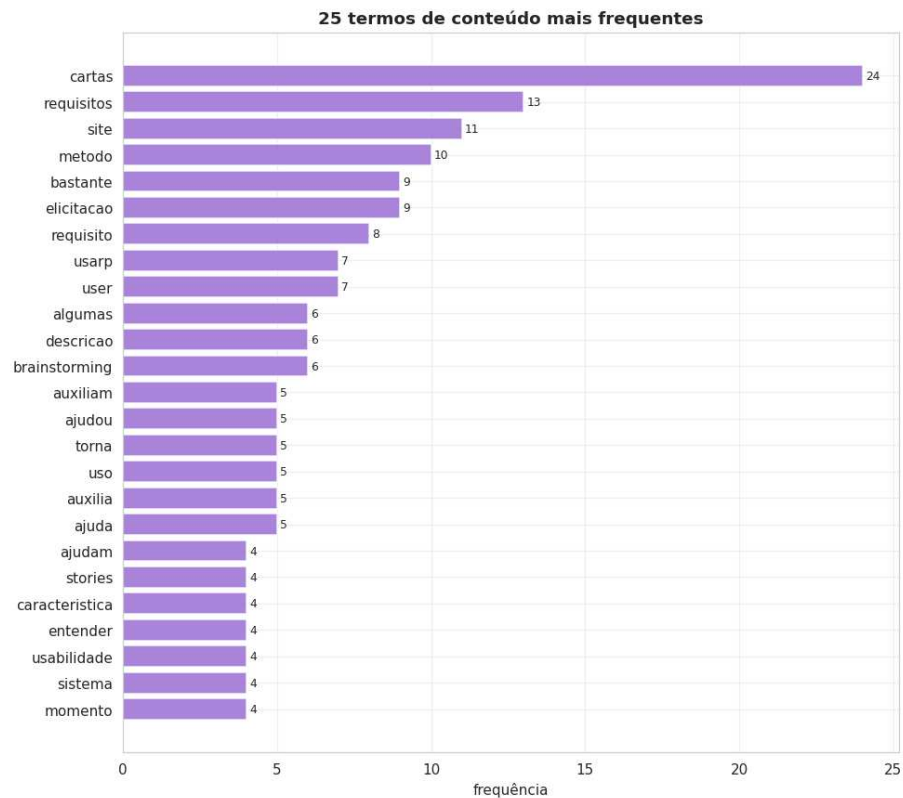
A classificação de polaridade (Figura 27) indica predomínio de respostas positivas (27), seguidas de respostas mistas (16) e neutras (16), com apenas 2 respostas negativas. A distribuição por bloco é coerente com a natureza de cada pergunta: o bloco de características concentra 79% de respostas positivas ou mistas e o do site, 87%, enquanto o bloco de adaptações é majoritariamente neutro, o que é esperado, dado que reúne sugestões e não avaliações. Os termos de conteúdo mais frequentes (Figura 28) reforçam o foco nos artefatos do método: “cartas” (24 ocorrências), “requisitos” (13), “site” (11) e “método” (10) lideram a contagem.

Figura 27 – Polaridade das respostas por pergunta e por contexto



Fonte: Elaborada pelo autor (2026).

Figura 28 – Termos de conteúdo mais frequentes nas respostas abertas



Fonte: Elaborada pelo autor (2026).

6.10.6 Síntese Qualitativa e Relação com os Estudos Primários

A Tabela 18 e a Figura 29 consolidam a leitura dos três blocos. Antes delas, a Tabela 17 apresenta, de forma sistematizada, a contagem de menções por tema, discriminada por bloco de pergunta e por contexto, complementando a leitura gráfica da Figura 24.

Tabela 17 – Menções por tema, segundo o bloco de pergunta e o contexto

Tema	Características	Site	Adaptações	Total	Acadêmico	Indústria
Cartas e guias (estrutura)	16	7	2	25	25	0
Organização e padronização	7	2	0	9	8	1
Curva de aprendizado e dificuldades	7	2	0	9	9	0
Colaboração em equipe	7	0	1	8	4	4
Personas e user stories	7	0	1	8	5	3
Agilidade e produtividade	5	0	1	6	4	2
Compleitude da elicitação	4	0	0	4	2	2
Contexto de uso e mecanismos	3	0	1	4	4	0
Total de atribuições	56	11	6	73	61	12

Fonte: Elaborada pelo autor (2026). Nota: como a codificação admite rótulos múltiplos, o total de atribuições excede o número de respostas (61).

O bloco de características é dominado pelo tema das cartas e guias; o de adaptações, pela ausência de mudanças propostas; e o do site, pela valorização positiva do material das

cartas, cuja principal limitação é a navegação por vezes confusa. Em conjunto, os achados qualitativos desenharam um quadro consistente com as evidências dos quatro estudos primários. A percepção de estrutura, organização e completude proporcionada pelas cartas reproduz os relatos de Marques *et al.* (2023a) e Fiori (2022); a dimensão colaborativa ecoa o uso do método em sessões com profissionais de diferentes papéis descrito por Marques *et al.* (2023a); e as tensões relativas à curva de aprendizado e ao volume e à linguagem das cartas remontam às dificuldades originalmente registradas por Oliveira *et al.* (2020) e reiteradas por Fiori (2022) e Simões (2022).

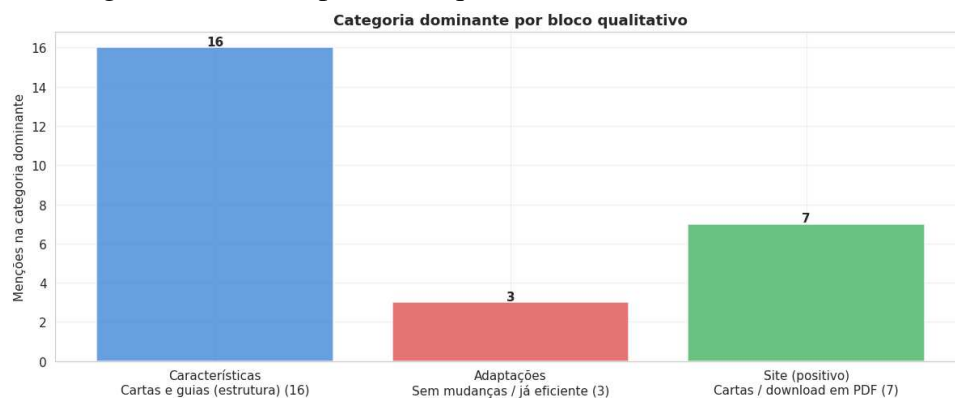
Um ponto merece destaque pela convergência entre os estudos: a subutilização das cartas de prototipação. Simões (2022) observam que os aspectos de prototipação foram pouco explorados, e Marques *et al.* (2023b), na análise documental dos trabalhos produzidos, identificaram que a carta de especificação de prototipação foi percebida como um obstáculo por parte dos participantes e que alguns mecanismos, como os de preferências e de área de objetos pessoais, foram subutilizados. O mesmo padrão emerge na presente síntese, tanto qualitativamente, pela menor presença do tema, quanto quantitativamente, já que o guia de prototipação apresenta a menor média entre os itens de cartas (5,52). Esse alinhamento entre os relatos discursivos, a análise de artefatos de Marques *et al.* (2023b) e as medidas numéricas fortalece a robustez do achado.

Tabela 18 – Síntese qualitativa por bloco

Bloco	n	Categoria dominante	% positivo/misto
Características	34	Cartas e guias (estrutura) (16)	79%
Adaptações	12	Sem mudanças / já eficiente (3)	25%
Site (positivo)	15	Cartas / download em PDF (7)	87%

Fonte: Elaborada pelo autor (2026).

Figura 29 – Categoria dominante por bloco qualitativo

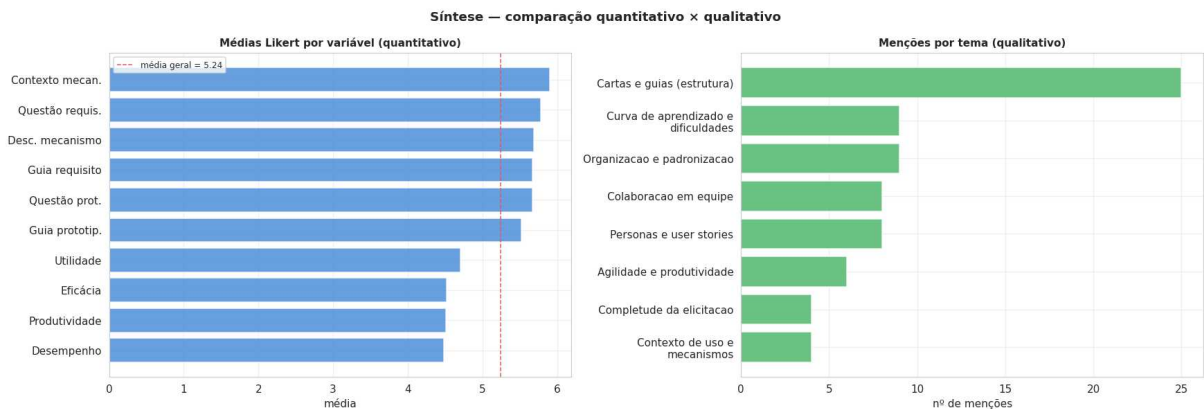


Fonte: Elaborada pelo autor (2026).

6.11 Discussão Integrada dos Resultados

Esta seção articula as evidências quantitativas e qualitativas e as relaciona aos objetivos do trabalho. A Figura 30 confronta, lado a lado, as médias dos itens do levantamento e a frequência dos temas qualitativos.

Figura 30 – Síntese integrada: médias do levantamento e frequência dos temas qualitativos



Fonte: Elaborada pelo autor (2026).

A leitura integrada revela forte coerência entre os dois conjuntos de evidência. Os itens com maiores médias no levantamento são os relativos ao conteúdo das cartas, com destaque para o contexto do mecanismo (5,90), a questão de requisitos (5,78) e a descrição do mecanismo (5,69), e o tema mais citado qualitativamente é justamente o das cartas e guias como estrutura do método, com 25 menções no total. A convergência se estende às tensões: as dificuldades apontadas nos relatos, isto é, a curva de aprendizado, a quantidade de cartas e a linguagem técnica de UI/UX, ajudam a explicar por que os itens de percepção operacional do TAM (desempenho, 4,48; produtividade, 4,51; eficácia, 4,52), embora positivos, recebem médias inferiores às dos itens de conteúdo das cartas. Em outras palavras, o conteúdo das cartas é altamente valorizado, ao passo que a experiência operacional de aplicá-las ainda esbarra na curva de aprendizado.

Essa convergência também esclarece o único contraste estatístico robusto identificado na análise quantitativa. O item de contexto do mecanismo de usabilidade, no qual a coorte industrial de 2022 diverge das turmas acadêmicas, corresponde, no plano qualitativo, ao tema de contexto de uso e mecanismos e à própria sugestão de incluir, no checklist, uma descrição de cada mecanismo. Trata-se, portanto, da dimensão mais sensível à maturidade de uso e ao grau de familiaridade com o método, o que é compatível com a sua trajetória de evolução, das USEPs densas de Oliveira *et al.* (2020) para as cartas e a etapa de filtragem em checklist propostas por

Marques *et al.* (2023b). Cabe notar que a dificuldade de seleção e de compreensão das cartas é justamente a motivação central do *checklist* e do quadro propostos por Marques *et al.* (2023b), o que reforça a leitura de que o item mais sensível da síntese coincide com o ponto que a própria evolução do método buscou endereçar. A leitura conjunta, na qual o conteúdo qualitativo ajuda a explicar os padrões numéricos, reforça a validade dos achados pela convergência entre as evidências (Braun; Clarke, 2006).

À luz dos objetivos específicos, os resultados permitem as seguintes conclusões parciais. Quanto ao primeiro objetivo, a percepção de utilidade, produtividade e eficácia mostrou-se elevada e homogênea entre contextos, ainda que os itens operacionais recebam médias um pouco inferiores às do conteúdo das cartas. Quanto ao segundo objetivo, a caracterização do uso percebido dos mecanismos evidenciou as cartas e os guias como o principal motor da elicitación e revelou a subutilização das cartas de prototipação, padrão convergente entre os relatos e as médias. Quanto ao terceiro objetivo, observaram-se mais semelhanças do que diferenças entre os contextos acadêmico e industrial, com a única divergência robusta restrita ao item de contexto do mecanismo. Quanto ao quarto objetivo, a análise de componentes principais explicitou as estruturas latentes, com uma dimensão de percepção positiva geral e contrastes secundários entre os blocos de cartas e os itens de percepção operacional. Por fim, as implicações práticas e as limitações, objeto do quinto objetivo, são aprofundadas no capítulo de considerações finais.

7 CONSIDERAÇÕES FINAIS

Este trabalho investigou empiricamente a aplicação do método USARP na elicitación de requisitos de usabilidade, por meio de uma síntese empírica que integrou evidências de quatro estudos primários conduzidos em contextos acadêmicos e industriais. A motivação partiu de uma lacuna concreta: embora estudos isolados indicassem benefícios do método, não havia uma síntese que avaliasse sua aplicação em diferentes contextos e identificasse padrões recorrentes de percepção, uso e adoção. Para preencher essa lacuna, consolidaram-se, no nível do respondente, os dados de levantamento dos estudos que adotaram o instrumento comum, e essa base foi submetida a uma reanálise quantitativa, com estatística descritiva, redução de dimensionalidade e testes não-paramétricos, e qualitativa, com análise temática, articulada com as evidências reportadas por Oliveira *et al.* (2020), Marques *et al.* (2023a), Fiori (2022) e Simões (2022).

No plano quantitativo, a percepção sobre o USARP mostrou-se elevada e marcadamente homogênea entre contextos e coortes, com forte efeito de teto em todos os dez itens analisados. A redução de dimensionalidade revelou uma primeira dimensão de percepção positiva geral, responsável por cerca de um terço da variância, acompanhada de contrastes secundários entre os blocos de cartas (requisitos, mecanismos e prototipação) e os itens de percepção operacional. Nem a segmentação não supervisionada, que agrupou os respondentes por nível geral de avaliação e não por contexto, nem as comparações inferenciais, que não acusaram diferenças robustas após a correção de Holm, sustentaram uma distinção marcante entre os ambientes acadêmico e industrial. A única diferença que resistiu ao controle estatístico ocorreu no item de contexto do mecanismo de usabilidade, entre a coorte industrial de 2022 e duas turmas acadêmicas.

No plano qualitativo, a análise temática das respostas abertas confirmou as cartas, o checklist e os guias de especificação como o principal motor da elicitación, seguidos pela organização e pela padronização do processo, pela colaboração entre profissionais de diferentes papéis e pelo uso conjunto de personas e *user stories*. As tensões concentraram-se na curva de aprendizado, no volume de cartas e na linguagem técnica de UI/UX, e as adaptações sugeridas foram majoritariamente incrementais, sem propostas de revisão estrutural do método. Um achado convergente entre os estudos e entre as duas naturezas de evidência foi a subutilização das cartas de prototipação, menos citadas nos relatos e com a menor média entre os itens de cartas.

Esses resultados atendem aos objetivos específicos propostos: sintetizaram-se as evidências de percepção de utilidade, produtividade e eficácia (primeiro objetivo); caracterizou-se

o uso percebido dos mecanismos, com destaque para a centralidade das cartas e a subutilização da prototipação (segundo objetivo); evidenciaram-se mais semelhanças do que diferenças entre contextos (terceiro objetivo); extraíram-se as estruturas latentes associadas à aplicação do método por meio da Análise de Componentes Principais (quarto objetivo); e discutiram-se as implicações práticas e as limitações (quinto objetivo). Atende-se, assim, ao objetivo geral de investigar empiricamente a aplicação do USARP a partir da síntese de evidências de contextos acadêmicos e industriais.

Como contribuições, o trabalho oferece uma síntese empírica crítica do USARP, que reúne, de forma comparável e no nível do respondente, evidências antes dispersas em estudos individuais; consolida e documenta o instrumento de avaliação derivado do *Technology Acceptance Model* e da utilidade percebida das cartas; disponibiliza um *pipeline* de análise reproduzível, com critérios e diagnósticos explícitos; e fornece subsídios práticos para a evolução do método, ao identificar tanto seus pontos fortes quanto os elementos que ainda impõem maior carga de uso.

O estudo apresenta limitações que devem ser consideradas na interpretação de seus resultados. A amostra é de tamanho reduzido, e a subamostra industrial, em particular, é pequena, com onze respondentes, o que limita o poder estatístico das comparações entre contextos; nesse sentido, a ausência de diferenças significativas indica ausência de evidência de diferença, e não equivalência comprovada. O forte efeito de teto e o valor modesto da medida de Kaiser-Meyer-Olkin recomendaram o uso exploratório, e não confirmatório, da Análise de Componentes Principais. Além disso, por se tratar de uma síntese de evidências baseada em dados secundários, coletados por instrumentos não idênticos entre todos os estudos e sem grupo de comparação controlado, não se formulam afirmações causais, e a generalização permanece circunscrita a cenários semelhantes aos estudados. Essas e outras ameaças à validade foram detalhadas no Capítulo 5.

Como trabalhos futuros, recomenda-se a condução de estudos observacionais sobre o uso efetivo do método, capazes de complementar as medidas de percepção com evidências de comportamento, conforme já sugerido por Marques *et al.* (2023a). A ampliação da base industrial, com coortes maiores, permitiria comparações entre contextos com maior poder estatístico. No plano do próprio método, os achados apontam direções concretas de refinamento: reduzir a carga associada ao volume de cartas, por exemplo aprofundando a etapa de filtragem assistida e as descrições contextuais, na forma de tooltips, no *checklist* proposto por Marques *et al.* (2023b);

revisitar as cartas de prototipação, consistentemente subutilizadas; e investigar em que medida o método contribui para a prevenção de problemas de usabilidade e para a redução de retrabalho ao longo do ciclo de desenvolvimento. Estudos longitudinais poderiam, ainda, examinar como a maturidade de uso afeta a percepção dos mecanismos, dimensão em que se observou a maior variação entre as coortes.

Em síntese, as evidências reunidas indicam que o USARP é percebido de maneira consistentemente positiva e similar em contextos acadêmicos e industriais, tendo nas cartas e nos guias o seu principal valor percebido, e na curva de aprendizado e no volume de artefatos os seus principais desafios. Ao consolidar e interpretar essas evidências de forma integrada, espera-se que este trabalho contribua para a adoção informada do método e para a continuidade de sua evolução, em benefício da incorporação sistemática da usabilidade desde as fases iniciais do desenvolvimento de software.

REFERÊNCIAS

- ALMAZROI, A. A. A systematic mapping study of software usability studies. **International Journal of Advanced Computer Science and Applications (IJACSA)**, v. 12, n. 9, p. 228–234, 2021. Disponível em: <https://dx.doi.org/10.14569/IJACSA.2021.0120927>.
- ARDITO, C.; COSTABILE, M. F.; ANGELI, A. D.; LANZILOTTI, R.; MATERA, M.; PICCINNO, A.; DESOLDA, G. Usability engineering: Theory and practice. **International Journal of Human-Computer Studies**, v. 72, n. 6, p. 526–548, 2014.
- BARTLETT, M. S. Tests of significance in factor analysis. **British Journal of Statistical Psychology**, v. 3, n. 2, p. 77–85, 1950.
- BRAUN, V.; CLARKE, V. Using thematic analysis in psychology. **Qualitative Research in Psychology**, v. 3, n. 2, p. 77–101, 2006.
- CARVAJAL, L.; MORENO, A. M.; FRANCH, X.; QUER, C. A systematic mapping study on usability requirements. **Journal of Systems and Software**, v. 86, n. 7, p. 1788–1803, 2013.
- CATTELL, R. B. The scree test for the number of factors. **Multivariate Behavioral Research**, v. 1, n. 2, p. 245–276, 1966.
- CHOMA, J.; ZAINA, L. A.; BERALDO, D. Userx story: Incorporating ux aspects into user stories elaboration. In: **Proceedings of the 18th International Conference on Human-Computer Interaction**. [S. l.]: Springer, 2016. p. 131–140.
- CRUZES, D. S.; DYBÅ, T. Synthesizing evidence in software engineering research. In: **Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '10)**. New York, NY, USA: ACM, 2010. p. 335–338. Disponível em: <https://doi.org/10.1145/1852786.1852788>.
- DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-1, n. 2, p. 224–227, 1979.
- DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, v. 13, n. 3, p. 319–340, 1989.
- FERREIRA, B.; BARBOSA, S. D. J.; CONTE, T. Creating personas focused on representing potential requirements to support the design of applications. In: **Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC)**. [S. l.]: ACM, 2018. p. 1–9.
- FIELD, A. **Discovering Statistics Using IBM SPSS Statistics**. 5. ed. London: SAGE, 2018.
- FIORI, M. V. S. **Uso da técnica USARP (Usability Requirements with Personas and User Stories) para elicitação de requisitos de usabilidade e prototipação: um estudo de caso no contexto acadêmico**. Russas, CE: [S. n.], 2022. Trabalho de Conclusão de Curso (Graduação em Engenharia de Software) – Universidade Federal do Ceará, Campus de Russas. Disponível em: <http://repositorio.ufc.br/handle/riufc/70283>.
- GONÇALVES, D.; ROCHA, A. R. A process to support the development of intelligent user interfaces. **Journal of Systems and Software**, v. 152, p. 21–38, 2019.

- HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate Data Analysis**. 8. ed. Boston: Cengage Learning, 2019.
- HARRIS, C. R.; MILLMAN, K. J. *et al.* Array programming with NumPy. **Nature**, v. 585, p. 357–362, 2020.
- HOLM, S. A simple sequentially rejective multiple test procedure. **Scandinavian Journal of Statistics**, v. 6, n. 2, p. 65–70, 1979.
- HUNTER, J. D. Matplotlib: a 2D graphics environment. **Computing in Science & Engineering**, v. 9, n. 3, p. 90–95, 2007.
- ISO. **ISO/IEC 25010: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE)**. Geneva: [S. n.], 2011. International Standard.
- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010.
- JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. New York: Springer, 2002.
- JURISTO, N.; MORENO, A. M.; SÁNCHEZ-SEGURA, M.-I. Guidelines for eliciting usability functionalities. **IEEE Transactions on Software Engineering**, v. 33, n. 11, p. 744–758, 2007.
- KAISER, H. F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, v. 20, n. 1, p. 141–151, 1960.
- KAISER, H. F. An index of factorial simplicity. **Psychometrika**, v. 39, n. 1, p. 31–36, 1974.
- KITCHENHAM, B.; BRERETON, P.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering – a systematic literature review. **Information and Software Technology**, v. 51, n. 1, p. 7–15, 2009.
- KITCHENHAM, B.; CHARTERS, S. **Guidelines for Performing Systematic Literature Reviews in Software Engineering**. Durham, UK, 2007. Version 2.3.
- KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583–621, 1952.
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, 1967. p. 281–297.
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. **Annals of Mathematical Statistics**, v. 18, n. 1, p. 50–60, 1947.
- MARQUES, A. B.; COSTA, A.; SANTOS, I. D. S.; ANDRADE, R. Enriching user stories with usability features in a remote agile project: a case study. In: . [S. l.: s. n.], 2023. p. 1–10.
- MARQUES, A. B.; FIORI, M. V.; FERREIRA, B. Evolving the usarp method to support usability requirements elicitation and specification. In: **Anais do XXVI Workshop em Engenharia de Requisitos (WER)**. Porto Alegre: SBC, 2023.
- MCKINNEY, W. Data structures for statistical computing in Python. In: **Proceedings of the 9th Python in Science Conference (SciPy)**. [S. l.: s. n.], 2010. p. 56–61.

MORENO, A. M.; YAGÜE, A. Agile user stories enriched with usability. In: **Proceedings of the 13th International Conference on Agile Software Development (XP)**. [S. l.]: Springer, 2012. p. 168–176.

NIELSEN, J. **Usability Engineering**. San Francisco: Morgan Kaufmann, 1993.

OLIVEIRA, G. F. d.; FERREIRA, B.; MARQUES, A. B. Usarp method: eliciting and describing usability requirements with personas and user stories. In: . [S. l.: s. n.], 2020. p. 437–446.

PEDREGOSA, F. *et al.* Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PREECE, J.; ROGERS, Y.; SHARP, H. **Interaction Design: Beyond Human-Computer Interaction**. 4. ed. Chichester: Wiley, 2015.

RODRIGUES, C. L.; ZIEGELMANN, P. K. Metanálise: conceitos e aplicações. **Revista HCPA**, v. 30, n. 1, p. 63–70, 2010.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987.

RUNESON, P.; HÖST, M.; RAINER, A.; REGNELL, B. **Case Study Research in Software Engineering: Guidelines and Examples**. Hoboken: Wiley, 2012.

SALDAÑA, J. **The Coding Manual for Qualitative Researchers**. 2. ed. London: SAGE, 2013.

SARAIVA, L. S. **Identificação de candidatos a padrões de requisitos de usabilidade emergentes da adoção do método USARP**. Dissertação (Trabalho de Conclusão de Curso (Graduação em Engenharia de Software)) – Universidade Federal do Ceará, Campus de Russas, Russas, 2023.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591–611, 1965.

SILVA, E. B.; ANDRADE, M. E.; LIMA, A. M.; MARQUES, A. B. Usarp tool: uma ferramenta de apoio à elicitação e especificação de requisitos de usabilidade. In: **Anais do XXVII Workshop em Engenharia de Requisitos (WER)**. Porto Alegre: SBC, 2024.

SIMÕES, J. d. S. **Utilizando a Técnica USARP (USAbility Requirements with Personas and user stories) com equipes trabalhando remotamente: um estudo de caso na indústria**. Russas, CE: [S. n.], 2022. Trabalho de Conclusão de Curso (Graduação em Engenharia de Software) – Universidade Federal do Ceará, Campus de Russas. Disponível em: <http://repositorio.ufc.br/handle/riufc/70294>.

SOMMERVILLE, I. **Software Engineering**. 9. ed. Boston: Addison-Wesley, 2011.

TEIXEIRA, F.; SANTOS, P. S. M.; TRAVASSOS, G. H. A requirements engineering process based on prototyping for health systems. In: **Proceedings of the Ibero-American Conference on Software Engineering (CIbSE)**. [S. l.: s. n.], 2014. p. 1–14.

VENKATESH, V.; BALA, H. Technology acceptance model 3 and a research agenda on interventions. **Decision Sciences**, v. 39, n. 2, p. 273–315, 2008.

- VIRTANEN, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, p. 261–272, 2020.
- WIERINGA, R. Design science as nested problem solving. In: **Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST)**. New York: ACM, 2009. p. 1–12.
- WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. **Experimentation in Software Engineering**. Berlin: Springer, 2012.