



UNIVERSIDADE FEDERAL DO MARANHÃO  
Programa de Pós-Graduação em Ciência da Computação

Italo Fernandes Serra da Silva

**Desenvolvendo um processo de Teste A/B para  
avaliação de Usabilidade e Experiência do  
Usuário**

São Luís - MA

2023

Italo Fernandes Serra da Silva

# **Desenvolvendo um processo de Teste A/B para avaliação de Usabilidade e Experiência do Usuário**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Maranhão

Orientador: Prof. Dr. Luis Jorge Enrique Rivero Cabrejos

São Luís - MA

2023

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Fernandes Serra da Silva, Italo.

Desenvolvendo um processo de Teste A/B para avaliação de Usabilidade e Experiência do Usuário / Italo Fernandes Serra da Silva. - 2023.

141 p.

Coorientador(a): Anselmo Cardoso de Paiva.

Orientador(a): Luis Jorge Enrique Rivero.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luis, 2023.

1. Comparação de software. 2. Experiência do usuário.  
3. Refinamento. 4. Teste A/B. 5. Usabilidade. I.  
Cardoso de Paiva, Anselmo. II. Jorge Enrique Rivero,  
Luis. III. Título.

Italo Fernandes Serra da Silva

## **Desenvolvendo um processo de Teste A/B para avaliação de Usabilidade e Experiência do Usuário**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Avaliação da dissertação em 10 de Abril de 2023, em São Luís - MA:

---

**Prof. Dr. Luis Jorge Enrique Rivero Cabrejos**  
Orientador  
Universidade Federal do Maranhão

---

**Prof. Dr. Anselmo Cardoso de Paiva**  
Coorientador  
Universidade Federal do Maranhão

---

**Prof. Dr. Davi Viana dos Santos**  
Examinador Interno  
Universidade Federal do Maranhão

---

**Prof. Dr. Raimundo da Silva Barreto**  
Examinador Externo  
Universidade Federal do Amazonas

São Luís - MA  
2023

*Dedico esta dissertação à minha esposa Maurícia Paz Aguiar, pelo apoio incondicional e constante incentivo e aos meus filhos Alexandre Augusto e Aquiles Fernandes pelo carinho que me foi dado por toda esta jornada.*

# Agradecimentos

A minha esposa, Maurícia Paz Aguiar, por me apoiar durante a minha jornada nesta pesquisa, sempre me incentivando a prosseguir com o trabalho, mesmo nos momentos mais difíceis. Obrigado por fazerem parte deste momento.

Aos integrantes do Núcleo de Computação Aplicada – NCA por todo o apoio. Vocês me deram memórias tão boas ao longo desses anos, que é difícil escolher apenas uma. Obrigado por estarem lá, quando preciso, para discutir as pesquisas. Certamente me tornei um pesquisador e uma pessoa melhor devido ao vosso apoio. Espero que nossas amizades persistam por muitos anos. Obrigado a todos.

Aos alunos da Universidade Federal do Maranhão (UFMA) José Florêncio, Rui Serra, Anderson Lopes e Gabriel Silva por todo apoio dado para o fechamento dos estudos de caso. Obrigado pelas discussões e pelos esforços empenhados para apoiar este trabalho. Espero que nossa parceria possa continuar. Obrigado!

A Universidade Federal do Maranhão e ao Programa de Pós Graduação em Ciência da Computação pelo apoio durante minha pesquisa.

Ao meu coorientador Anselmo Paiva, pelo constante apoio, pelos excelentes ensinamentos que levo para a minha vida, pelos momentos gastos comigo. Muito Obrigado por tudo!

Aos membros da banca, Davi Viana dos Santos e Raimundo da Silva Barreto, que dedicaram seu tempo para realizar a leitura da dissertação e por estarem presentes na minha defesa.

Por fim, tenho um agradecimento especial ao meu orientador, Luis Rivero. Fico muito feliz por chegar onde cheguei e como penso que só foi possível por sua causa. Aprendi muito com você, como: ser organizado, responsável, disciplinado e muito mais. Agradeço por todo o apoio. Tudo o que posso dizer é o quanto estou grato por todas as reuniões e direcionamentos realizados e por toda movimentação e esforço gerado para chegarmos até aqui. Muito obrigado!

*"Se você não pode medir, você não pode gerenciar."*

(Peter Drucker)

# Resumo

Este trabalho propõe um processo de avaliação de software utilizando o teste A/B, usabilidade e experiência do usuário. Tradicionalmente, a decisão de adquirir um software é baseada em funcionalidades, custo e tempo de implantação, sem considerar a opinião dos usuários que o utilizarão no dia a dia. Este estudo buscou desenvolver um processo genérico para avaliar dois softwares, utilizando critérios de usabilidade e experiência do usuário, a fim de auxiliar na tomada de decisão de compra. Foi realizada uma revisão da literatura, identificando seis artigos relevantes sobre testes de comparação entre softwares em termos de usabilidade e experiência do usuário. Em seguida, foram identificadas etapas que poderiam ser substituídas ou adicionadas ao processo, tornando-o flexível para adaptações. Com base nesse estudo, foi criada a primeira versão de uma nova técnica de avaliação de software chamada ABTUX, um processo que utiliza o teste A/B, usabilidade e experiência do usuário. O processo passou por refinamentos e adaptações com a orientação de engenheiros de software, resultando em uma segunda versão. Para validar o processo ABTUX, ele foi aplicado nas plataformas de ensino Udemy e Coursera, que são concorrentes e possuem funcionalidades similares. As respostas dos usuários foram coletadas, analisadas objetivamente e subjetivamente, e os resultados indicaram que a Coursera obteve a melhor avaliação em termos de usabilidade e experiência do usuário, de acordo com as premissas e objetivos do teste. Além disso, um questionário de avaliação foi aplicado aos avaliadores do teste ABTUX para identificar oportunidades de melhoria no processo e em seus artefatos, tanto qualitativamente quanto quantitativamente, para orientar estudos futuros.

**Palavras-chave:** Teste A/B, experiência do usuário, usabilidade, avaliação de software, comparação de software, tomada de decisão, refinamento.

# Abstract

This work presents a proposal for a software evaluation process using A/B testing, usability, and user experience. Traditionally, when acquiring software, factors such as functionality, cost, and implementation time are considered for the purchase decision, without taking into account the opinions of the users who will use the software in their daily activities. This study aimed to develop a generic process to evaluate two software options by employing usability and user experience criteria, thus supporting the decision-making process. A literature review was conducted to identify the state-of-the-art in comparing software in terms of usability and user experience, resulting in the identification of six relevant scientific articles. The study mapped their steps, techniques, methods, platforms, and resources used in testing. Subsequently, stages that could be replaced or added to the process were identified, making the new technique flexible for adaptation and customization according to specific needs. Building upon this study, a first version of a new software evaluation technique was created, named ABTUX, which stands for A/B Testing, Usability, and User Experience Evaluation Process. The process underwent adaptations and refinements with the guidance and support of software engineers, resulting in a second version. To validate the ABTUX process, it was applied to the Udemy and Coursera e-learning platforms, as they are competitors with similar functionalities. Responses were collected, and both objective and subjective answers were analyzed to assess the results. Ultimately, Coursera was identified as the software with the highest evaluation in terms of usability and user experience, based on the test's premises and objectives. Furthermore, a questionnaire was administered to assess the evaluators who conducted the ABTUX test, aiming to qualitatively and quantitatively evaluate the process and its artifacts, identifying opportunities for improvement and guiding future studies.

**Keywords:**A/B testing, User experience, Usability, Software Evaluation, Software Comparison, Decision-making, refinement.

# Lista de ilustrações

Figura 1 – Metodologia aplicada para o desenvolvimento do processo de avaliação de usabilidade e UX através do Teste A/B: ABTUX . . . . .	17
Figura 2 – Exemplo de um modelo de Teste A/B simples . . . . .	25
Figura 3 – Fluxo do Processo ABTUX v1.0 . . . . .	73
Figura 4 – Ilustração do DE-PARA do processo ABTUX (v1) e (v2 ACTIV07) . . . . .	98
Figura 5 – Ilustração do DE-PARA do processo ABTUX (v1) e (v2 - ACTIV08) . . . . .	98
Figura 6 – Macro Fluxo da versão 2.0 do processo ABTUX . . . . .	101
Figura 7 – UEQ: Gráfico com dos resultados da Coursera . . . . .	127
Figura 8 – UEQ: Gráfico com dos resultados da Udemy . . . . .	129
Figura 9 – Comparativo entre Coursera e Udemy . . . . .	130

# Lista de tabelas

Tabela 5 – Processo de Teste A/B: Atividades obrigatórias . . . . .	35
Tabela 6 – Processo de Teste A/B: Atividades obrigatórias . . . . .	41
Tabela 7 – Processo de Teste A/B: Atividades obrigatórias . . . . .	46
Tabela 8 – Processo de Teste A/B: Atividades obrigatórias . . . . .	49
Tabela 10 – Processo de Teste A/B: Atividades obrigatórias . . . . .	54
Tabela 11 – Processo de Teste A/B: Atividades obrigatórias . . . . .	59
Tabela 12 – Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais . . . . .	68
Tabela 13 – Proposta A/B testando processo por Usabilidade e UX avalia- ção - Atividades Extras Opcionais que podem ser realizadas se necessário por a desenvolvimento equipe . . . . .	69
Tabela 14 – Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais . . . . .	78
Tabela 17 – Questionamento dos alunos de Engenharia de software novatos sobre os artefatos do processo ABTUX v1.0 . . . . .	90
Tabela 18 – Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais . . . . .	95
Tabela 19 – Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais . . . . .	99
Tabela 27 – Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais . . . . .	114

# Lista de abreviaturas e siglas

UX	<i>User Experience</i>
SUS	<i>System usability scale</i>
QUEST	<i>Quaid-e-Awam University of Engineering, Science and Technology</i>
UEQ	<i>User Experience Questionarie</i>
UFMA	Universidade Federal do Maranhão
ABTUX	<i>Teste A/B para avaliar a Usabilidade e Experiência do Usuário (UX)</i>
ERP	<i>Enterprise Resource Planning</i>
OEC	<i>Overall Evaluation Criterion</i>
SEQ	<i>Single Ease Question</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	Contextualização	14
1.2	Problema e justificativa	15
1.3	Contribuições	16
1.4	Metodologia	16
1.5	Organização do Trabalho	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
2.1	Usabilidade	20
2.2	Experiência do Usuário	21
2.3	Tipos de Avaliações	22
2.3.1	Avaliações Comparativas	22
2.3.2	Avaliação de heurística	23
2.3.3	Avaliação de usabilidade	23
2.3.4	Teste A/B	24
2.3.4.1	Overall Evaluation Criterion (OEC)	24
2.3.4.2	Parâmetros	24
2.3.4.3	Variantes	25
2.4	Resumo do Capítulo	25
<b>3</b>	<b>REVISÃO DE LITERATURA</b>	<b>27</b>
3.1	Introdução ao Capítulo	27
3.2	Planejamento do Protocolo da Revisão	28
3.2.1	Questão de Pesquisa	28
3.2.2	Termos de Busca	28
3.2.3	Fontes de Busca	29
3.2.4	Critérios de Seleção	29
3.3	Execução da Revisão de Literatura	30
3.4	Resultados e Discussão	31
3.4.1	Modelo de extração de informações dos artigos	33
3.4.2	Processo de extração de informações dos artigos	34
3.4.2.1	Artigo 01: Melhoria da usabilidade por meio de testes A/B e refatoração	35
3.4.2.2	Artigo 02: Garantindo a qualidade da interface da Web por meio de testes de divisão baseados em usabilidade	41
3.4.2.3	Artigo 03: Tendências de design da Web e sua usabilidade pelo método de teste A B	46

3.4.2.4	Artigo 04: AB4Web: um testador A/B on-line para comparar alternativas de design de interface de usuário . . . . .	49
3.4.2.5	Artigo 05: Avaliação de usabilidade de sistemas ERP: uma comparação entre SAP S/4 Hana e Oracle Cloud . . . . .	51
3.4.2.6	Artigo 06: Avaliação da Experiência do Usuário por meio de Teste A/B Automático	54
<b>3.5</b>	<b>Mapeamento das etapas do processo . . . . .</b>	<b>57</b>
<b>3.6</b>	<b>Resumo do Capítulo . . . . .</b>	<b>71</b>
<b>4</b>	<b>PROPOSTA DO PROCESSO ABTUX V1.0 . . . . .</b>	<b>72</b>
<b>4.1</b>	<b>Processo ABTUX e seus Artefatos - Versão 1.0 . . . . .</b>	<b>72</b>
4.1.1	Fase de Iniciação . . . . .	74
4.1.2	Fase de Planejamento . . . . .	74
4.1.3	Fase de Execução . . . . .	76
4.1.4	Fase de Resultados e Conclusão . . . . .	77
4.1.5	Avaliação da Usabilidade . . . . .	83
4.1.5.1	Questões objetivas . . . . .	85
4.1.5.2	Questões subjetivas . . . . .	87
4.1.6	Questionário de Experiência do usuário . . . . .	88
<b>4.2</b>	<b>Resumo do Capítulo . . . . .</b>	<b>89</b>
<b>5</b>	<b>REFINAMENTO DO PROCESSO ABTUX . . . . .</b>	<b>90</b>
<b>5.1</b>	<b>Processo ABTUX e seus artefatos - Versão 2.0 . . . . .</b>	<b>90</b>
5.1.1	Revisão dos artefatos . . . . .	90
5.1.2	Ajustando o processo . . . . .	97
<b>5.2</b>	<b>Resumo do Capítulo . . . . .</b>	<b>106</b>
<b>6</b>	<b>APLICAÇÃO PRÁTICA DO PROCESSO ABTUX V2.0 . . . . .</b>	<b>107</b>
<b>6.1</b>	<b>Objetivo da Aplicação . . . . .</b>	<b>107</b>
6.1.1	Executando o processo ABTUX versão 2.0 . . . . .	107
<b>6.2</b>	<b>Análise dos Resultados . . . . .</b>	<b>114</b>
6.2.1	Perguntas objetivas de Usabilidade e UX: . . . . .	114
6.2.2	Análise dos resultados das Questões objetivas . . . . .	115
6.2.3	Análise das perguntas subjetivas . . . . .	118
<b>6.3</b>	<b>Questionário de Avaliação da Experiência do usuário . . . . .</b>	<b>123</b>
6.3.1	Análise do questionário UEQ na Coursera . . . . .	127
6.3.2	Análise do questionário UEQ na Udemy . . . . .	128
6.3.3	Análise comparativa da Udemy e Coursera, através da UEQ . . . . .	130
<b>6.4</b>	<b>Avaliação do processo ABTUX através do Feedback dos Avaliadores</b>	<b>131</b>
<b>7</b>	<b>CONCLUSÃO E TRABALHOS FUTUROS . . . . .</b>	<b>135</b>
<b>7.1</b>	<b>Epílogo . . . . .</b>	<b>135</b>

<b>7.2</b>	<b>Contribuições . . . . .</b>	<b>136</b>
<b>7.3</b>	<b>Perspectivas futuras . . . . .</b>	<b>137</b>
	<b>Referências . . . . .</b>	<b>138</b>

# 1 INTRODUÇÃO

Este capítulo apresenta a introdução a esta pesquisa. Além de uma contextualização do tema, são apresentados a motivação, os objetivos e a metodologia de pesquisa. Também, é apresentada a estrutura e organização deste documento no que se refere ao processo de avaliação da usabilidade e experiência do usuário de sistemas computacionais, através do teste A/B.

## 1.1 Contextualização

A usabilidade é um dos fatores de qualidade mais importantes para o projeto de sistemas interativos, pois impacta na forma como os utilizamos nas atividades da vida diária (Firmenich et al. 2019). Um sistema interativo com baixo grau de usabilidade pode tornar-se difícil de usar e tornar o usuário menos produtivo na realização de tarefas (Fernandez et al. 2011).

Além disso, o *User Experience (UX)* é um conceito para entender e estudar a qualidade no uso de produtos interativos. De acordo com AllAboutUX.org, existem pelo menos 27 definições sobre UX. Por exemplo, UX é definido pela *International Organization for Standardization (ISO)* (ISO and STANDARD 2010) como as percepções e respostas de uma pessoa que resultam do uso, ou da aplicação antecipada de um produto, sistema ou serviço.

De acordo com Borsci et al. (2013) a relação entre experiência do usuário e usabilidade ainda está em discussão; e alguns pesquisadores de UX [(Kujala et al. 2011), (Gegner and Runonen 2012), (Vermeeren et al. 2010)] afirmam que UX vai além da abordagem orientada a tarefas da usabilidade tradicional, trazendo aspectos como beleza, diversão, prazer e crescimento pessoal que satisfazem as necessidades humanas gerais, mas têm pouco valor instrumental. Consequentemente, medidas tradicionais de usabilidade relacionadas à facilidade de uso e utilidade de um aplicativo podem ser empregadas para avaliar a experiência do usuário. No entanto, avaliar os atributos tradicionais de usabilidade pode não ser suficiente, pois a UX considera atributos além do instrumental, como as emoções dos usuários antes, durante e após sua interação com um aplicativo de software (Vermeeren et al. 2010)].

Portanto, para transmitir uma experiência positiva, é necessário que os aplicativos disponibilizem funcionalidades e meios fáceis de acessá-la (Hassenzahl and Tractinsky 2006). Assim, os princípios de usabilidade devem ser respeitados ao desenvolver a interface do usuário de um aplicativo, considerando também

as emoções e reações dos usuários em relação ao aplicativo.

Considerando a necessidade de atender a diferentes padrões de qualidade, o teste de software é uma área importante para identificar defeitos de software (Garousi et al. 2020). Vários métodos de avaliação têm sido propostos para capturar os pensamentos, expectativas e emoções dos usuários em relação aos aplicativos de software (Bargas-Avila and Hornbæk 2011). Dentre as diferentes abordagens de teste de software, o teste A/B surgiu como uma alternativa para comparar diferentes tipos de software em um determinado contexto (Kohavi and Longbotham 2017).

O Teste A/B propõe dividir o universo de usuários em dois ou mais grupos com o objetivo de testar diferentes versões de uma ideia com cada grupo e comparar o efeito de cada versão (Kohavi and Longbotham 2017). As informações obtidas nesses testes podem ser usadas para definir o impacto de mudanças em diferentes versões de um software ou avaliar o impacto da escolha entre dois ou mais sistemas de informação simultâneos. Tradicionalmente, o teste A/B permite comparar os requisitos funcionais de um sistema de informação. No entanto, novas abordagens foram desenvolvidas para permitir avaliar usabilidade e experiência do usuário ao comparar dois ou mais sistemas de informação (Speicher et al. 2014).

Apesar da proposta de diferentes abordagens para testes A/B no contexto de Usabilidade e avaliação de UX, ainda existem algumas oportunidades de melhoria. Por exemplo, não identificamos um processo de teste A/B que pudesse ser aplicado para identificar problemas de usabilidade e experiência do usuário (Albert and Tullis 2013).

Considerando o exposto, este trabalho propõe um novo processo de teste A/B que permite avaliar atributos de usabilidade e experiência do usuário. Para isso, para cada um dos processos de teste A/B encontrados em uma revisão de literatura, coletamos dados sobre a categorização do processo, suas atividades, os papéis dentro do processo e os artefatos que foram empregados durante o planejamento, condução e relato do teste.

## 1.2 Problema e justificativa

Ao buscar a aquisição ou utilização de um software ou serviço computacional, usuários e empresas geralmente se concentram na pesquisa das funcionalidades oferecidas, bem como em considerações de custo e tempo de implantação. No entanto, muitas vezes negligenciam a importância fundamental da usabilidade e da experiência do usuário associada ao produto, o que pode resultar em insatisfação e dificuldades na curva de aprendizado durante sua utilização (Niranjanamurthy et al. 2014).

Essa pesquisa é motivada pela necessidade de reconhecer que os critérios de usabilidade e experiência do usuário como requisitos importantes para a decisão de

compra de um software em detrimento a outro. A questão de pesquisa abordada neste estudo se propõe a analisar e investigar de forma cientificamente embasada:

*"Como avaliar comparativamente dois ou mais softwares através da usabilidade e experiência do usuário, para o apoio a tomada de decisão de qual software adquirir?"*

### 1.3 Contribuições

Para este estudo, temos as seguintes contribuições foi desenvolvida para melhorar o processo de avaliação de Teste A/B com usabilidade e experiência do usuário:

- Desenvolvimento de um processo de avaliação comparativa de dois sistemas computacionais, com foco na usabilidade e experiência do usuário, utilizando teste A/B.
- Definição das atividades obrigatórias para ser realizado um teste A/B, seguindo o estado da arte encontrada na literatura.
- Mapeamento de atividades opcionais, que podem substituir atividades obrigatórias de um processo genérico para poder adequar o processo as necessidades dos usuários.
- Mapeamento de atividades adicionais que poderão ser incluídas em qualquer parte do teste para atender a cenários específicos, adicionando passos dentro do teste A/B com usabilidade e UX.
- Apresentação de evidências empíricas da aplicação da tecnologia proposta em diferentes tipos de sistemas computacionais, com o objetivo de recomendar a melhor solução entre as avaliadas.

### 1.4 Metodologia

A metodologia utilizada nesta pesquisa foi através do uso de estudos aplicados, que é uma técnica de pesquisa que visa aplicar teorias e conceitos existentes para solucionar problemas reais (Babbie, 2016). Diferentemente da pesquisa básica, que busca expandir o conhecimento em uma área, a pesquisa aplicada tem como objetivo criar soluções práticas para problemas específicos (Straub, 2012).

Esses estudos são frequentemente realizados em áreas como negócios, tecnologia, saúde e ciências sociais (Babbie, 2016). Uma das principais vantagens dos estudos aplicados é a sua aplicação prática dentro da indústria ou sociedade (Straub, 2012), pois os resultados obtidos por meio desses estudos podem ser aplicados para resolver problemas do cotidiano (Babbie, 2016).

Por outro lado, uma das desvantagens dos estudos aplicados é que eles podem ser limitados pela falta de generalização dos resultados para outras situações (Babbie, 2016). No entanto, essa limitação pode ser superada por meio da realização de estudos múltiplos em diferentes contextos (Straub, 2012).

Para realizar um estudo aplicado, é necessário seguir algumas etapas específicas para o desenvolvimento e refinamento do processo proposto para avaliação de sistemas computacionais. A figura 1 representa os passos que iremos percorrer para a montagem do novo processo de avaliação de software, conforme a seguir.

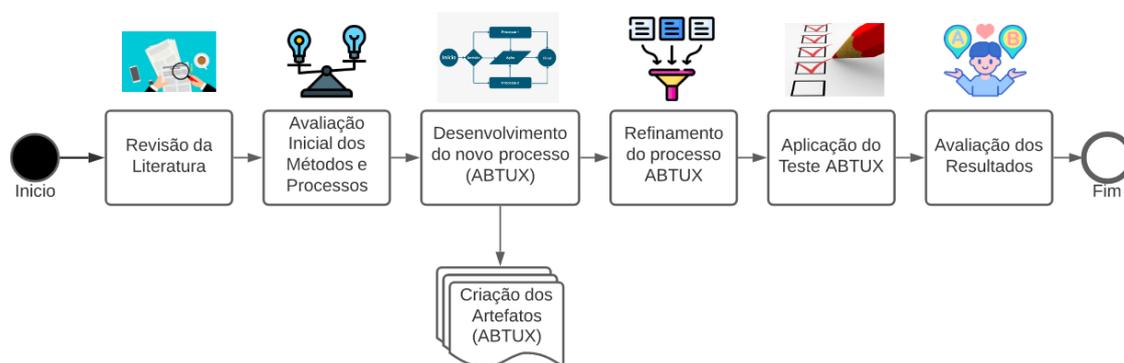


Figura 1 – Metodologia aplicada para o desenvolvimento do processo de avaliação de usabilidade e UX através do Teste A/B: ABTUX

**(A) Revisões da Literatura e Trabalhos Relacionados:** Nesta primeira etapa, realizou-se estudos secundários a fim de conhecer e identificar os conceitos, processos e instrumentos utilizados para a aplicação de um teste A/B padrão. Em seguida, foram identificados os Testes A/B que utilizam em suas etapas atributos de Usabilidade ou UX. Também foi pesquisado artigos e estudos que falam sobre os atributos de UX ou usabilidade genéricos (heurísticas) capazes de influenciar a usabilidade e a experiência do usuário através do uso do Testes A/B.

**(B) Avaliação Inicial dos Métodos e Processos Existentes:** Durante esta fase, foram realizados estudos iniciais para avaliar as propostas existentes para a avaliação de software com o objetivo de alcançar uma usabilidade e experiência do usuário de alta qualidade. Durante esta avaliação, selecionamos um subconjunto dos métodos identificados na etapa de Revisão da Literatura, a fim de identificar seus pontos fortes e reunir um conjunto de recursos a serem incorporados em nossa proposta.

**(C) Desenvolvimento do processo ABTUX:** Durante esta etapa, incorporamos as características identificadas tanto na Revisão da Literatura quanto nas Etapas de Análise do Processo e Atributos para propor um conjunto de tecnologias com o objetivo de avaliar e sugerir melhorias para o processo de teste de software. Por meio delas, desenvolvemos uma

tecnologia para avaliação sistemas computacionais, através de um processo de avaliação de software utilizando em conjunto o Teste A/B, Usabilidade e UX. Esta ação gerou a primeira versão do processo, denominado de ABTUX (v1.0).

**(D) Refinamento do Processo ABTUX:** O objetivo desta etapa é condensar os dados coletados no processo ABTUX para fazer ajustes e melhorias no processo desenvolvido e que atenda às necessidades atuais na área de pesquisa de usabilidade e UX, auxiliando os engenheiros de software no processo de avaliação de softwares, ao mesmo tempo em que sugere o que pode ser alterado ou adicionado para atender aos diversos cenários de teste desenhados para o novo processo, em que foi possível gerar a segunda versão do processo ABTUX (v2.0).

**(E) Execução do Teste ABTUX (v2.0):** Para validar do ponto de vista prático, a viabilidade do uso da proposta do processo, foi realizada uma prova de conceito, aplicando todo o processo desenvolvido em dois softwares semelhantes ou concorrentes para encontrar defeitos. Para efeito de comparação, serão utilizados as técnicas e os artefatos criados a fim de comparar os resultados e definir o sistema computacional vencedor.

**(F) Avaliação dos Resultados:** Ao final do Teste, realizamos a coleta dos resultados obtidos através dos questionários de usabilidade e UX, além da coleta de informações como o tempo de execução e outras informações observadas durante o teste para definir o software vencedor. Em paralelo, durante a avaliação dos resultados do processo ABTUX (v2.0) disponibilizamos um questionários para os responsáveis pela aplicação do teste (avaliadores) para podermos obter informações do processo desenvolvido, tais como: (a) nas dificuldades encontradas pelos engenheiros de software ao aplicar o ABTUX, visando alcançar a usabilidade e UX de alta qualidade; (b) requisitos sugeridos na revisão da literatura; e (c) resultados de nossa avaliação inicial do processo ABTUX. Por fim, realizamos consolidamos estes dados para obter um melhor desempenho no processo de forma que ele possa ser utilizado por usuários ou na indústria de software para trabalhos futuros e novas oportunidades de melhoria.

## 1.5 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 trata da fundamentação teórica das técnicas utilizadas. São abordados conceitos de usabilidade, experiência do usuário, testes comparativos e o teste A/B que fazem parte do objeto de pesquisa.
- O Capítulo 3. São apresentados os artigos pesquisados que envolvem de forma parcial ou total o uso de teste A/B, a experiência do usuário e usabilidade em sistemas computacionais para identificarmos se a pesquisa proposta já possui alguma aplicação

da técnica que estamos propondo, ou se ela é inédita. Caso não seja inédita, iremos propor melhorias nas técnicas existentes e disponibilizar todos os artefatos para a aplicação desta nova técnica aprimorada.

- O Capítulo 4 irá demonstrar o processo ABTUX, com suas atividades, artefatos e atributos de usabilidade e UX utilizados. Também será apresentado as etapas do processo.
- O Capítulo 5 irá apresentar a evolução e refinamento do processo ABTUX através de uma revisão por pares com engenheiros de software.
- No Capítulo 6 será realizada a aplicação prática do processo ABTUX, a consolidação de todos os dados extraídos do processo aplicado, avaliação e discussão dos resultados.
- O Capítulo 7 irá apresentar as considerações finais sobre os resultados e trabalhos futuros e os artigos científicos desenvolvidos.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os principais conceitos relacionados e trabalhos relevantes para fundamentar esta proposta de pesquisa. Nas próximas seções, são apresentados os conceitos de Usabilidade, Teste de usabilidade, Experiência do Usuário, Tipos de Testes (Comparativos, Avaliação por heurística e Teste de Usabilidade) e o Teste A/B. Compreender esses conceitos é crucial para evoluir as práticas atuais de um forma que se adéquem a nova tecnologia proposta no processo de avaliação de um sistema computacional.

### 2.1 Usabilidade

A Usabilidade ([Bargas-Avila and Hornbæk 2011](#)) pode ser aplicada através de uma avaliação, que consiste basicamente em observar ou entrevistar usuários de um produto ou software para identificar potenciais problemas e dificuldades no seu uso. Os testes buscam medir o grau de dificuldade durante a experiência do usuário pelo produto. Estas são heurísticas que têm sido utilizadas em avaliações de usabilidade tradicionais [([Borsci et al. 2013](#)), ([Fernandez et al. 2011](#))]:

1. Visibilidade do Status do Sistema: O sistema deve sempre manter os usuários informados sobre o que está acontecendo na tela em que ele se encontra, em tempo real, por meio de atualizações instantâneas, dentro de um prazo aceitável.
2. Compatibilidade entre o sistema e o mundo real: O sistema deve utilizar sinais, símbolos (como ícones) e linguagem em que o usuário utiliza no dia a dia.
3. Controle e liberdade para o usuário: No software é importante que o usuário tenha a opção de desfazer ou refazer algo que deseja. Em muitos casos a ação involuntária ou um clique em algum local indesejado ou por engano possa acontecer, então o software precisa entregar esta facilidade para o usuário para que ele possa sair daquela situação indesejada ou retorne ao passo anterior.
4. Consistência e Padronização: No software, estabelecer padrões de botões, navegação é muito importante para uma boa experiência de uso do produto. Para isso é preciso desenhar o software de forma que todo o produto esteja esteticamente organizado, evitando confusões de uso.
5. Prevenção de erros: Essa heurística trata as situações em que é realizado o tratamento de mensagens em caso de erros e direcionando o usuário a executar uma ação de confirmação ou aviso antes que o software avance para uma próxima ação.

6. Reconhecimento em vez de memorização: Devemos evitar a necessidade do usuário ter que gravar tudo que faz, então para isso, objetos importantes devem ficar visíveis, e de fácil acesso, para que o usuário reconheça sua função.
7. Eficiência e flexibilidade de uso: A interface precisa ser simples e de fácil acesso. Usuários sem experiência no produto precisam compreender e conseguir executar suas ações com a menor orientação possível e permitir que os usuários avançados realizem sem precisar repassar novamente pelas etapas básicas.
8. Estética e design minimalista: Temos que ter cuidado com a quantidade de informações que queremos passar para o usuário, e isso poderá cansá-lo a ponto de abandonar o uso da aplicação por achar confuso demais.
9. Ajude os usuários a reconhecerem, diagnosticarem e recuperarem-se de erros: Caso dê algo de errado, é importante mostrar ao usuário qual foi o erro e como se recuperar dele, logo, as mensagens de erro precisam ser claras e diretas, em linguagem simples e informar qual foi a ação que causou o erro.
10. Ajuda e documentação: O usuário poderá necessitar de apoio maior em algum momento e mesmo que a ajuda e a documentação seja pouco utilizada, elas devem estar disponíveis. Elas podem também ajudar o usuário resolver um problema por um guia assistido.

Estas heurísticas serão utilizadas em nosso estudo para ajudar a avaliar quais sistemas computacionais proporcionam uma melhor usabilidade para o usuário (Nielsen 2012).

## 2.2 Experiência do Usuário

A experiência do usuário (UX) aborda à(s) experiência(s) resultante(s) do encontro do usuário com sistemas computacionais (Vermeeren et al. 2010). UX inclui encontros com sistemas – não apenas de forma ativa durante o seu uso, mas também de forma mais passiva, como observar alguém usando um software. A UX é única para um indivíduo, pois ela é influenciada por uma bagagem de experiências e expectativas que cada um traz em si e fica entrelaçado em um contexto social e cultural. Enquanto os aspectos tradicionais de usabilidade se concentram no desempenho e na interação perfeita, os fatores experiência do usuário estão mais voltados com o afeto, a interpretação e o significado (Vermeeren et al. 2010). A usabilidade e a experiência do usuário estão portanto intimamente ligadas. Alguns pontos são muito importantes para garantir uma boa interação, e para isso os designers precisam se esforçar para refletir sobre o conjunto atributos de qualidade que precisam ser avaliados (Saffer 2006), sendo eles:

1. **Confiável:** É essencial que o usuário tenha confiança de forma simples e eficiente antes de usar um produto ou serviço, e essa decisão quanto a confiabilidade de um produto é por muitas vezes tomada segundos antes de interagir com o produto. É nesse tempo que criamos nossos primeiros julgamentos. Normalmente, quando temos confiança em algo ou alguém, somos muito mais propensos a nos envolver ainda mais com ele".
2. **Apropriado:** Os softwares projetados pelos designers precisam perceber o cenário em que o produto ou serviço irá funcionar, também considerando a parte sentimental do momento da interação.
3. **Inteligente:** Softwares precisam ser espertos de forma a impedir que usuários cometam erros primários ou façam cometam mais erros do que necessário, principalmente deixa-lo pensar muito possa executar uma ação;
4. **Esperto:** O software que traz uma sensação boa logo de início, de forma a quase prever a intenção do usuário. Então a ideia é de que os softwares atendam as necessidades dos usuários e cumpram facilmente suas tarefas.
5. **Agradável:** Softwares devem ser agradáveis para que o usuário queira novamente utiliza-lo.

A Experiência do usuário aborda as opções envolvidas na transmissão de informações de comportamento. Eles são sobre entender as pessoas, seus padrões de comportamento e pensamento. Ao construir esse pensamento na estrutura do sistema computacional, ajudamos a garantir uma experiência bem-sucedida para aqueles que o usam, ou seja, pensando nos itens que formam o processo de Experiência do Usuário, formam-se os padrões, o fluxo de informações e funcionalidades que são entregues aos usuários, focado em como o usuário vai de fato usar e completar tarefas ([Garrett 2011](#)).

## 2.3 Tipos de Avaliações

### 2.3.1 Avaliações Comparativas

Nos relatórios de estudos comparativos de usabilidade disponíveis na literatura, diferentes entidades são utilizadas como objetos de comparação. Alguns estudos comparam diferentes métodos de avaliação, comparam diferentes implementações do mesmo sistema ou protótipo e outros comparam diferentes grupos de usuários usando o mesmo sistema. Qualquer estudo que compare entidades deve seguir alguns princípios básicos para garantir que a comparação seja válida. Segundo ([Kruger et al. 2016](#)) em um estudo de literatura comparativa para uma UX ou avaliação de usabilidade produz as seguintes diretrizes:

- Entidades comparáveis devem ser incluídas no estudo; isso significa que as entidades devem servir a um propósito semelhante, operar em ambientes semelhantes e basear-se em experiências semelhantes.
- Os testes para as várias entidades devem ser conduzidos de forma semelhante. Isso implica usar o mesmas ferramentas, procedimentos e variáveis.
- Os resultados dos testes devem ser de natureza comparável.

### 2.3.2 Avaliação de heurística

Uma das formas de se validar uma interface é através de uma avaliação heurística, que foi desenvolvida por Nielsen. É uma avaliação feita por profissionais e não por usuários, que visa entender se os elementos da interface do usuário, como por exemplo se menus e estrutura de navegação, estão conforme aos princípios já testados de usabilidade, que são as heurísticas(Preece et al. 2015).

### 2.3.3 Avaliação de usabilidade

Segundo (Preece et al. 2015) avaliar é uma parte integrante do processo de design e que se concentra tanto na usabilidade projeto, onde analisa-se a facilidade de aprender e usar, quanto a experiência dos usuários que interagem com ele, entendendo por exemplo quão motivadora e divertida é a interação, com o objetivo de melhorar o design. Existem diversas formas de avaliação, mas basicamente os avaliadores vão coletar informações sobre experiências dos usuários ao interagirem com um protótipo ou um sistema computacional. Nessa avaliação é fundamental explicar ao usuário a importância do teste e também definir os aspectos que serão avaliados.

Conforme apresentado por (Kruger et al. 2016), em um teste de usabilidade mostra-se o sistema computacional ou uma parte específica do sistema para o usuário e pede-se a ele que primeiramente entenda o que é, então posteriormente o usuário tente executar as tarefas indicadas pelo responsável pelo teste. Lembrando também de pedir ao usuário que fale o que estiver pensando quanto à interface, para que seja possível entender a percepção que ele tem e assim identificar os erros e chances de melhoria para uma melhor compreensão ou execução de tarefas.

Um dos testes mencionados é o teste de compreensão. Neste teste é mostrado ao usuário o sistema computacional desejado para ver se ele compreende o propósito do sistema, como funciona, como ele está organizado. E há outra forma de avaliação, que é selecionar uma tarefa e pedir ao usuário que execute as tarefas.

O foco em nosso estudo é como uma avaliação comparativa de UX e usabilidade entre dois sistemas computacionais, com os mesmos requisitos funcionais pode beneficiar

a indústria ou qualquer outro tipo de segmento. Definimos um estudo comparativo de UX e usabilidade entre como o ato de examinar, o comportamento do usuário ao usar as interfaces tecnológicas de diferentes sistemas computacionais que tenha propósitos comparáveis e definir qual o vencedor com a melhor usabilidade e experiência do usuário (UX).

### 2.3.4 Teste A/B

Por meio do processo de teste A/B, os pesquisadores de experiência do usuário (UX) e usabilidade podem tomar decisões iterativas centradas no usuário e orientadas por dados para sistemas computacionais. O “teste A/B” e sua extensão “teste A/B/n” representam uma notação abreviada para um experimento controlado simples, no qual os usuários recebem aleatoriamente uma das duas ou mais variações de um produto ou serviço: controle (A), tratamento (B) e qualquer número de variações adicionais (n), onde as variações apresentam uma única variável de projeto isolada, ilustrado na figura 2.

Com um design de experimento adequado, a variação de melhor desempenho pode ser identificada de acordo com métricas predefinidas (Kohavi and Longbotham 2017). Dentro de uma estrutura diversificada de metodologias de pesquisa de UX e usabilidade, o teste A/B oferece uma técnica produtiva para responder a estas perguntas. Grande parte da pesquisa existente que explora as estruturas estatísticas associadas ao teste A/B pode ser encontrada na literatura da ciência da computação, com investigação específica de técnicas de amostragem, algoritmos de randomização, métodos de atribuição e coleta de dados brutos (Young 2014).

#### 2.3.4.1 Overall Evaluation Criterion (OEC)

O *overall evaluation criterion* (critério de avaliação) geral é uma medida quantitativa do objetivo do experimento. apresenta um exemplo e que o OEC pode ser a média de vezes que o usuário interage com o aplicativo (abre a tela inicial ou faz alguma ação), assim como aumentar este OEC faz com que os usuários os usuários utilizem mais o aplicativo, o que é um excelente resultado. O OEC deve poder ser mensurado em um curto período de tempo, ainda que se acredite que conduza objetivos de longo prazo (Scroferneker 2021) .

#### 2.3.4.2 Parâmetros

Também chamados de fatores ou variáveis, os parâmetros são basicamente as mudanças que acreditamos que alterarão o OEC. Em testes A/B simples, temos apenas um parâmetro com dois valores. Por exemplo, podemos testar os efeitos no tráfego de usuários ao colocar propagandas em uma página ou aplicativo. Neste caso, teríamos dois valores para o parâmetro propaganda: sim e não (Scroferneker 2021).

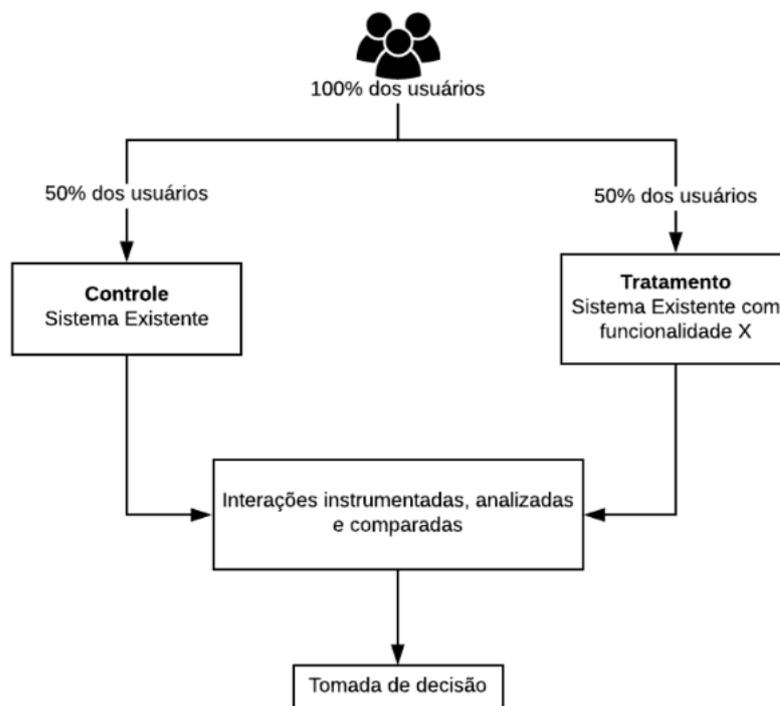


Figura 2 – Exemplo de um modelo de Teste A/B simples

#### 2.3.4.3 Variantes

Em um teste A/B simples, A e B são duas variantes, chamadas de controle e tratamento. A variante de controle é normalmente atribuída à versão já existente, em que utilizamos como base para o experimento, no entanto, nem sempre teremos uma versão anterior, podemos testar duas versões novas para descobrir qual delas obtém melhores resultados no OEC, já a variável de tratamento, representa a alteração que queremos testar no experimento (Scroferneker 2021).

Para melhor compreensão dos conceitos já citados, podemos considerar um ensaio para o teste de um adubo para plantas. Dividimos o grupo de plantas em duas populações: os que receberão o adubo e os que não tomarão. Sendo assim, avaliaremos quantas plantas tiveram o crescimento maior entre os grupo depois de um tempo (Magazine 2012).

Neste exemplo, nossa OEC é a quantidade de plantas que não tiveram crescimento no período dos testes, nosso parâmetro que podemos alterar é o tipo de adubo e as variantes possíveis para o parâmetro é com adubo ou sem adubo (Magazine 2012).

## 2.4 Resumo do Capítulo

Neste capítulo foram apresentados vários conceitos fundamentais para a realização desta pesquisa. Com essa fundamentação ficou evidente que a avaliação da usabilidade e experiência do usuário (UX) é parte importante no processo de avaliação de um sistema

computacional, pois utilizando estes atributos de qualidade, podemos realizar alterações em um produto, sem alterar a sua funcionalidade, sobre cada um dos aspectos apresentados e realizar uma comparação entre eles, para assim ter um resultado estatístico ou por feedback dos usuários se a mudança melhorou a interação do homem com o computador (HCI). Com estes resultados, é possível também realizar o mapeamento dos problemas de usabilidade e UX de sistemas computacionais. Para termos a confirmação que a união destes três atributos sobre a avaliação de um software é de fato utilizado na comunidade científica, realizamos uma revisão de literatura para obtermos esta resposta, conforme veremos no capítulo a seguir.

## 3 REVISÃO DE LITERATURA

Este capítulo apresenta e discute os resultados de um estudo secundário da literatura para identificar atributos de qualidade utilizáveis em avaliação de software através de um teste comparativo através da usabilidade e/ou experiência do usuário.

### 3.1 Introdução ao Capítulo

Nos últimos anos, o teste A/B tem se tornado uma abordagem cada vez mais popular e eficaz para avaliar a usabilidade e a experiência do usuário (UX) em diferentes contextos. Essa técnica permite comparar duas versões de um design ou de uma interface de usuário, conhecidas como versão A e versão B, a fim de identificar qual delas oferece uma melhor performance e satisfação para os usuários. Nesta revisão de literatura, examinaremos os principais estudos e trabalhos relacionados ao uso do teste A/B para avaliação de usabilidade e UX, destacando suas contribuições e insights relevantes (Krug 2014).

O teste A/B, também conhecido como teste dividido, envolve a criação de duas versões diferentes de uma interface, com apenas uma única diferença controlada entre elas. Um grupo de usuários é exposto à versão A, enquanto outro grupo é exposto à versão B. Com base nos dados coletados, é possível determinar qual versão apresenta melhores resultados em termos de usabilidade, taxa de conversão, satisfação do usuário e outros indicadores relevantes (Scroferneker 2021).

A usabilidade e a experiência do usuário são fatores cruciais para o sucesso de um produto ou serviço. A usabilidade refere-se à facilidade com que os usuários podem interagir com um sistema, enquanto a UX abrange aspectos emocionais e subjetivos, como a satisfação, o prazer e a eficácia percebida pelo usuário durante a interação (Nielsen 1994). O teste A/B é uma abordagem valiosa para avaliar e aprimorar essas dimensões, permitindo que os pesquisadores identifiquem as melhores práticas de design (Anderson 2016).

Diversas pesquisas têm explorado as aplicações do teste A/B na área de usabilidade e UX. Alguns estudos se concentraram na comparação de diferentes designs de interface para identificar qual versão proporciona uma experiência mais eficiente e agradável aos usuários. Outros investigaram a influência de fatores específicos, como cores, layout, disposição de elementos e chamadas para ação, utilizando o teste A/B para medir o impacto dessas variações nas métricas de usabilidade e UX (Anderson 2016).

Uma variedade de métodos e técnicas tem sido empregada para realizar testes A/B em usabilidade e UX. Entre os mais comuns estão: testes de tempo de conclusão de tarefas, testes de satisfação do usuário (por meio de questionários ou escalas), análise de

métricas de conversão (como taxa de cliques, taxa de abandono e taxa de conversão) e monitoramento do comportamento do usuário por meio de técnicas como eye-tracking e gravação de tela (Kujala et al. 2011).

Os estudos revisados nesta pesquisa apresentaram diversas contribuições e insights relevantes para o uso do teste A/B na avaliação de usabilidade e UX. Alguns trabalhos destacaram a importância de realizar testes A/B em diferentes estágios do processo de design, desde as fases iniciais até a implementação final. Isso permite identificar problemas e oportunidades de melhoria em cada etapa, garantindo uma abordagem iterativa e orientada por dados. Outros estudos ressaltaram a necessidade de considerar o contexto e o perfil dos usuários ao realizar testes A/B. Diferentes grupos demográficos, habilidades e preferências podem influenciar a percepção e a interação dos usuários com uma interface, e é fundamental levar esses aspectos em consideração ao interpretar os resultados e fazer ajustes no design.

## 3.2 Planejamento do Protocolo da Revisão

Nesta seção é descrito o planejamento do protocolo de estudo secundário sobre heurísticas de usabilidade e experiência do usuário para comparação de dois softwares sobre o teste A/B. Isso inclui: as questões de pesquisa, os termos de busca, as fontes de pesquisa e os critérios de seleção dos artigos. Nas subseções a seguir são detalhadas as a parte deste protocolo.

### 3.2.1 Questão de Pesquisa

O objetivo desta revisão é buscar o estado da arte em relação a heurísticas de usabilidade e UX sobre o Teste A/B para avaliação de dois sistemas computacionais, de forma a identificar problemas e através da comparação definir o melhor software entre eles, porém devido a dificuldade da crise da COVID-19, conseguimos realizar a revisão de literatura através do *Google Scholar*.

Para pesquisar o processo de Teste A/B e as metodologias de comparação de software existentes, foram pesquisadas buscando realizar o mapeamento de como o processo foi realizado, seus artefatos para buscar responder a questão principal: **QP:** *Como avaliar comparativamente dois ou mais softwares através da usabilidade e experiência do usuário, para o apoio a tomada de decisão de qual software adquirir?*

### 3.2.2 Termos de Busca

Nesta subseção são apresentados os termos para se montar a *String* de busca, que identificasse publicações científicas que tratem do nosso tema. Com o objetivo de obter o

maior número de publicações relevantes para validar a *String*, para responder a questão principal utilizamos palavras-chaves e seus sinônimos, conforme a Tabela 1 abaixo:

Tabela 1: Palavras-Chave e Sinônimos para montagem da *String* de Busca

Palavras-Chave	Sinônimos
Heuristic evaluation	process; evaluation, inspection, steps
Usability	heuristic, affordance, SUS, learnability, efficiency, memorability, satisfaction
User Experience	UX, questionnaire experience, UEQ, UX surveys, UMUX
A/B Test	split test, comparative test, multivariate testing, testing.

Com base nos termos de busca, a String de busca base resultante foi: (*"heuristic evaluation"OR "process evaluation", OR "inspection"*) AND (*"Usability"OR "heuristic"OR, "learnability"OR "affordance"OR "efficiency"OR "memorability"OR "satisfaction"*) AND (*"user experience"OR "UX"OR "questionnaire experience"*) AND (*"split test"OR "comparative test"OR "multivariate testing"OR "testing"*). Esta string foi utilizada no motor de busca do Google Scholar.

### 3.2.3 Fontes de Busca

Esta revisão de literatura foi limitada a artigos e publicações acessíveis via bibliotecas digitais, indexadas pelo Google Scholar, dentre os quais identificamos diversos artigos disponíveis nas plataformas Scopus Digital, Library, ACM Digital Library, Spring Link Data Library, IEEE Xplore Digital Library e Science Digital Library.

### 3.2.4 Critérios de Seleção

Os trabalhos considerados nesse levantamento bibliográfico são publicações que passaram por uma revisão por pares para que fossem selecionadas as publicações com maior representatividade, que respondam a questão de pesquisa. Foram definidos os seguintes critérios de inclusão e exclusão:

- Critérios de Inclusão (CI) - Todos precisam ser atendidos:
  - **CI1:** O artigo deve possuir heurísticas de usabilidade ou UX para avaliação de um software;
  - **CI2:** Os testes necessitam ter uma comparação entre dois ou mais softwares;
  - **CI2:** Os testes de comparação poderão ser automatizados ou genéricos;

- **CI2:** Os testes devem ser conduzidos de forma semelhante, logo necessitam usar o mesmas ferramentas, procedimentos e variáveis;
- **CI3:** Os resultados dos testes devem ser de natureza comparável.
  - Critérios de Exclusão (CE):
- **CE1:** A publicação não possui detalhes de como o processo é realizado;
- **CE2:** A publicação não possui nenhum tipo de comparação entre dois ou mais softwares;
- **CE3:** A publicação não possuir heurísticas de usabilidade ou UX.

### 3.3 Execução da Revisão de Literatura

Para organizar a etapa de mapeamento dos processos dos artigos, foi criado um formulário de extração de dados apresentado no formulário de extração é utilizado para auxiliar na análise das extrações dos dados oriundas das publicações que são consideradas no resultado final. O formulário possui os itens e descrições, conforme recomendado por (Rivero and Conte 2017) conforme a Tabela 2.

Tabela 2: Formulário de extração de dados dos artigos científicos

Item	Descrição
Nome da abordagem	Descreve se o artigo pesquisado aborda os temas da pesquisa
Atributos avaliados	Identifica se o artigo avalia os itens de usabilidade, experiência do usuário (UX) e teste A/B;
Tipos de problemas identificados	Relato das palavras-chave dos problemas identificados pelos artigos para realizar uma avaliação sobre os atributos definidos
Local de avaliação	Informa se o teste foi realizado de forma presencial, remota ou híbrida.

Após filtrar os artigos utilizando a *String* de busca pelo *Google Scholar*, no período de março de 2022 a abril de 2022, foi realizado o *download* ao todo de 37 publicações.

Tabela 3: Filtros de pesquisa por artigos científicos

Biblioteca	Qtd. Geral	Filtro 1	Filtro 2
Google Scholar	37	19	6

Os 37 artigos retornados passaram por um primeiro filtro, onde ocorreu a leitura do título e do resumo da publicação. Conforme critérios de inclusão e exclusão foram eliminadas ou incluídas a lista de artigos que iremos utilizar no estudo. Logo após este primeiro filtro, 19 publicações restaram como sendo de conteúdo relevante para a extração das informações, sendo que cada uma delas foi lida na íntegra.

Conforme era feita a leitura, foi visto que algumas publicações não apresentavam as heurísticas de usabilidade ou UX para a comparação entre dois software ou duas versões, mas sim de recomendações de como melhorar as heurísticas ou criando novas heurísticas para incluir como atributo de avaliação.

Em outros casos, o artigo buscava obter informações estatísticas um Teste A/B de forma massiva, sem explicar os critérios de usabilidade, apenas apontando o resultado final. Contudo, alguns artigos apresentaram heurísticas de avaliação de usabilidade ou UX com o Teste Comparativo, ou Teste A/B em contextos de páginas web, aplicativos móveis, cliente-servidor e genérica, sendo aplicados de forma remota, híbrida ou presencial. Desta forma, ao todo, 6 publicações foram selecionadas com conteúdo relevante e que atendia todos os critérios de seleção.

### 3.4 Resultados e Discussão

Esta revisão da literatura identificou atributos de qualidade que impactam a usabilidade e a experiência do usuário ao utilizar um teste A/B, para que então seja possível propor tecnologias que contribuam com a melhoria da qualidade destes processos. É importante ressaltar que se um atributo se mostrava semelhante a outros de fontes diferentes, estes eram combinados em um único atributo, garantindo que o processo mais completo fosse selecionado e / ou complementado com detalhes dos demais processos e seus atributos.

Tradicionalmente, o teste A/B permite comparar os requisitos funcionais de um sistema de informação. No entanto, uma abordagem semelhante pode ser usada para melhorar a usabilidade. Assim, novas abordagens foram desenvolvidas para permitir avaliar a usabilidade e a experiência do usuário ao comparar dois ou mais sistemas de informação (Speicher et al. 2014).

Ao todo, seis publicações (1) - (Firmenich et al. 2019), (2) - (Gardey and Garrido 2020), (3) - (Jabbar et al. ), o (4) - (Prasetyo and Soliman 2021),

(5) - (Speicher et al. 2014), (6) - (Vanderdonckt et al. 2019)] se destacaram pela aplicação de teste A/B com usabilidade ou experiência do usuário (UX). A partir destes artigos iremos identificar de forma detalhada as atividades realizadas para a execução dos testes, os seus atributos e os artefatos.

Dentre essas abordagens, Firmenich et al. (2019) propõe um método iterativo suportado por um kit de ferramentas que permite que especialistas em usabilidade projetem testes de usuários, executem-nos remotamente, analisem resultados e avaliem soluções alternativas para problemas de usabilidade de forma semelhante ao teste A/B. De acordo com os autores, cada solução é criada aplicando refatorações web do lado do cliente, ou seja, mudanças nas páginas web no cliente que visam melhorar a usabilidade. O principal benefício de sua abordagem é que ela reduz o custo geral do teste do usuário e, principalmente, do teste A/B, aplicando refatorações para criar soluções alternativas sem modificar o código do servidor do aplicativo.

Em outro artigo, Speicher et al. (2014) desenvolveu o Teste de Divisão baseado em Usabilidade, que é uma abordagem de teste A/B com alavancagem considerando as interações do usuário para derivar métricas quantitativas de usabilidade. Os autores desenvolveram o WaPPU, que é uma ferramenta realizada como um serviço com um repositório central, que permite aos desenvolvedores criar e monitorar projetos de teste A/B. Na ferramenta, é possível escolher entre uma gama de 27 recursos de interação predefinidos, incluindo cliques, comprimento da trilha do cursor e quantidade de rolagem. A ferramenta foi capaz de detectar a diferença prevista de usabilidade com significância estatística para o maior e mais representativo grupo de usuários.

Finalmente, alguns artigos descrevem abordagens de testes A/B de usabilidade usando testes A/B tradicionais considerando instrumentos tradicionais de usabilidade (por exemplo, escala de usabilidade do sistema) (Jabbar et al. ) ou atributos (facilidade de operação, número de etapas, ergonomia, preferência visual, outros) [(Prasetyo and Soliman 2021),(Vanderdonckt et al. 2019)], ou recursos específicos de UX em modelos de aprendizado de máquina automatizados incorporados (Gardey and Garrido 2020). Por exemplo, Jabbar et al. (2021) realiza um teste A/B para melhorar a interatividade de um site. Em sua pesquisa, as tendências de design foram aplicadas no site da (*Quaid-e-Awam University of Engineering, Science and Technology (QUEST)*) para torná-lo mais interativo para os alunos. A usabilidade da versão antiga e da nova versão foi medida para recomendar melhores tendências de design. O estudo foi elaborado com questionário pré-teste e pós-teste elaborado utilizando escala de usabilidade do sistema (SUS) medindo três atributos de usabilidade que são eficácia, eficiência e satisfação.

Apesar da proposta ou aplicação de diferentes abordagens para testes A/B no contexto de Usabilidade e avaliação de UX, ainda existem algumas oportunidades de

melhoria. Por exemplo, após realizar uma revisão de literatura combinando os termos “teste A/B”, “teste dividido”, “teste multivalor”, “teste comparativo”, “usabilidade” e “experiência do usuário”, não identificamos Processo de teste A/B que pode ser aplicado para identificar problemas de usabilidade e experiência do usuário. Além disso, a maioria dos processos de teste A/B no contexto de interação humano-computador não coleta dados qualitativos, o que dificulta identificar o que melhorar na versão selecionada de um software para atender às necessidades dos usuários. Por fim, embora o teste remoto seja importante para coletar dados reais de uso, a maioria dos processos de teste A/B identificados foi realizada dessa maneira, em vez de em laboratório ou em campo, o que, por sua vez, pode ser útil para coletar mais informações do ponto de visão dos usuários, como feedback do usuário e respostas emocionais.

Considerando o exposto, há a necessidade de desenvolver abordagens específicas de avaliação de testes A/B de Usabilidade/UX que possam ser úteis na avaliação de diferentes versões de uma ideia de software ou software concorrente. Além disso, essas abordagens podem ser úteis para definir o conjunto de etapas necessárias para realizar um teste A/B.

### 3.4.1 Modelo de extração de informações dos artigos

Seguindo as atividades de revisão de literatura, em relação a processo de execução dos testes em cada artigo, criamos um modelo de extração de informações em que é possível capturar itens relevantes para a pesquisa, as atividades e os artefatos utilizados durante a execução do seu teste, conforma a Tabela 4 abaixo:

Tabela 4: Formulário de extração de dados dos artigos científicos detalhada

Nome da abordagem:				
Atributos avaliados:				
Tipos de problemas identificados:				
Local de avaliação:				
Tipo de Sistema Avaliado:				
Plataforma do Sistema Avaliado:				
Código	Opcional	Descrição da atividade	Função	Recursos
ACTV01	Sim / Não (Campo vazio)			
ACTV02	Sim / Não (Campo vazio)			
ACTVNN	Sim / Não (Campo vazio)			

Para cada leitura de artigo, foi realizado um mapeamento onde padronizamos as informações, para saber se já existia na literatura avaliações de usabilidade e UX com o

teste A/B, sendo elas:

- O Nome da abordagem: descreve se o artigo pesquisado aborda os temas da dissertação
- Atributos avaliados: identifica se o artigo avalia os itens de usabilidade, experiência do usuário (UX) e teste A/B;
- Tipos de problemas identificados: relato das palavras-chave dos problemas identificados pelos artigos para realizar uma avaliação sobre os atributos definidos.
- Local de avaliação: Informa se o teste foi realizado de forma presencial, remota ou híbrida.
- Tipo de Sistema Avaliado: Identifica se o artigo foi desenvolvido para atender qualquer tipo de sistema (genérico) ou se ele está focado para algum tipo específico de plataforma (Web, Mobile, Cliente-servidor), entre outros que possa vir a surgir.
- Plataforma do Sistema Avaliado: Apresenta se o estudo de caso aplicado no artigo foi realizado sobre alguma plataforma específica.
- Código: Acrônimo criado para mapear a sequência de atividades realizadas pelo artigo, organizado pela ordem cronológica em que os eventos ocorrem durante o experimento. Para este template, criamos a sigla ACTVNN, onde ACTV significa a palavra ATIVIDADE.
- Opcional: Isso ocorreu devido ao fato de que, em alguns casos, os autores realizaram atividades com ou sem o uso de ferramentas específicas que não eram genéricas.
- Descrição da atividade: Descreve as atividades que estão sendo realizadas, em sequência.
- Função: informa qual o perfil de usuário está realizando a atividade do teste.
- Recursos: Informa quais artefatos ou sistemas computacionais são utilizados durante a avaliação.

A partir deste modelo, iniciamos o processo de extração de partes do texto de cada um dos seis artigos, conforme iremos apresentar a seguir.

### 3.4.2 Processo de extração de informações dos artigos

Para realização de um teste de software é importante identificarmos quais são os passos que foram seguidos para a aplicação do teste de software e a partir do modelo de extração desenvolvido foi realizada uma nova leitura sobre cada uma das publicações

destacadas para podermos entender como foi realizado o processo de avaliação comparativa em de cada um dos artigos.

Para podermos demonstrar como foi realizado o processo para os seis artigos, iremos detalhar o processo de extração realizado no artigo de Firmenich et al, para exemplificar o trabalho.

### 3.4.2.1 Artigo 01: Melhoria da usabilidade por meio de testes A/B e refatoração

Em nossa primeira atividade de extração de dados selecionamos o artigo publicado por (Firmenich et al. 2019), que utilizou o método de teste comparativo para resolver problemas de usabilidade. Para este artigo, conseguimos extrair os dados, conforme a Tabela 5. Após a apresentação da tabela, iremos explicar as atividades extraídas.

Tabela 5: Processo de Teste A/B: Atividades obrigatórias

Nome da Abordagem	Teste de usabilidade A/B e método			
Atributos Avaliados	Usabilidade			
Tipos de problemas identificados	Página saturada, página grande, conteúdo obsoleto, conteúdo duplicado, caminhos de navegação longos, organização confusa, rolagem excessiva, eficácia da funcionalidade, eficiência da funcionalidade, satisfação da funcionalidade			
Local de avaliação	Pessoalmente, Remoto			
Tipo de Sistema Avaliado	Generico			
Plataforma do Sistema Avaliado	Web			
<b>Código</b>	<b>Opcional</b>	<b>Descrição da atividade</b>	<b>Função</b>	<b>Recursos</b>

Continuação na próxima página

Tabela 5: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART01_ ACTV01		Selecionar um caso de uso específico a ser testado nesta interação, o definindo como "Teste de Usabilidade"	Especialista em Designer	Scenario Recorder
ART01_ ACTV02	Sim	Decompor a funcionalidade do teste de Usabilidade selecionado a nível de tarefas a serem executadas (Por Exemplo: "Create a new user account "Login"), etc	Especialista em Designer	Scenario Recorder
ART01_ ACTV03	Sim	Associar as tarefas em grupos para criar um Modelo de Cenário de Teste de Usabilidade	Especialista em Designer	Scenario Editor

Continuação na próxima página

Tabela 5: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART01_ ACTV04		Definir as métricas e quais ferramentas serão utilizadas no cenário de teste, quando este for executado	Especialista em Designer	Scenario Editor
ART01_ ACTV05		Recrutar os participantes para executar os cenários de teste	Especialista em Designer	Não definido
ART01_ ACTV06		Responder o questionário demográfico com os participantes	Participantes	Scenario Player
ART01_ ACTV07		Realizar os testes dos cenários de teste seguindo os passos guiados (Por exemplo: Create a new user account “Login”, “Search for products”, “Add products to cart” e “Check-out”)	Participantes	Scenario Player

Continuação na próxima página

Tabela 5: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART01_ ACTV08		Responder ao final do teste um questionário de usabilidade para os participantes	Participantes	Scenario Player / Questionário SUS
ART01_ ACTV09		Analisar os resultados para descobrir problemas de usabilidade (Por exemplo: “Abandoned form”, “Undescriptive element”, and “Misleading link”)	Especialista Projetista	Test Analysis Tool (Visual IA)
ART01_ ACTV10		Criar versões diferentes do aplicativo como soluções para tentar solucionar problemas de usabilidade	Especialista Projetista	Test Analysis Tool (Visual IA)
ART01_ ACTV11		Reaplicar o teste de usabilidade	Especialista em Designer	Test Analysis Tool (Visual IA)
ART01_ ACTV12		Analisar os novos resultados para descobrir problemas de usabilidade	Especialista Projetista	Test Analysis Tool (Visual IA)

Continuação na próxima página

Tabela 5: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART01_ ACTV13		Decidir qual a melhor solução avaliada (A/B/n)	Especialista Projetista	Não definido
ART01_ ACTV14		Implementar as melhorias dos melhores resultados	Especialista Projetista	Não definido
ART01_ ACTV15		Disponibilizar a interface no ambiente de produção	Especialista Projetista	Não definido

No artigo, Fiermenich, criou um processo de teste de usabilidade com o método A/B para avaliar atributos de usabilidade para plataformas *Web* genéricas (qualquer tipo de sistema desenvolvido para a *Web*).

Neste artigo, identificamos inicialmente que para realizar o teste de software foi definido a criação de uma versão de em que se definiu suas funcionalidades (caso de uso) que seriam avaliados durante todo o teste e assim após executarmos o primeiro processo sobre a versão inicial, o qual é denominada Controle (ART1\_ACTV01).

Neste processo o caso de uso foi decomposto a nível de funções da tela para avaliar o local de botões, menus, posição dos campos e cores da tela (ART1\_ACTV02).

Após mapear as funções por tela, cada tela foi agrupada e associada a realização de tarefas para realizar uma ação dentro do software para compor cada cenário de teste (ART1\_ACTV03).

Em seguida avaliou-se que para realizarmos um teste é necessário definir as métricas de como iremos avaliar o caso de uso proposto, como o tempo proposto para cada tarefa do teste e como serão quantificadas as respostas sobre o produto e as ferramentas utilizadas, como exemplo a sua pontuação (ART1\_ACTV04).

O processo de recrutamento de participantes (ART1\_ACTV05) é uma etapa essencial para a realização de um teste, onde é feito o preenchimento do termos de consentimento para autorizar a realização do teste. Também é necessário responder questões para podermos avaliar se eles estão habilitados a realizar o teste sem vício (ART1\_ACTV06).

Seguindo o processo de extração, identificamos ações de descrever a execução das tarefas guiadas de acordo com as funcionalidades que estão presentes no protótipo, como exemplo temos as tarefas: (a) Create a new user account “Login”; (b) “Search for products”; (c) “Add products to cart” e (d) “Check-out” (ART1\_ACTV07).

Após a conclusão do teste de software, é solicitado o preenchimento de um questionário de usabilidade para coletar informações dos usuários que realizaram o teste (ART1\_ACTV08).

Para identificar os problemas de usabilidade, como por exemplo (a) “Abandoned form”, (b) “Undescriptive element”, (c) “Misleading link” foi feita uma análise dos resultados e identificar quais problemas os usuários tiveram em cada caso de uso (ART1\_ACTV09).

Somente após executar e obter os resultados e análise do primeiro teste (Controle), é proposto alterações no software que está sendo observado, gerando assim uma segunda versão para que possa ser possível avaliar se a segunda versão (Variação) é melhor que a versão original (ART1\_ACTV10).

Com a nova versão disponível o teste é reaplicado e desta vez é realizada uma nova avaliação dos resultados (ART1\_ACTV11) para então fazer uma nova análise, agora comparando as duas versões para avaliar qual delas obteve o melhor resultado (ART1\_ACTV12) e assim definir qual a melhor versão gerada (ART1\_ACTV13).

Por fim, ao ser decidida a melhor versão, é construída uma especificação de projeto para implementar o melhor cenário, que resultará na geração do código-fonte da versão vencedora (ART1\_ACTV14) que será disponibilizada no ambiente em produção (ART1\_ACTV15).

Para executarmos o teste, temos os seguintes atores envolvidos no processo de avaliação de software sendo eles o Especialista em Designer para as atividades elaboração e execução do teste (ART1\_ACTIV01, ART1\_ACTIV02, ART1\_ACTIV03, ART1\_ACTIV04, ART1\_ACTIV05 e ART1\_ACTIV11), o participante que preenche os dados demográfico, executa os cenários de teste conforme passos guiados e respondem ao final o questionário de usabilidade (ART1\_ACTV06, ART1\_ACTV07, ART1\_ACTV08) e para criar uma nova versão e desenvolver a solução para o ambiente de produção ficou o responsável o Especialista Projetista (ART1\_ACTV09, ART1\_ACTV10, ART1\_ACTV12, ART1\_ACTV13, ART1\_ACTV14, ART1\_ACTV15).

Por fim, para este teste foi utilizado um software chamado *Visual IA* que facilitava a geração dos cenários de teste através dos seus módulos (Scenário Player, Scenário Recorder e Scenário Editor). Este software facilita a análise dos testes por conseguir a partir dele alterar as telas do sistema e disponibilizar as versões (A/B) para avaliação dos usuários. Além de gravar informações de como foi o uso e os tempos para a realização de cada tarefa.

### 3.4.2.2 Artigo 02: Garantindo a qualidade da interface da Web por meio de testes de divisão baseados em usabilidade

Em nossa segunda atividade de extração de dados selecionamos o artigo publicado por (Speicher et al. 2014), desenvolveu o Teste de Divisão baseado em Usabilidade, que é uma abordagem de teste A/B com alavancagem considerando as interações do usuário para derivar métricas quantitativas de usabilidade. Os autores desenvolveram o WaPPU, que é uma ferramenta realizada como um serviço com um repositório central, que permite aos desenvolvedores criar e monitorar projetos de teste A/B.

Na ferramenta, é possível escolher entre uma gama de 27 recursos de interação predefinidos, incluindo cliques, comprimento da trilha do cursor e quantidade de rolagem. Além disso, ela capaz de detectar a diferença prevista de usabilidade com significância estatística para o maior e mais representativo grupo de usuários.

Para este artigo, a Tabela 6. detalha como foi realizada a extração de dados por atividade.

Tabela 6: Processo de Teste A/B: Atividades obrigatórias

Nome da Abordagem	Teste de divisão baseado em usabilidade			
Atributos Avaliados	Usabilidade			
Tipos de problemas identificados	Página saturada, página grande, conteúdo obsoleto, conteúdo duplicado, caminhos de navegação longos, organização confusa, rolagem excessiva, eficácia da funcionalidade, eficiência da funcionalidade, satisfação da funcionalidade			
Local de avaliação	Pessoalmente, Remoto			
Tipo de Sistema Avaliado	Generico			
Plataforma do Sistema Avaliado	Web			
<b>Código</b>	<b>Opcional</b>	<b>Descrição da atividade</b>	<b>Função</b>	<b>Recursos</b>

Continuação na próxima página

Tabela 6: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART2_ ACTV01		Definir os sites ou páginas a serem avaliados no teste	Especialista em Designer	WaPPU
ART2_ ACTV02		Para um cenário do site a ser avaliado, criar variações no cenário e no site para posteriormente avaliar a versão original e modificada	Especialista em Designer	WaPPU
ART2_ ACTV03		Determinar a intenção do usuário	Especialista em Designer	WaPPU
ART2_ ACTV04		Fornecer um modelo de usabilidade comum por variação gerada (ou identificada) e intenção	Especialista em Designer	Não definido
ART2_ ACTV05		Recrutar participantes para o teste	Especialista em Designer	Lista de E-mail
ART2_ ACTV06		Definir tarefas semi-estruturadas para executar as tarefas no site (“Find a birthday present for a good friend that does not cost more than 50 Euros.”)	Especialista em Designer	Navegador Web

Continuação na próxima página

Tabela 6: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART2_ ACTV07		Orientar os participantes a desbloquear os AD Blocks de seus navegadores Web (browsers)	Especialista em Designer	WaPPU
ART2_ ACTV08		Responder o questionário de dados demográficos	Participantes	WaPPU / SERP
ART2_ ACTV09		Dividir os participantes entre as duas interfaces disponíveis do teste	Especialista em Designer	WaPPU
ART2_ ACTV10		Realizar o teste A/B	Especialista em Designer	WaPPU
ART2_ ACTV11		Acompanhar o comportamento do usuário nas interfaces de um teste durante a execução	Especialista em Designer	WaPPU
ART2_ ACTV12		Responder o questionário de usabilidade (R1) (R2) (R3)	Participantes	UNIT / WaPPU

Continuação na próxima página

Tabela 6: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART2_ ACTV13		Pressionar o botão "finalizar o teste" quando o objetivo for alcançado - Isso pode ser útil para mensurar o tempo de realização de uma tarefa, obtendo dados sobre a eficiência do usuário.	Participantes	WaPPU
ART2_ ACTV14		Analisar os resultados obtivos	Especialista em UX	WaPPU
ART2_ ACTV15		Redesenhar a aplicação com apoio de profissionais de UX	Especialista em UX	SERPS

Em Definir os sites ou páginas a serem avaliadas no teste o artigo quer avaliar a usabilidade de uma interface de busca de um site, onde as versões testadas foram a versão padrão da página de resultados de busca (SERP) e uma versão redesenhada da mesma página. O teste foi realizado utilizando a ferramenta WaPPU para rastrear a interação dos usuários com as duas versões da SERP (ART02\_ACTIV01).

Para um cenário do site a ser avaliado, criar variações no cenário e no site para posteriormente avaliar a versão original e modificada, foram definidos dois componentes principais da SERP para fins de rastreamento: o contêiner que contém todos os resultados da busca (serpResults) e o contêiner que contém a caixa de busca (searchForm)(ART2\_ACTIV02).

A seguir o texto apresenta que a intenção do usuário é detectar diferenças na usabilidade da interface e treinar modelos de qualidade para melhorar o software WaPPU (ART2\_ACTIV03).

Para avaliar o site, foi utilizado o questionário baseado no Inuit para avaliar a usabilidade, assim como o rastreamento de interação, como a barra de navegação, e isolando o restante da página como um componente separado (ART2\_ACTIV04).

A avaliação foi realizada como um estudo de usuário remoto assíncrono cujo fluxo de trabalho é orientado por participantes recrutados por meio de listas de e-mail da empresa (ART2\_ACTIV05) e foram definidas tarefas semi-estruturada para simular que todos os participantes agissem de acordo com uma intenção comum, por exemplo: "Encontre um presente de aniversário para um bom amigo que não custe mais de 50 euros"(ART2\_ACTIV06). Cada participante foi apresentado aleatoriamente a uma das duas interfaces da SERP para concluir sua tarefa.

Os participantes foram instruídos a desativar qualquer bloqueador de anúncios (*AD Blocks*) para reduzir o contexto para diferentes tamanhos de tela apenas e minimizar as flutuações na intenção do usuário (ART2\_ACTIV07).

Antes de iniciar o teste foram coletados dados demográficos dos participantes pela ferramenta WaPPU para identificação do perfil do usuário, como sexo, idade, escolaridade, entre outros (ART2\_ACTIV08) e em seguida de forma aleatória dividiu os participantes entre as duas versões da interface (a original e a redesenhada por especialistas em usabilidade) em que cada participante recebeu apenas uma das versões do software para avaliar (ART2\_ACTIV09).

Após a divisão dos grupos, o teste A/B foi iniciado com a disponibilização da página a ser avaliada (ART2\_ACTIV10).

Durante o teste de usabilidade, foi acompanhado o comportamento dos usuários nas interfaces testadas utilizando o WaPPU para registrar a interação com os componentes da página do site (ART2\_ACTIV11)

Após completar as tarefas, os participantes avaliaram a usabilidade da página do site usando o questionário Inuit exibido pelo WaPPU (ART2\_ACTIV12).

Quando o objetivo da tarefa foi alcançado é importante medir o tempo de realização da tarefa para obter dados sobre a eficiência do usuário. O interessante deste teste é que foi criado um botão para encerrar o teste de forma controlada e garantir que todos os dados necessários para a análise tenham sido coletados (ART2\_ACTIV13).

A seguir os resultados foram analisados usando o questionário Inuit para avaliar a usabilidade percebida pelos participantes após interagirem com as interfaces A e B. Além disso, os dados de interação foram coletados e analisados usando o WaPPU para rastrear o comportamento do usuário em interfaces web. A análise dos dados de interação permitiu aos pesquisadores avaliar o desempenho dos usuários nas diferentes interfaces, incluindo o tempo gasto em cada tarefa, as ações realizadas e a taxa de sucesso (ART2\_ACTIV14).

Combinando as avaliações de usabilidade percebida e de desempenho, os pesquisadores foram capazes de avaliar a eficácia das interfaces A e B e determinar qual delas era mais eficaz em termos de usabilidade e desempenho para assim identificar pontos de melhorias para criar uma nova versão com seu redesenho, baseado nas experiências coletadas com o apoio de profissionais de UX (ART2\_ACTIV15).

Os resultados apresentados no artigo sugerem que a abordagem proposta é eficaz para detectar diferenças na usabilidade da interface e treinar modelos de qualidade, mesmo com um conjunto limitado de dados.

Embora o texto forneça detalhes sobre o design do teste A/B e as etapas do estudo de usuário realizado, não fica claro como os dados foram analisados e como as diferenças entre as versões da interface foram avaliadas quantitativamente. Além disso, não há informações detalhadas sobre as métricas de usabilidade utilizadas para avaliar as versões da interface.

### 3.4.2.3 Artigo 03: Tendências de design da Web e sua usabilidade pelo método de teste A/B

O artigo de (Jabbar et al. ) apresenta um estudo em que o uso de design eficiente e atualizado pode melhorar significativamente a experiência do usuário e a interatividade de um website. Nesse sentido, este estudo aplicou as tendências atuais de design no site da Universidade de Engenharia, Ciência e Tecnologia Quaid-e-Awam (QUEST) em Nawabshah, Sindh, Paquistão, com o objetivo de torná-lo mais interativo para os estudantes, através de um teste A/B e neste artigo, a Tabela 7 detalha as atividades de teste A/B identificadas no texto.

Tabela 7: Processo de Teste A/B: Atividades obrigatórias

Nome da Abordagem	Teste de usabilidade A/B e método
Atributos Avaliados	Usabilidade
Tipos de problemas identificados	Página saturada, página grande, conteúdo obsoleto, conteúdo duplicado, caminhos de navegação longos, organização confusa, rolagem excessiva, eficácia da funcionalidade, eficiência da funcionalidade, satisfação da funcionalidade
Local de avaliação	Remoto

Continuação na próxima página

Tabela 7: Processo de Teste A/B: Atividades obrigatórias (Continuação)

Tipo de Sistema Avaliado	Generico			
Plataforma do Sistema Avaliado	Web			
<b>Código</b>	<b>Opcional</b>	<b>Descrição da atividade</b>	<b>Função</b>	<b>Recursos</b>
ART3_ ACTV01		Determinar qual das versões de site é a mais eficaz	Especialista em Designer	Software Customizado
ART3_ ACTV02		Selecionar aleatoriamente os visitantes para verem a versão do site A ou a opção B	Especialista em Designer	Software Customizado
ART3_ ACTV03		Deterinar a intenção do usuário:A opção B resultaria em uma taxa de conversão mais alta do que a opção A	Especialista em Designer	Software Customizado
ART3_ ACTV04		Definir tarefas semi-estruturadas para executar as tarefas no site (“Quick access bar menu.”)	Especialista em Designer	Software Customizado
ART3_ ACTV05		Dividir os participantes e realizar o teste A/B durante 1 semana	Especialista em Designer	Software Customizado
ART3_ ACTV06		Coletar o tempo de execução das tarefas automaticamente	Especialista em Designer	Software Customizado

Continuação na próxima página

Tabela 7: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART3_ ACTV07		Responder questionário pós-teste para medir a satisfação do usuário	Especialista em Designer	Questionário SUS
ART3_ ACTV08		Avaliar de usabilidade: eficiência, eficácia e satisfação entre a opção A e opção B	Participantes	Questionário SUS
ART3_ ACTV09		Apresentar o software vencedor	Especialista em Designer	Não informado
ART3_ ACTV10		delete	Especialista em Designer	Não informado
ART3_ ACTV11		delete	Especialista em Designer	Não informado

O estudo aplicou as tendências atuais de design no site da Universidade de Engenharia, Ciência e Tecnologia Quaid-e-Awam (QUEST) em Nawabshah, Sindh, Paquistão, com o objetivo de avaliar qual a melhor versão do site (ART3\_ACTV01).

Para avaliar a eficácia do novo design, foi conduzido um estudo de usabilidade com 260 estudantes de graduação e pós-graduação, que realizaram quatro tarefas relacionadas à usabilidade em ambas as versões do site (ART3\_ACTV02).

Para podermos comparar ambas as versões do site, foi criada uma nova versão do site em que a intenção do usuário seria que esta nova versão seria a melhor, pois ela traria em sua interface novas tendências de usabilidade (ART3\_ACTV03).

Em seguida foram definidas as tarefas semi-estruturadas que seriam apresentadas aos participantes durante o teste, sendo elas: (1) encontrar o menu de acesso rápido; (2) encontrar os resultados de graduação; (3) encontrar os resultados de pós-graduação; e (4) encontrar o aviso de licitação (ART3\_ACTV04).

Após isso foram separados os participantes em dois grupos de forma automatizada para que cada grupo tivesse acesso a uma das versões do site. Em seguida o teste foi aplicado durante 1 semana de forma a controlar o acesso dos participantes do estudo (ART3\_ACTV05).

Durante a execução, a eficiência das tarefas foi registrada automaticamente (ART3\_ACTV06), e um questionário pós-teste foi aplicado para ambas as versões, utilizando a escala de usabilidade do sistema (SUS), que mede três atributos de usabilidade:

efetividade, eficiência e satisfação (ART3\_ACTV07).

O estudo seguiu analisando os dados coletados através do questionário mencionado para avaliar como foram as interações dos usuários em ambas as versões do sistema (ART3\_ACTV08).

Os resultados indicaram que a nova versão do site, com as tendências de design atualizadas, foi considerada significativamente melhor em termos de eficiência, efetividade e satisfação do usuário, em comparação com a versão antiga (ART3\_ACTV09).

Por fim o artigo define que é essencial seguir as tendências de design atuais ao projetar *sites* para melhorar sua usabilidade e experiência do usuário e pode fornecer informações úteis para profissionais da área que buscam aprimorar suas habilidades de design e tornar seus *sites* mais eficazes.

#### 3.4.2.4 Artigo 04: AB4Web: um testador A/B on-line para comparar alternativas de design de interface de usuário

O texto apresenta a introdução do AB4Web, uma engine baseada em web que implementa uma versão balanceada e randomizada dos testes A/B multivariados, especificamente projetada para que os profissionais possam comparar prontamente as preferências dos usuários finais para alternativas de interface do usuário, como layout de menus, widgets, controles, formulários ou comandos de entrada visual.

Neste artigo apresentamos na Tabela 8 as etapas do teste A/B identificados no estudo

Tabela 8: Processo de Teste A/B: Atividades obrigatórias

Nome da Abordagem	AB4Web
Atributos Avaliados	Usabilidade
Tipos de problemas identificados	Testar diferentes alternativas de design de interface do usuário
Local de avaliação	Remoto
Tipo de Sistema Avaliado	Generico

Continuação na próxima página

Tabela 8: Processo de Teste A/B: Atividades obrigatórias (Continuação)

Plataforma do Sistema Avaliado	Web			
<b>Código</b>	<b>Opcional</b>	<b>Descrição da atividade</b>	<b>Função</b>	<b>Recursos</b>
ART4_ ACTV01		Apresentar o software AB4Web	Especialista em Designer	AB4Web
ART4_ ACTV02		Testar diferentes alternativas de design de interface do usuário	Especialista em Designer	AB4Web
ART4_ ACTV03		Escolher das variantes de design a serem testadas	Especialista em Designer	AB4Web
ART4_ ACTV04		seleção das métricas de avaliação	Especialista em Designer	AB4Web
ART4_ ACTV05		Criar duas variantes de design	Especialista em Designer	AB4Web
ART4_ ACTV06		Dividir os usuários aleatoriamente em dois grupos	Especialista em Designer	AB4Web
ART4_ ACTV07		Executar o teste A/B com oAB4Web	Especialista em Designer	AB4Web
ART4_ ACTV08		Coletou dados sobre o comportamento dos usuários, como o tempo gasto na navegação e o número de cliques no menu	Participantes	AB4Web
ART4_ ACTV09		Analisar para determinar se existe uma diferença significativa entre as duas variantes de design.	Especialista em Designer	AB4Web

Continuação na próxima página

Tabela 8: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART4_ ACTV10		Decidir qual variante de design deve ser adotada	Especialista em Designer	AB4Web
-----------------	--	--	--------------------------	--------

O artigo de Vanderdonck teve como objetivo apresentar o software AB4WEB como uma ferramenta para teste A/B automatizado (ART4\_ACTV01), e para apresentar sua efetividade, seu estudo no artigo foi para determinar se a posição do menu (à esquerda ou à direita da página) afetava o tempo de navegação do usuário e a eficácia do menu (ART4\_ACTV02).

Para conduzir o teste A/B, foram criadas duas variantes de design: uma em que o menu estava posicionado à esquerda da página e outra em que o menu estava posicionado à direita da página (ART4\_ACTV03), seguido da definição das métricas de avaliação em que iria ser monitorado e coletado informações estatísticas do tempo da tarefa (ART4\_ACTV04).

Para o teste foram selecionadas duas variantes de design dentro das criadas para apresentar versões do menu de forma randômica a esquerda e a direita (ART4\_ACTV05) e dividiu-se os participantes em dois grupos de teste, no modelo de teste A/B multivariado de forma aleatória, divididos em dois grupos, sendo que um grupo viu a variante de design com o menu à esquerda e o outro grupo viu a variante de design com o menu à direita (ART4\_ACTV06).

Em seguida após a preparação do teste iniciou-se a execução do teste A/B através do software AB4Web (ART4\_ACTV07) e durante sua execução coletou dados sobre o comportamento dos usuários, como o tempo gasto na navegação e o número de cliques no menu (ART4\_ACTV08). Após a conclusão do teste, os dados foram analisados para determinar se havia uma diferença significativa no tempo de navegação e na eficácia do menu entre as duas variantes de design (ART4\_ACTV09).

Os resultados da análise mostraram que a posição do menu não teve um efeito significativo no tempo de navegação do usuário ou na eficácia do menu. Isso significa que para aquelas variantes o designer pode escolher a posição do menu com base em outras considerações de design, sem se preocupar com o impacto na navegação do usuário ou na eficácia do menu (ART4\_ACTV10).

#### 3.4.2.5 Artigo 05: Avaliação de usabilidade de sistemas ERP: uma comparação entre SAP S/4 Hana e Oracle Cloud

O artigo apresenta uma metodologia utilizada em um estudo comparativo ou teste A/B entre dois softwares de gestão empresarial: SAP S/4 HANA e Oracle Cloud ERP. O

estudo visa diferenciar simultaneamente os dois softwares do ponto de vista ergonômico e medir sua usabilidade por meio da Escala de Usabilidade do Sistema (SUS) e perguntas adicionais. O texto descreve os participantes, o procedimento experimental e a análise estatística realizada, conforme extração realizada na Tabela 9.

Tabela 9: Extração dos dados do artigo de (Vanderdonckt et al. 2019)

Nome da Abordagem	Comparativo entre dois softwares de gestão empresarial (ERP)			
Atributos Avaliados	Usabilidade			
Tipos de problemas identificados	ergonomia			
Local de avaliação	Pessoalmente, Remoto			
Tipo de Sistema Avaliado	Generico			
Plataforma do Sistema Avaliado	ERP			
Código	Opcional	Descrição da atividade	Função	Recursos
ART5_ACTV01		Comparar dois softwares de gestão empresarial (ERP)	Especialista em Designer	Não informado
ART5_ACTV02		Exclusão de participantes com menos de 2 anos de experiência	Especialista em Designer	Não informado
ART5_ACTV03		Convocação de participantes	Especialista em Designer	Ligações e reuniões <i>online</i>

Continuação na próxima página

Tabela 9: Extração dos dados do artigo de (Vanderdonckt et al. 2019) (Continuação)

ART5_ACTV04		Preparar o Testes com a explicação do escopo do estudo aos participantes	Especialista em Designer	Não informado
ART5_ACTV05		Realizar o teste comparativo (Duplo teste)	Especialista em Designer	Não informado
ART5_ACTV06		Responder questionário	Especialista em Designer	Não informado
ART5_ACTV07		Análisar os dados e obter as estatísticas	Participantes	Avaliação estatística (t-test)
ART5_ACTV08		Apresentar os resultados do teste	Especialista em Designer	Ligações e reuniões <i>online</i>

O texto tem como objetivo de criar uma metodologia de teste comparativo entre softwares de gestão empresarial sobre a visão de especialistas dos softwares de mercado SAP S/4 HANA e Oracle Cloud ERP (ART5\_ACTV01).

Os participantes foram escolhidos com base em sua experiência em gerenciar projetos com esses softwares, em que aqueles com menos de 2 anos de experiência foram excluídos para garantir alta precisão nos resultados (ART5\_ACTV02).

Após aplicação dos critérios de inclusão e exclusão, foram convocados 27 gerentes de projeto de ERP que aceitaram participar do estudo (ART5\_ACTV03).

Em seguida, é realizada a preparação do teste com a explicação do escopo do estudo aos participantes (ART5\_ACTV04), para o início do teste comparativo (ART5\_ACTV05).

Os participantes responderam às mesmas perguntas duas vezes, uma vez para cada software, a usabilidade do software foi medida usando a Escala de Usabilidade do Sistema (SUS) e algumas perguntas adicionais foram incluídas no questionário para avaliar a ergonomia do software (ART5\_ACTV06).

As respostas dos participantes foram analisadas usando estatísticas descritivas para mostrar se há diferença entre as médias dos dados de cada software (ART5\_ACTV07).

E ao final verificar se foi possível diferenciar os dois softwares de uma perspectiva ergonômica e apresentou o software melhor avaliado entre eles (ART5\_ACTV08).

## 3.4.2.6 Artigo 06: Avaliação da Experiência do Usuário por meio de Teste A/B Automático

Este artigo aborda a importância da usabilidade em sistemas de software e como o teste A/B e a refatoração podem ser utilizados para melhorar a experiência do usuário. O artigo discute o processo de teste A/B e como ele pode ser usado para comparar diferentes versões de um sistema, identificando a mais eficaz em termos de usabilidade, com os passos apresentados na Tabela 10.

Tabela 10: Processo de Teste A/B: Atividades obrigatórias

Nome da Abordagem	Melhoria da usabilidade por meio de testes A/B e refatoração			
Atributos Avaliados	Usabilidade			
Tipos de problemas identificados	Página saturada, página grande, caminhos de navegação longos, organização confusa, rolagem excessiva, eficácia da funcionalidade, eficiência da funcionalidade, satisfação da funcionalidade			
Local de avaliação	Remoto			
Tipo de Sistema Avaliado	Generico			
Plataforma do Sistema Avaliado	Web			
<b>Código</b>	<b>Opcional</b>	<b>Descrição da atividade</b>	<b>Função</b>	<b>Recursos</b>
ART6_ ACTV01		Gerar pontuação entre as duas versões da tarefa e interface sendo avaliada	Especialista em Designer	Client-side web refactorings (CSWRs)

Continuação na próxima página

Tabela 10: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART6_ ACTV02		Desenvolver método baseado em Machine Learning (ML) para análise de interação dos eventos de logs de usuarios de interação com web sites	Especialista em Designer	Não Informado
ART6_ ACTV03		Criar versões de aplicações web com técnicas de avaliação semi-automática de problemas de usabilidade sejam encontrados através das análises de interações	Especialista em Designer	Não Informado
ART6_ ACTV04		Aplicar o teste A/B	Especialista em Designer	Client-side web refactorings (CSWRs)
ART6_ ACTV05		Coletar uma grande quantidade de eventos de interações para descobrir problemas de usabilidade para análise de logs	Especialista em Designer	Client-side web refactorings (CSWRs)
ART6_ ACTV06		Explorar comportamento positivo (soluções) e negativos (erros) através de ferramenta de usabilidade automatizada	Especialista em Designer	Client-side web refactorings (CSWRs)

Continuação na próxima página

Tabela 10: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ART6_ ACTV07		Selecionar o elemento com problema de usabilidade	Especialista em Designer	Client-side web refactorings (CSWRs)
ART6_ ACTV08		Investigar o método para fazer as alterações de eficácia através de ferramenta web	Especialista em Designer	Client-side web refactorings (CSWRs)
ART6_ ACTV09		Avaliar o esforço de interação dos usuários em diferentes interfaces	Especialista em Designer	UX-Painter
ART6_ ACTV10		Descobrir micro medidas para que cada elemento possa ser medido e avaliado para a avaliação de UX semi-automática	Especialista em Design	UX-Painter
ART6_ ACTV11		Definir critérios da árvore de decisão através da micro-medidas	Especialista em Designer	UX-Painter
ART6_ ACTV12		Retroalimentar o algoritmo de Machine Learning para melhorar a análise semi-automática	Desenvolvedor	Não Informado

As etapas do teste A/B, conforme discutidas no artigo, começam com a identificação de um objetivo claro para o teste. Isso envolve a definição de métricas (pontuação) de desempenho que serão medidas, como taxa de conversão ou tempo médio de permanência na página (ART6\_ACTV01).

Após a definição de métricas é desenvolvido um método baseado em Machine Learning para análise dos logs que serão gerados durante o teste para coleta de dados de

forma automatizada (ART6\_ACTV02).

Em seguida duas versões da da página web são criadas para que seja possível coletar os dados distintos de ambas as versões para captura do comportamento do usuário sobre as diferenças das duas versões através de ferramentas de análise para rastrear o comportamento dos usuários e comparar as métricas de desempenho entre as versões (ART6\_ACTV03).

Em seguida, os autores aplicam o teste A/B (ART6\_ACTV04) de forma massiva, gerando uma grande quantidade de dados e eventos armazenados em *logs* que serão utilizados para descobrir o comportamento dos usuários dentro das duas versões da página web (ART6\_ACTV05).

Através da análise dos *logs*, com o apoio das ferramentas de análise e a Machine Learning foram explorados os resultados positivos e negativos de usabilidade observadas nas duas versões do software (ART6\_ACTV06).

Realizada a análise dos pontos positivos e negativos, de forma automatizada são isolados os problemas de usabilidade (ART6\_ACTV07) para que um especialista em design possa investigar quais elementos precisam ser alterados e refatorados (ART6\_ACTV08). Durante esta atividade é também observado o esforço dos usuários em interagir com as interfaces (ART6\_ACTV09).

Após está análise são geradas novas micro medidas para cada elemento para que possam ser adicionados novas avaliações sobre os pontos identificados com problemas de usabilidade (ART6\_ACTV10). Estas medidas são adicionadas a Machine Learning como critérios da árvore de decisão (ART6\_ACTV11).

Por fim, o algoritmo retroalimentado é novamente executado para avaliar se houve evolução ou regressão da solução, visando gerar a recomendação automática de ajustes de usabilidade da aplicação (ART6\_ACTV12).

### 3.5 Mapeamento das etapas do processo

Os artigos selecionados descrevem abordagens de testes A/B de usabilidade usando testes A/B tradicionais considerando instrumentos tradicionais de usabilidade (por exemplo, escala de usabilidade do sistema) (Jabbar et al. ) ou atributos (facilidade de operação, número de etapas, ergonomia, preferência visual, outros) [(Prasetyo and Soliman 2021),(Vanderdonckt et al. 2019)], ou recursos específicos de UX em modelos de aprendizado de máquina automatizados incorporados (Gardey and Garrido 2020). Como a proposta de nosso novo processo é realizar um teste A/B com usabilidade e UX para qualquer tipo de software, foi avaliado em cada um dos artigos o que teríamos em comum as suas atividades. Baseado nesta premissa, na

extração dos seis artigos apresentados, observe que algumas atividades no formulário de extração foram marcadas como opcionais.

Isso ocorreu devido ao fato de que, em alguns casos, os autores realizaram atividades específicas que não eram genéricas, ou que exigiriam o uso de ferramentas específicas para execução do teste.

Quando isso acontecia, marcamos essas atividades como opcionais, para indicar sempre que requisitos ou ações específicas pudessem ser necessárias. Observe que este exemplo também indica quem executou a atividade ou é o ator principal para sua execução, além de indicar os recursos que foram necessários para sua execução.

Caso os autores de um artigo analisado não tenham fornecido esta informação, inferimos esta informação. Quando não havia como inferir as informações, indicamos “Não informado”. Ao final do processo de análise, todas as atividades obrigatórias e opcionais foram registradas em tabelas, cada uma para o artigo analisado em nossa revisão de literatura.

Para montagem do nosso processo, baseados em (Rivero and Conte 2017), realizamos a análise de cada uma das atividades dos seis artigos e criamos um novo termo para agrupar todas as atividades que executavam a mesma ação, mas descritas de forma diferente em cada um dos artigos. Feito isso começamos a identificar quais artigos descreviam ou destacavam cada etapa ou atividade relevante que pudesse contribuir na criação da nossa proposta.

Para facilitar a identificação dos artigos e o que cada um contribuiria em cada atividade, criamos uma codificação para cada um dos artigos, de forma a facilitar sua visualização.

Iniciamos a ação da nossa proposta selecionando as atividades consideradas obrigatórias para qualquer tipo de avaliação para identificar nosso pilar principal de sustentação do nosso novo processo.

Ao considerar todas as atividades dentro do processo, a Tabela 11 mostra todas as atividades que devem ser utilizadas para realizar o teste A/B com usabilidade ou avaliação de UX. É importante destacar que apesar de que vários artigos possam mencionar algumas etapas do processo que eles realizaram, os artigos que detalham apresentam informações relevantes para contribuição do novo processo foram marcados.

Após a apresentação da tabela, iremos explicar com mais detalhes a montagem desta proposta de processo.

Tabela 11: Processo de Teste A/B: Atividades obrigatórias

Código	Substituto	Descrição da atividade	[1]	[2]	[3]	[4]	[5]	[6]
ACTV01		Selecionar um caso de uso específico a ser testado nesta interação, o definindo como "Teste de Usabilidade"	X					
ACTV02		Determinar a intenção do usuário		X				
ACTV03	ATV03 pode ser substituído pela: OP_ATV_01	Disponibilizar a versão atual do sistema para o teste			X			
ACTV04		Definir as métricas e quais ferramentas serão utilizadas no cenário de teste, quando este for executado	X			X		
ACTV05		Definir pontuação para as variantes e métricas e interface sendo avaliada						X
ACTV06		Definir tarefas semi-estruturadas para execução das tarefas no site		X				

Continuação na próxima página

Tabela 11: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ACTV07		Preparar o experimento com variantes de design e objetos				X		
ACTV08		Definir as datas de início e fim do teste, instruções textuais, gráficas, termo de consentimento, número de comparações e os dados demográficos a serem capturados				X		
ACTV09		Definir critérios de inclusão e exclusão de participantes					X	
ACTV10		Recrutar os participantes para a aplicação dos cenários de teste	X	X	X		X	
ACTV11		Coletar dados demográficos por entrevista ou questionário	X	X	X			
ACTV12		Apresentar ao participante as orientações para execução do teste				X		

Continuação na próxima página

Tabela 11: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ACTV13		Dividir os participantes entre as duas interfaces disponíveis do teste		X	X	X		
ACTV14		Acompanhar o balanceamento dos participantes por versão da aplicação				X		
ACTV15	ATV15 pode ser substituído pela: OP_ATV_03	Realizar os testes dos cenários de teste seguindo os passos guiados	X	X	X	X		X
ACTV16		Acompanhar o comportamento do usuário nas interfaces de um teste durante a execução		X				
ACTV17		Avaliar o esforço de interação dos usuários em diferentes interfaces						X
ACTV18		Informar a conclusão do teste		X				

Continuação na próxima página

Tabela 11: Processo de Teste A/B: Atividades obrigatórias (Continuação)

ACTV19		Responder ao final do teste um questionário de usabilidade (para os participantes)	X	X	X		X	
ACTV20	ATV20 pode ser substituído pela: OP_ATV_02	Analisar os resultados para descobrir problemas de usabilidade	X	X	X	X	X	X
ACTV21		Decidir qual a melhor solução avaliada (A/B/n)	X			X	X	
		<b>Número total de contribuições por artigo</b>	<b>8</b>	<b>10</b>	<b>7</b>	<b>9</b>	<b>5</b>	<b>4</b>

LEGENDA: [1] - (Firmenich et al. 2019), [2] - (Speicher et al. 2014), [3] - (Jabbar et al. ), [4] - (Vanderdonckt et al. 2019), [5] - (Prasetyo and Soliman 2021), [6] - (Gardey and Garrido 2020)

Como primeira atividade (ACTV01), identificamos que é essencial definirmos casos de uso. Eles importantes em testes de usabilidade porque ajudam a orientar o teste, identificar problemas de usabilidade, priorizar problemas e estabelecer critérios de sucesso. Eles permitem que o pesquisador saiba quais tarefas devem ser realizadas durante o teste e ajudam a garantir que o teste seja direcionado e eficiente (Rubin and Chisnell 2008).

Determinar a intenção do usuário é importante em um teste de usabilidade e UX porque ajuda a orientar o design do sistema, identificar problemas de usabilidade e melhorar a satisfação do usuário (Norman 2013). Sobre esta atividade (Speicher et al. 2014) demonstra que o contexto a ser utilizado o software é muito importante para o teste, o artigo explica a intenção do usuário em seu caso de teste (ACTV02).

Disponibilizar a versão do sistema para o teste A/B permite comparar e avaliar diferentes abordagens de design e funcionalidades para determinar qual delas é mais eficaz em termos de satisfação do usuário, desempenho e usabilidade (Anderson 2016). Em nosso processo (Jabbar et al. ) explica de forma clara como foi realizado esta ação e os cuidados

que devemos ter para disponibilizar as duas versões para caracterizar o teste como um teste A/B (ACTV03).

Devemos utilizar métricas em um teste A/B de usabilidade e UX para coletar dados quantitativos que possam ser usados para comparar o desempenho e a eficácia de duas versões do software. Algumas das principais razões pelas quais é importante utilizar métricas em um teste A/B para (a) Obter dados objetivos e quantitativos; (b) Identificar pontos fracos e fortes; (c) Justificar decisões de design (Nielsen 1994). Durante a extração dos dados do artigo, (Firmenich et al. 2019) e (Vanderdonckt et al. 2019) apresentaram as métricas e ferramentas com informações relevantes para contribuição do processo (ACTV04).

Definir pontuação para as variantes e métricas de interface sendo avaliada em um teste A/B de usabilidade e UX é importante porque permite comparar objetivamente o desempenho de diferentes versões do software, além disso, essa abordagem ajuda a entender qual das duas versões é mais eficaz e a justificar decisões de design com base em dados objetivos (Souza and Silva 2020). Entre os seis artigos (Gardey and Garrido 2020) possui informações relevantes de quais métricas e variantes foram utilizadas em seu estudo (ACTV05).

Definir tarefas semi-estruturadas é uma prática importante para garantir que os participantes do teste tenham uma experiência consistente e possam ser avaliados de forma objetiva. Elas permitem que os participantes tenham liberdade para realizar a tarefa a totalmente, seguindo poucas orientações (Krug 2014).

A definição de tarefas semi-estruturadas são exemplificadas nos artigos de (Firmenich et al. 2019) e (Speicher et al. 2014) com exemplos bem claros quais atividades (passos guiados) os usuários estavam realizando dentro do teste para completar suas tarefas, o que contribui na montagem de nossos artefatos (ACTV06).

Preparar o experimento com variantes de design e objetos ajuda a garantir que as tarefas sejam relevantes e representativas para serem aplicadas de forma prática (Preece et al. 2015). Algumas possíveis variantes possuem diferentes: (a) *layouts*; (b) fluxos; (c) tipos e tamanhos de fontes; (d) formas de exibição; (e) opções de entradas de dados; (f) *feedback* para o usuário.

Em nossos artigos essas variantes foram combinadas de várias maneiras para criar as diferentes versões do teste A/B. Por exemplo, (Vanderdonckt et al. 2019), (Speicher et al. 2014) e (Firmenich et al. 2019) criaram versões do teste de layouts de interface diferentes, mas com as mesmas fontes e cores; (Jabbar et al. ) criou duas versões do teste podem com layouts semelhantes, mas com fluxos de navegação diferentes. O importante é garantir que as variantes testadas sejam relevantes e significativas para os objetivos do teste e para as necessidades dos usuários (ACTV07).

Definir as datas de início e fim do teste é importante para garantir que o teste seja realizado dentro de um prazo definido e que os resultados possam ser analisados em tempo hábil, assim como coletar dados demográficos, como idade, gênero, experiência prévia com o que será testado e nível de escolaridade, ajuda a garantir que o teste inclua uma amostra representativa de usuários e a identificar possíveis padrões ou tendências nos resultados (Nielsen 1994) (ACTV08).

O objetivo de definir critérios de inclusão e exclusão de participantes em um teste A/B é garantir que o público-alvo seja adequadamente representado e que os resultados do teste sejam aplicáveis ao grupo de interesse. Isso pode ajudar a garantir que os resultados do teste sejam mais confiáveis e representativos do público-alvo do produto ou serviço sendo testado (Magazine 2012). O artigo de (Prasetyo and Soliman 2021) apresenta como foi realizado a inclusão e exclusão de participantes e as justificativas utilizadas para não utilizar parte dos entrevistados para executar o teste, evitando assim vícios ou respostas fora do escopo e do contexto pesquisado (ACTV09).

Já a parte de recrutamento dos participantes com base em critérios definidos pode ajudar a garantir que os resultados do teste reflitam as necessidades e expectativas dos usuários do teste. Para isso, mesmo após definirmos os critérios de inclusão e exclusão é preciso ter cuidado ao separar os participantes em grupos para realização do teste e assim minimizar a influência de fatores externos, como nível de experiência ou conhecimento do produto, no resultado dos testes (Magazine 2012). Os artigos (Firmenich et al. 2019), (Speicher et al. 2014), (Jabbar et al. ) e (Prasetyo and Soliman 2021) apresentaram informações relevantes de como foram recrutados os participantes para a realização do teste, que serão base para construção dos nossos artefatos em nossa proposta de um novo processo (ACTV10).

Ao coletar essas informações de dados demográficos, é possível entender melhor as necessidades e expectativas dos diferentes grupos de usuários, o que pode ajudar a identificar problemas de usabilidade e UX específicos para cada grupo. Além disso, os dados demográficos também podem ser usados para segmentar os resultados do teste e identificar padrões e tendências entre os diferentes grupos de usuários. Os artigos (Firmenich et al. 2019), (Speicher et al. 2014) e (Jabbar et al. ) apresentaram os cuidados tomados para esta atividade com o objetivo de balancear os grupos para a aplicação do teste (ACTV11).

É importante apresentar ao participante as orientações para execução do teste de usabilidade e UX porque isso pode influenciar diretamente na qualidade dos dados coletados. as orientações também podem ajudar a garantir que todos os participantes sigam as mesmas instruções e realizem as mesmas tarefas, o que é fundamental para garantir a consistência dos resultados. As orientações devem ser claras e objetivas, incluindo informações sobre o que o participante deve fazer, como fazer e quais os objetivos da tarefa

(Rubin and Chisnell 2008).

Para esta atividade, apesar de todos executarem esta etapa, apenas (Vanderdonckt et al. 2019) apresentou exemplos e orientações mais aprofundadas de como receber o participante e que orientações seriam necessárias para que ele pudesse iniciar o teste. Os demais deixaram apenas indicações que este processo foi realizado (ACTV12).

Ao dividir os participantes em um teste A/B, uma parte é exposta a uma das interfaces (grupo A) e outra parte é exposta à outra interface (grupo B). Ao fazer isso, é possível avaliar qual das interfaces é mais eficaz em atender aos objetivos do usuário e proporcionar uma melhor experiência. Além disso, essa divisão permite identificar com mais precisão o impacto de cada variável testada nas interfaces, o que é importante para tomar decisões informadas sobre o design e o desenvolvimento de produtos (Scroferneker 2021). Os artigos (Speicher et al. 2014), (Jabbar et al. ) e (Vanderdonckt et al. 2019) demonstram como é realizada esta atividade de divisão em seu texto, onde apresentam de forma simultânea os cenários A e B para avaliação de seus usuários (ACTV13).

Acompanhar o balanceamento dos participantes por versão da aplicação é uma atividade importante, pois é necessário garantir que o número de participantes em cada grupo (versão A ou B) seja equilibrado e suficiente para que os resultados sejam estatisticamente significativos.

Caso ocorra um desequilíbrio entre o número de participantes em cada grupo, os resultados podem ser afetados e não refletir corretamente a experiência do usuário com cada versão da aplicação e também ajuda a garantir a validade estatística do teste (Rubin and Chisnell 2008). Sobre este aspecto, o único a apresentar informações de como se manter o balanceamento entre os participantes foi (Speicher et al. 2014) explicando os cuidados para que a divisão fique em 50% para cada versão de forma dinâmica, pois neste caso foi aplicado utilizando o software *WaPPU* (ACTV14).

Realizar os testes dos cenários de teste seguindo os passos guiados é importante para garantir que todos os participantes executem as mesmas tarefas de maneira consistente. Os passos guiados são úteis para avaliar aspectos específicos do produto, como a capacidade do usuário de encontrar um recurso ou concluir uma tarefa específica (Nielsen 1994). Neste aspecto, todos os artigos informaram que utilizaram a técnica de passos guiados, exceto (Prasetyo and Soliman 2021).

É importante deixar claro que os passos guiados são diferentes das tarefas semi-estruturadas, pois as semi-estruturadas são aquelas que oferecem mais flexibilidade para o participante em como alcançar o objetivo proposto (apresentado na ACTV06), enquanto passos guiados são tarefas que são claramente definidas em etapas e oferecem menos flexibilidade para o participante (ACTV15).

Acompanhar o comportamento do usuário durante a aplicação do teste A/B com

usabilidade e UX é importante para entender como eles interagem com as diferentes versões da interface e como suas experiências são influenciadas por essas variações. Isso permite que sejam identificados problemas e oportunidades de melhoria, possibilitando ajustes para aprimorar a usabilidade e a experiência do usuário (Nielsen 1994). Sobre este aspecto apenas o (Speicher et al. 2014) comentou sobre os resultados que obtiveram como por exemplo a captura do movimento do mouse e o posicionamento dos cliques sobre a tela para identificar dificuldades de usabilidade, através do software *WaPPU* (ACTV16).

Para avaliar o esforço de interação dos usuários em diferentes interfaces durante o teste A/B, pode-se utilizar métricas como tempo de conclusão de tarefas, taxa de erro e número de cliques, coleta de dados fisiológicos, como o monitoramento da atividade cerebral, globo ocular, entre outros para obter informações mais precisas sobre o esforço do usuário durante a interação com a interface (Nielsen 1994). Entre os artigos (Gardey and Garrido 2020) comenta sobre a importância de realizar este tipo de avaliação, o qual desenvolveu uma técnica de avaliação do esforço da interação do usuário (ACTV17) manual. Os demais artigos avaliam este item apenas na etapa de análise de resultados ou utilizam softwares automatizados.

A atividade de conclusão do teste é uma etapa essencial do teste A/B com usabilidade e UX, pois permite que o pesquisador obtenha feedback dos participantes sobre a experiência de uso das diferentes variantes da aplicação testadas (Kujala et al. 2011). no artigo de (Speicher et al. 2014) comenta da importância de observar a conclusão do teste para obter os dados iniciais da percepção fisiológica e de tempo que o usuário levou para executar as tarefas (ACTV18).

A atividade de preencher o questionário de UX e usabilidade é uma etapa crucial para obter feedback dos participantes após a realização do teste A/B. Nesse questionário, são avaliados diversos aspectos, tais como a facilidade de uso, a clareza da informação apresentada, a navegabilidade da interface, a estética visual, entre outros fatores relevantes para a experiência do usuário (Lund 2001).

Neste item (Firmenich et al. 2019), (Jabbar et al. ) e (Prasetyo and Soliman 2021) utilizaram o questionário SUS para avaliar a usabilidade; Já (Speicher et al. 2014) informou que utilizou o questionário INUIT, o informam que utilizaram o questionário SUS, o (Gardey and Garrido 2020) e (Vanderdonckt et al. 2019) não deixou claro qual questionário foi aplicado de usabilidade (ACTV19). Em relação ao questionário de UX, não ficou evidenciado qual questionário foi aplicado, apenas questões voltadas para a avaliação do adjetivo Perspicuidade, que é utilizado no questionário UEQ (Nakamura et al. 2021).

Para analisar os resultados e descobrir problemas de usabilidade e UX é por meio da análise quantitativa e qualitativa dos dados coletados durante o teste A/B. A análise quantitativa pode ser feita por meio da análise estatística dos dados, enquanto a análise qualitativa pode ser realizada por meio da observação e análise dos comportamentos

dos usuários durante o teste e também por meio da análise dos *feedbacks* e comentários fornecidos pelos participantes (Rubin and Chisnell 2008).

Nesta etapa todos os artigos apresentaram ações de análise de seus resultados, a exemplo (Firmenich et al. 2019), (Speicher et al. 2014), (Vanderdonckt et al. 2019), (Prasetyo and Soliman 2021) e (Gardey and Garrido 2020) apresentaram uma avaliação sobre métodos quantitativos, em que o cálculo baseava-se na maioria das vezes em informações coletadas pelas ferramentas automatizadas em que eles estavam utilizando. Já (Jabbar et al. ) além de dados quantitativos, ele analisou dentro de sua pesquisa questões qualitativas, como as respostas dos usuários sobre a avaliação do novo portal desenvolvido pela equipe de desenvolvedores (ACTV19).

Por fim, ao tomar a decisão de qual é o melhor software após o teste A/B, é importante considerar não apenas as métricas quantitativas, mas também a usabilidade e a experiência do usuário (UX).

Uma interface que apresente boas métricas de desempenho pode não ser necessariamente a mais usável ou agradável para o usuário final. Portanto, é importante considerar as percepções e *feedbacks* dos usuários durante o teste, bem como os dados de usabilidade e UX coletados. Isso permitirá que se tome uma decisão mais informada e equilibrada sobre qual software deve ser escolhido (Nielsen 2012).

Sobre o que está acima explicado, o artigo de (Firmenich et al. 2019) apresentou que a versão vencedora foi a variação (B) após a refatoração aplicada e trabalho dos especialistas em *Design*; (Vanderdonckt et al. 2019) apresentou que a melhor solução foi a *Frequency-based Menu*; (Prasetyo and Soliman 2021) apresentou que a melhor solução avaliada, baseada em seus critérios o software de *ERP* da *Oracle*. Os demais artigos não apresentaram softwares vencedores, apenas informações estatísticas da análise dos resultados em cada cenário de teste estudado (ACTV20).

Observamos que mesmo tendo em nosso foco um processo de avaliação de software genérico é necessário flexibilizar o processo para que ele possa absorver cenários específicos, sendo necessária substituição de algumas atividades por outras (opcionais) ou adicionando novas atividades para atender a um determinado cenário de teste.

Diante disso, adicionamos estas variações ao nosso estudo, gerando como resultado uma lista de tabelas 12 e 13 que mostram o processo proposto, que se diferencia de outros combinando as atividades básicas dos processos, artefatos e critérios de avaliação, sugerindo métricas para avaliar a usabilidade e a experiência do usuário.

Tabela 12: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais

Código	Descrição da atividade	[1]	[2]	[3]	[4]	[5]	[6]
OP-ACTV-01	Criar versões de aplicações web com técnicas de avaliação de problemas de usabilidade			X			X
OP-ACTV-02	Utilizar técnicas de Machine Learning para análise de interação						X
OP-ACTV-03	Criar versões de aplicações web com técnicas de avaliação semi-automática de problemas de usabilidade sejam encontrados através das análises de interações						X
	<b>Total Número de contribuições</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>3</b>

LEGENDA: [1] - (Firmenich et al. 2019), [2] - (Speicher et al. 2014), [3] - (Jabbar et al. ), [4] - (Vanderdonckt et al. 2019), [5] - (Prasetyo and Soliman 2021), [6] - (Gardey and Garrido 2020)

No entanto, como já mencionado anteriormente, algumas atividades podem ser substituídas, como o ACTV20, em que uma análise pode ser realizada por meio de algoritmos de aprendizado de máquina. Se for esse o caso, a equipe de desenvolvimento pode verificar a referência (Gardey and Garrido 2020) para obter informações sobre como realizar essa atividade. Essas informações são fornecidas nas Tabelas 11 e 12, verificando o código de possíveis atividades substitutas.

Por fim, as atividades extras que não são obrigatórias nem substitutas das atividades obrigatórias podem ser encontradas na Tabela 13. Essas atividades não foram ordenadas, pois podem aparecer de acordo com as necessidades específicas da equipe de desenvolvimento.

Para podermos ilustrar a possibilidade de uso de atividades adicionais, poderíamos utilizar a atividade AD-ATV-18 que explora o comportamento do usuário através de ferramentas automatizadas, desta forma a atividade ACTV16 e ACTV17 na Tabela 11 poderiam receber esta atividade para coleta de dados automatizados.

Tabela 13 – Proposta A/B testando processo por Usabilidade e UX avaliação - Atividades Extras Opcionais que podem ser realizadas se necessário por a desenvolvimento equipe

Código	Descrição da atividade	[1]	[2]	[3]	[4]	[5]	[6]
AD_ATV_01	Decompor a funcionalidade do teste de Usabilidade selecionado a nível de tarefas a serem executadas	X					
AD_ATV_02	Associar as tarefas em grupos para criar um Modelo de Cenário de Teste de Usabilidade	X					
AD_ATV_03	Criar versões diferentes do aplicativo como soluções para tentar solucionar problemas de usabilidade	X					
AD_ATV_04	Reaplicar o teste de usabilidade	X					
AD_ATV_05	Produzir a especificação e projeto de implementação da melhor solução identificada	X					
AD_ATV_06	Implementar as melhorias dos melhores resultados	X		X			
AD_ATV_07	Disponibilizar a interface no ambiente de produção	X					
AD_ATV_08	Orientar os participantes a desbloquear os AD Blocks de seus navegadores Web (browsers)		X				
AD_ATV_09	Redesenhar a aplicação com apoio de profissionais de UX		X				
AD_ATV_10	Carregar as variantes em um diretório com o pool de artefatos a serem utilizados no experimento				X		

AD_ATV_11	Realizar o teste na ferramenta (por Exemplo: clicks on a variant or on the "It's a draw")				X		
AD_ATV_12	Realizar um teste estatístico para validação de dados coletados da pesquisa					X	
AD_ATV_13	Explorar comportamento positivo (soluções) e negativos (erros) através de ferramenta de usabilidade automatizada						X
AD_ATV_14	Selecionar o elemento com problema de usabilidade						X
AD_ATV_15	Investigar o método para fazer as alterações de eficácia através de ferramenta web						X
AD_ATV_16	Descobrir micro medidas para que cada elemento possa ser medido e avaliado para a avaliação de UX semi-automática						X
AD_ATV_17	Definir critérios da árvore de decisão através da micro-medidas						X
AD_ATV_18	Retroalimentar o algoritmo de Machine Learning para melhorar a análise semi-automática						X
	<b>Total número de contribuições</b>	<b>7</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>6</b>

LEGENDA: [1] - (Firmenich et al. 2019), [2] - (Speicher et al. 2014), [3] - (Jabbar et al. ), [4] - (Vanderdonckt et al. 2019), [5] - (Prasetyo and Soliman 2021), [6] - (Gardey and Garrido 2020)

Diante do estudo do processo de mapeamento do teste dos artigos, observou-se a necessidade de criar um novo processo capaz de realizar um Teste A/B com usabilidade e UX, com a criação de novos artefatos simplificados para atender a todos os sistemas computacionais.

## 3.6 Resumo do Capítulo

Neste capítulo, foi realizado um levantamento bibliográfico sobre processos de Testes Comparativos com usabilidade e/ou experiência do usuário (UX). Foram encontrados 6 artigos relevantes que utilizavam estas técnicas de comparação entre softwares ou suas versões. Durante a revisão de literatura foi realizado o mapeamento de como os artigos científicos executaram seus testes de forma a conhecermos o estado da arte da aplicação de testes A/B com Usabilidade e/ou UX, para assim descobrir a existência destes testes e a sua técnica de aplicação.

Sobre o mapeamento dos processos utilizados nos artigos, como mencionado antes, aplicamos o mapeamento de cada etapa descrita nos artigos, assim como seus atributos e artefatos utilizados. Caso uma atividade tenha sido citada mais de uma vez entre os processos considerados, combinamos sua descrição com as demais atividades para fazer uma única atividade que fosse mais compreensível e detalhada. Observando a criação de processos mapeados, percebeu-se a necessidade da criação de um novo processo, chamado ABTUX que consideraria todas as atividades relevantes estudadas e seus artefatos para a aplicação de um teste de usabilidade e UX por um Teste A/B, que iremos demonstrar no próximo capítulo.

## 4 PROPOSTA DO PROCESSO ABTUX v1.0

Neste capítulo apresentamos a proposta de fluxo para avaliação de testes de software com Usabilidade e UX, denominado ABTUX. Baseada nos principais autores que abordam nosso tema, após a revisão de literatura e fundamentação teórica realizada desenhamos um fluxo e determinamos as atividades.

### 4.1 Processo ABTUX e seus Artefatos - Versão 1.0

Com o mapeamento dos processos durante a revisão de literatura, na Tabela 11 e 12 obtivemos 24 atividades, sendo 21 obrigatórias e 3 opcionais para realizarmos o teste. Estas atividades foram organizadas em um diagrama de atividades para fazerem parte do processo de avaliação do teste de usabilidade e experiência do usuário através de um teste A/B.

A figura 3 que possui todas as atividades do processo. Iremos explicar suas etapas e como montamos cada atividade em seguida.

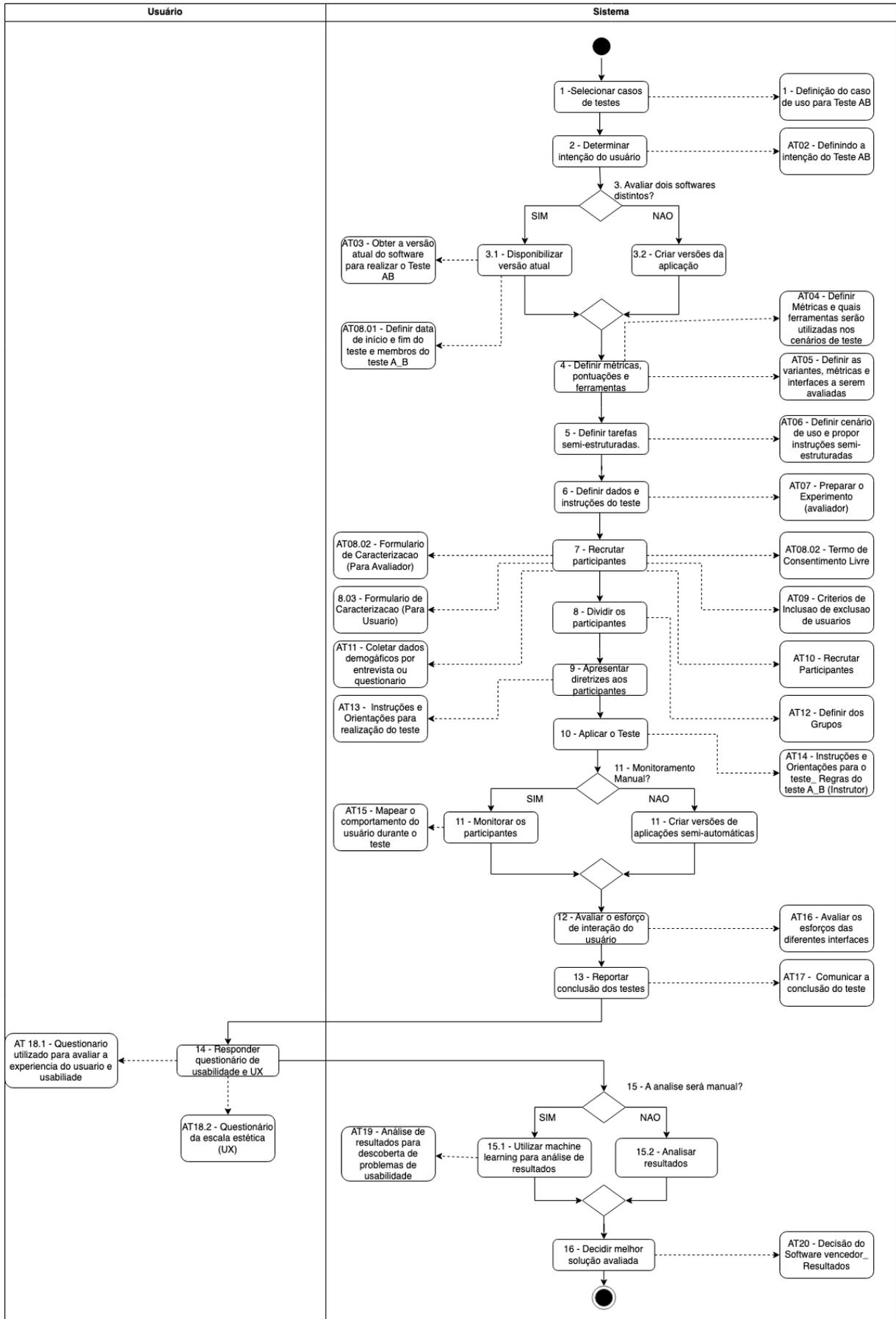


Figura 3 – Fluxo do Processo ABTUX v1.0

### 4.1.1 Fase de Iniciação

Inicialmente, para montagem do processo buscou-se seguir as atividades mapeadas nas tabelas 11 e 12, onde tivemos a primeira atividade do processo ABTUX "1 - Selecionar um caso de teste", com as mesmas características da associado a ACTV01 explicada anteriormente. Para complementar a explicação e orientações de como será realizado esta etapa do processo, criamos um artefato vinculado a esta atividade "AT01 - determinar o caso de uso para o Teste A/B" em que disponibiliza informações sobre a finalidade do teste, como, por exemplo, avaliar a eficácia de um novo layout ou a usabilidade de um novo recurso em um aplicativo.

Seguindo a nossa referência de processo, foi criado no processo ABTUX a atividade "2 - Determinar a intenção do usuário", com as mesmas características da ACTV02, sem alterações no processo proposto. Para complementar as orientações do processo foi criado o artefato "ART02 - Definindo a intenção do teste A/B" contendo informações do que é necessário definir, como: (a) objetivo do teste, ou seja, o que se deseja medir ou avaliar com o teste; (b) Identificação de problemas de usabilidade; (c) avaliação da satisfação do usuário; (d) Mensuração do desempenho em tarefas específicas, entre outros.

Observando o fluxo da 11 e 12 existe um mapeamento em que a próxima etapa do processo possui duas alternativas, sendo elas: ACTV03 e OP-ATV-01. Como temos a opção em seguir por um dos caminhos, foi necessário adicionarmos ao nosso fluxo um ponto de decisão para resolver a situação. Então inserimos no fluxo um operador de decisão com a seguinte Questão: "3 - Avaliar dois softwares distintos?". Caso a condição seja verdadeira (SIM) o item "3.1 - Disponibilizar a versão atual" é ativada e caso a condição seja falsa (NÃO) o item "3.2 - Criar versão de aplicação" é acionado.

Para ambas opções acima, foi criado o artefato "AT03 - Obter a versão atual do software para realizar o teste A/B, contendo informações do nome, da versão atual, funcionalidades de cada software.

### 4.1.2 Fase de Planejamento

A seguir foi criada a atividade "4 - Definir métricas, pontuações e ferramentas". Nesta atividade unificamos a ACTV04 e ACTV05 em uma única atividade, mas gerando dois artefatos, que iremos explicar em seguida.

O artefato "AT04 - Definir Métricas e quais ferramentas serão utilizadas nos cenários de teste", possui campos genéricos para inserir informações, como: (a) Taxa de conversão; (b) Taxa de rejeição; (c) Facilidade de uso; (d) Satisfação do usuário. Em relação a ferramentas, existem diversas disponíveis, como exemplo: Google Optimize, Google Analytics, SurveyMonkey, entre outras.

Completando a atividade foi elaborado o artefato "AT05 - Definir as variantes,

métricas e interfaces a serem avaliadas", com campos onde posamos preencher dados como as variantes, as métricas e as interfaces. As variantes podem ser, por exemplo as versões A e B de diferentes versões do sistema. Sobre as métricas podemos utilizar o tempo médio de conclusão da tarefa. Por fim as interfaces poderiam ser a página inicial de um site ou aplicativo ou um formulário de inscrição.

Avançando pelo processo, criamos a atividade "5 - Definir tarefas semi-estruturadas", com os conceitos herdados da ACTV06. Em seu artefato "AT06 - Definir cenário de uso e propor instruções semi-estruturadas-v2"exibe informações como a (a)Tarefa; (b) Requisito, (c) Descrição e Tempo estimado para execução da tarefa.

Após o preenchimento das informações acima mencionadas, é preciso "6 - definir dados e as instruções do teste". É importante estabelecer uma data e hora para o teste e garantir que o local esteja equipado com o necessário para realizar o teste, como um computador com acesso à internet, webcam, microfone, entre outros. Esta atividade unificou a ACTV08 e ACTV09 do mapeamento anterior .

Para a atividade acima, 3 artefatos foram criados devido a termos 3 passos importantes, sendo o primeiro passo que o avaliador receberá as informações de como deverá executar o teste prático e como deverá orientar os participantes; No segundo passo o avaliador preencherá informações como: (a) Caminho para chegar ao software; (b) Data da execução do teste; (c) Hora da execução do teste; (d) Versão do produto (e) Ano da versão do produto. Estas informações serão repassadas aos participantes do teste, criado no "AT08.01 - Definir data de início e fim do teste e membros do teste A/B".

Por fim, nesta atividade, temos o modelo do termo de consentimento em que os participantes deverão preencher para participar do teste, disponível no artefato "AT08.02 - Termo de consentimento Livre AT".

Chegamos agora na atividade "7 - Recrutar participantes"associada a ACTV10 e ACTV11 do mapeamento da tabela 11. Já mencionamos na ACTV10 informações relevantes sobre os critérios de inclusão e exclusão dos participantes existentes no artefato "ART09 - Critérios de inclusão e Exclusão de usuários". Para os demais iremos apresentar dois exemplos: "AT10 - Recrutar participantes": Recrutar participantes apresenta opções de acionamento do participante através de: (a) Convites por e-mail; (b) Publicar anúncios em redes sociais direcionados para o público-alvo do produto.

Seguindo a premissa dos critérios de inclusão e exclusão, para podermos "AT11 - Coletar dados demográficos por entrevista ou questionário"perguntamos informação aos possíveis participantes, como: (a) idade, (b) gênero, (c) ocupação, (d) nível de escolaridade, (e) se já utilizou o software antes, entre outras informações relevantes para entender o perfil dos usuários participantes.

### 4.1.3 Fase de Execução

Na execução do fluxo, seguimos para a atividade "8 - Dividir os participantes". Baseados nos critérios disponibilizados no AT12 - Definir grupos, os participantes são separados de forma a termos um balanceamento entre as versões que irão ser testadas.

A atividade acima está vinculada com a ACTV13 da tabela 13. Ajustamos esta etapa do fluxo para facilitar a operação do próximo passo. Iremos explicar a seguir o motivo.

Após a separação dos grupos iremos "9 - Apresentar diretrizes aos participantes", momento em que o avaliador irá repassar as informações de como será aplicado o teste, mas agora voltado para a visão do usuário. Estas instruções deverão ser preenchidas e estar disponíveis no "AT13 - Instruções e Orientações para realização do teste", vinculada com o ACTV12.

Observe que caso a orientação viesse antes de dividir os grupos, todos os participantes receberiam as instruções do cenário do teste (controle e tratamento). Esta ação causaria a anulação dos dados conforme premissas para execução de um teste A/B.

No momento que inicia a atividade "10 - Aplicar o Teste" os participantes iniciam sua atividade de avaliação do software. Criamos este marco dentro do processo ABTUX de forma isolada para podermos orientar o avaliador a seguir todos os passos e premissas para executar o teste A/B, de forma que ele não desvie em nenhum momento das premissas, com o risco de invalidar todo o processo. o artefato "AT14 - Instruções e Orientações para o teste-Regras do teste A/B (Instrutor)" contém um guia rápido dos pontos essenciais para validar e caracterizar um teste A/B.

Já na etapa seguinte, novamente temos no mapeamento da tabela 11 uma atividade substituta. Para poder avançar no fluxo, criamos o operador de decisão "11 - Monitoramento manual?", em que caso falso (NÃO) significa que o avaliador irá utilizar alguma ferramenta automatizada para capturar as ações dos participantes durante o teste. Como existe diversas ferramenta e técnicas de coleta, deixamos livre esta etapa para que o avaliador possa ter liberdade em realizar esta atividade. Esta atividade "11.2 - Criar versões de aplicações semi-automáticas" está vinculada a tabela 12: OP\_ATV\_03, já estudada.

Caso a resposta seja verdadeira (SIM), podermos seguir no processo "11.1 - Monitorar os participantes", associado a ACTV15, já explicada anteriormente. Para auxiliar o avaliador, criamos um artefato "AT15 - Mapear o comportamento do usuário durante o teste" com informações de como realizar o monitoramento manual dos participantes. No documento possui informações, como exemplo de como obter o tempo, observar o seu comportamento.

Complementando a atividade anterior, a atividade "12 - Avaliar o esforço de interação do usuário" associada com a ACTV16, onde criamos o artefato "AT16 - Avaliar

esforços das interfaces"com orientações de como auxiliar os participantes em caso de dúvidas, ou dificuldades de interação, sem afetar os resultados do teste, vinculada a ACTV17 .

Quando o teste é concluído "13 - Reportar a conclusão dos testes", associada a ACTV18, disponibilizamos orientações de como proceder quando o participante encerra suas atividades de forma a não impactar nos demais que ainda não conseguiram concluir todas as tarefas do teste através do artefato "AT17 - Comunicar a conclusão dos testes".

Em seguida, é necessário o participante "14 - Responder questionário de usabilidade e UX"sobre a sua experiencia de uso durante o teste. Nesta versão do ABTUX disponibilizamos dois tipos de Questionário, sendo o primeiro de uabilidade e o segundo de UX nos artefatos a seguir: (a) "18.1 - Questionário utilizado para avaliar a experiência do usuário e usabilidade"; (b) "AT18.2 - Questionário de escala estética (UX)". Esta atividade está vinculada a ACTV19.

#### 4.1.4 Fase de Resultados e Conclusão

Novamente, na tabela 11 temos no resultado do estudo a possibilidade de seguirmos duas formas para avaliar os resultados do teste, sendo uma de forma manual e outra automatizada, logo criamos um operador de decisão com a seguinte questão: "15 - A análise será manual?", caso falso (NÃO) iremos seguir com a atividade "15.1 - utilizar machine learning para análise de resultados", associado a tabela 12: OP\_ATV\_02. Desta forma o avaliador poderá ter liberdade no uso de qualquer software que facilite a coleta dos resultados.

Caso positivo (SIM) iremos para a atividade "15.2 - Analisar Resultados", vinculada a ACTV20 já explicada anteriormente. Para a análise manual, criamos o documento "19 - Análise de resultados para descoberta de problemas"com as dicas de como extrair estas informações e técnicas que podem ser utilizadas de forma quantitativa e qualitativa.

Por fim, devemos decidir a melhor solução avaliada "16 - Decidir melhor solução avaliada"para definir qual o software ou versão do software é a melhor, baseado na análise dos resultados e opinião dos participantes. O artefato "AT20 - Decisão do Software vencedor - resultados"é um template modelo de exemplo de informações com o resumo de todo o estudo realizado, contendo informações com: (a) Resumo; (b) Método; (c) Resultados - Qualitativos e Quantitativos; (d) Análise, discursão, descobertas e Recomendações; (f) Conclusão; (g) Apêndice. Esta atividade está vinculada com a ACTV20 da tabela 13.

Após passarmos por todas as atividades do processo ABTUX, acima mencionados, bem como cada um de seus artefatos produzidos para a execução do processo, foi elaborada a tabela 14 contendo um resumo do processo, com as informações essenciais de como foi estruturado o processo, contendo as seguintes informações: (a) Identificador da atividade

(ID); (b) Atividade;(c) Descrição da atividade (Resumo); (d) Artefato vinculado; (e) Referência.

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais

ID	Atividade	Descrição da atividade	Artefato vinculado
1	Selecionar Testes	Essa atividade definir os sistemas computacionais que queremos realizar o teste	AT01 - Definindo o caso de uso para o teste A/B
3.1	Disponibilizar versão atual	Neste item os sistemas a serem comparados são existentes, logo você deve informar qual a versão dos sistemas computacionais farão parte do teste	AT03 - Obter a versão atual do software para realizar o teste A/B
3.2	Criar versão da aplicação	Cria uma nova versão da aplicação que iremos comparar (variação) que fará parte do teste	

Continuação na próxima página

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

4	Definir métricas pontuações e ferramentas	Atividade em que iremos definir como iremos realizar as ferramentas e as medições de tarefas a serem realizadas no teste (tempo, pontuações, etc)	AT04 - Definir Métricas e quais ferramentas serão utilizadas nos cenários de teste AT05 - Definir as variantes, métricas e interfaces a serem avaliadas
5	Definir Tarefas semi-estruturadas	Nesta atividade iremos descrever a execução das tarefas programadas de acordo com as funcionalidades que estão presentes em ambos os sistemas	AT06 - Definir cenário de uso e propor instruções semi estruturadas-v2
6	Definir dados e instruções de teste	Define a data, local a forma de realização dos testes, e define as instruções para realização do teste	AT07 - Preparar o Experimento (avaliador) AT08.1 - Definir data de início e fim do teste e membros do teste A/B AT08.2 - Termo de Consentimento Livre AT

Continuação na próxima página

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

7	Recrutar participantes	Nesta atividade deverá ser necessário avaliar se os grupos e pessoas possuem algum critério que possa viciar ou prejudicar a avaliação da funcionalidade do sistema computacional	AT09 - Critérios de Inclusão e exclusão de usuários AT10 - Recrutar Participantes AT11 - Coletar dados demográficos por entrevista ou questionário
8	Dividir os participantes	Separa os participantes, conforme premissas estabelecidas de validação do Teste A/B, para que a avaliação esteja balanceada.	AT12 - Definir os Grupos
9	Apresentar diretrizes aos participantes	Apresentar o resumo de como o teste será aplicado e o que o participantes deverá executar	AT13 - Instruções e Orientações para realização do teste
10	Aplicar o Teste	Orientações de como aplicar o teste na visão do instrutor para o início da execução do teste e seu monitoramento	AT14 - Instruções e Orientações para o teste_Regras do teste A/B (Instrutor)

Continuação na próxima página

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

11.1	Monitorar os participantes	Apresenta orientações de como observar e coletar dados de comportamento dos participantes, durante a aplicação do teste	AT15 - Mapear o comportamento do usuário durante o teste
11.2	Criar versões de aplicações semi-automáticas	Atividade opcional, para caso quem esteja aplicando o teste utilize ferramentas para apoio ao monitoramento e coleta dos dados dos participantes.	N/A
12	Avaliar o esforço de interação do usuário	Durante a execução do teste é necessário avaliar e acompanhar as ações realizadas pelos participantes, suas dificuldades e dúvidas durante o teste	AT16 - Avaliar os esforços das diferentes interfaces
13	Reportar conclusão dos testes	Orientações do que fazer quando o teste for concluído tanto para o participante, quanto para o instrutor.	AT17 - Comunicar a conclusão do teste

Continuação na próxima página

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

14	Responder questionário de usabilidade e UX	Após a execução das tarefas, os testadores precisarão responder ao questionário para coleta dos dados do teste aplicado. Este ponto é muito importante para o nosso processo	AT18.1 - Questionário utilizado para avaliar a experiência do usuário e usabilidade AT18.2 - Questionário da escala estética (UX)
15.1	Utilizar <i>machine learning</i> para análise de resultados	Em nosso processo, deixamos como opcional o uso de ferramentas automatizadas para análise da usabilidade e experiência do usuário (UX), pois alguns testes já possuem ferramentas próprias para acelerar e facilitar a coleta e análise dos dados	N/A
15.2	Analisar resultados	Apresentar os resultados e divulgar qual foi o melhor entre as funcionalidades e nos sistemas computacionais em geral	AT19 - Análise de resultados para descoberta de problemas de usabilidade

Continuação na próxima página

Tabela 14: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

16	Decidir melhor solução avaliada	Apresentar os resultados e divulgar qual foi o melhor entre as funcionalidades e nos sistemas computacionais em geral	AT20 - Decisão do Software vencedor - Resultados
----	---------------------------------	---	--

Em virtude da nossa proposta do processo ABTUX utilizar teste A/B para avaliarmos com usabilidade e UX, iremos demonstrar estes dois artefatos com exemplos de informações preenchidas durante o estudo do processo.

#### 4.1.5 Avaliação da Usabilidade

O artefato de usabilidade "AT18.1 - Questionário utilizado para avaliar a experiência do usuário e usabilidade" foi desenvolvido sobre a perspectiva das heurísticas de Nielsen.

As heurísticas de Nielsen são um conjunto de princípios de design de interface de usuário que foram desenvolvidos por Jakob Nielsen e Rolf Molich. Essas heurísticas são uma lista de dez diretrizes amplas que podem ajudar a avaliar a qualidade geral da interface do usuário de um produto de software (Nielsen 1994).

Elaboramos um resumo das heurísticas de Nielsen na tabela 15 a seguir:

Tabela 15: Lista de Heurísticas de Nielsen

Item	Heurística	Item	Heurística
a	Visibilidade do status do sistema	f	Reconhecer ao invés de lembrar
b	Concordância do sistema com o mundo real	g	Flexibilidade e eficiência de uso
c	Controle e liberdade do usuário	h	Estética e design minimalista
d	Consistência e padrões	i	Ajudar os usuários a reconhecer, diagnosticar e corrigir erros
e	Prevenção de erros	j	Ajuda e documentação

Legenda: Tabela de heurísticas propostas por Nielsen

Detalhando a tabela, temos as seguintes informações:

- **a. Visibilidade do status do sistema:** O sistema deve manter os usuários informados sobre o que está acontecendo, fornecendo *feedback* adequado dentro de um tempo razoável.
- **b. Correspondência entre o sistema e o mundo real:** O sistema deve falar a linguagem do usuário, com palavras, frases e conceitos familiares para o usuário, em vez de termos orientados ao sistema.
- **c. Controle e liberdade do usuário:** Os usuários devem ser capazes de controlar a interação com o sistema, e o sistema deve oferecer uma maneira de cancelar ou desfazer ações.
- **d. Consistência e padrões:** Os usuários não devem precisar se perguntar se palavras, situações ou ações diferentes significam a mesma coisa em diferentes partes do sistema. As convenções de nomenclatura e interface devem ser consistentes em todo o sistema.
- **e. Prevenção de erros:** O sistema deve ser projetado de maneira que minimize a chance de erros do usuário, com medidas como confirmações e avisos de erro.
- **f. Reconhecimento em vez de lembrança:** O sistema deve minimizar a carga de memória do usuário, tornando objetos, ações e opções visíveis. O usuário não deve precisar se lembrar das informações de uma tela para outra.
- **g. Flexibilidade e eficiência de uso:** O sistema deve permitir que usuários experientes realizem tarefas mais rapidamente por meio do uso de atalhos ou outras formas de acelerar a interação.
- **h. Estética e design minimalista:** Os diálogos do sistema não devem conter informações irrelevantes ou raramente necessárias. Qualquer informação que não seja crucial deve ser suprimida ou apresentada de maneira menos proeminente.
- **i. Ajuda os usuários a reconhecer, diagnosticar e corrigir erros:** O sistema deve fornecer ajuda clara e específica sobre erros, com sugestões concretas sobre como corrigir o problema.
- **j. Ajuda e documentação:** Mesmo que seja preferível ter um sistema fácil de usar sem documentação, pode ser necessário fornecer ajuda e documentação adequadas para que os usuários entendam como realizar tarefas e usar recursos do sistema.

As heurísticas de Nielsen são amplamente utilizadas na avaliação da usabilidade de produtos de software e têm sido usadas para melhorar a experiência do usuário em

muitos sistemas diferentes. Embora não sejam uma lista exaustiva de diretrizes de design de interface, elas oferecem uma base sólida para avaliar a qualidade geral de um sistema de software (Nielsen 1994).

#### 4.1.5.1 Questões objetivas

A partir destas heurísticas, criamos uma tabela genérica, onde disponibilizamos para o avaliador inserir informações que ele gostaria de avaliar. Para cada pergunta, pelo menos uma das heurísticas de Nielsen precisa estar associada para que após a avaliação seja possível avaliar os resultados. Na tabela 16 apresentamos um exemplo de questionário para avaliar um site e as heurísticas associadas:

Tabela 16: Questionário x Heurísticas de Nilsen

Pergunta	Heurística
É fácil encontrar o que se procura no site?	Visibilidade do status do sistema
As informações são apresentadas de forma clara e organizada?	Correspondência entre o sistema e o mundo real
As funcionalidades do site são fáceis de usar?	Controle e liberdade do usuário
O tempo de carregamento do site é aceitável?	Consistência e padrões
O site oferece feedback suficiente ao usuário?	Prevenção de erros
As mensagens de erro são claras e ajudam o usuário a corrigir o problema?	Prevenção de erros
O site oferece opções de navegação intuitivas?	Flexibilidade e eficiência de uso
O design do site é agradável e atraente, sem distrações desnecessárias?	Estética e design minimalista
As informações de ajuda e suporte estão facilmente acessíveis?	Ajuda e documentação
O site é acessível em diferentes dispositivos e navegadores?	Compatibilidade com o usuário e controle do usuário

Em relação as perguntas, sobre as heurísticas, baseados em (Nielsen 1994):

- "O site informa claramente o status do processo de cadastro ou de outras

**ações do usuário?"** - Pergunta relacionada à heurística de visibilidade do status do sistema, pois o usuário precisa ser informado sobre o que está acontecendo no site. Caso contrário, pode ficar confuso e não saber se sua ação foi bem sucedida ou se precisa realizar alguma outra ação.

- **"O site oferece um layout visualmente claro e organizado?"** - Está relacionada à heurística de consistência e padrões, pois é importante que o layout do site siga uma estrutura visual coerente e organizada, de forma que o usuário consiga se orientar e identificar elementos de forma intuitiva.
- **"Os botões e links são facilmente identificáveis?"** - Questão relacionada à heurística de correspondência entre o sistema e o mundo real, pois é importante que os elementos do site correspondam ao que o usuário espera encontrar no mundo real. Por exemplo, um botão com o ícone de um envelope deve ser usado para enviar e-mails, já que isso é uma correspondência com o que é comum no mundo real.
- **"O site oferece uma navegação clara e intuitiva?"** - Está relacionada à heurística de flexibilidade e eficiência de uso, pois o usuário deve ser capaz de navegar pelo site de forma rápida e eficiente, sem precisar passar por vários passos desnecessários.
- **"O site apresenta informações de forma clara e compreensível?"** - Está relacionada à heurística de prevenção de erros, pois é importante que o site apresente informações claras e precisas para evitar que o usuário cometa erros. Se as informações forem apresentadas de forma confusa ou ambígua, o usuário pode cometer erros ao realizar ações no site.
- **"O site oferece opções de personalização para o usuário?"** - Questão relacionada à heurística de controle e liberdade do usuário, pois é importante que o usuário tenha controle sobre sua experiência no site. Oferecer opções de personalização pode aumentar a satisfação do usuário e fazê-lo sentir que tem controle sobre o que está acontecendo.
- **"O site oferece feedback para as ações do usuário?"** - Questão relacionada à heurística de feedback para o usuário, pois é importante que o usuário receba *feedback* sobre suas ações no site. Isso pode ajudá-lo a entender se a ação foi bem sucedida ou se precisa realizar alguma outra ação para completar sua tarefa.
- **"O site oferece ajuda e documentação suficientes para orientar o usuário?"** - Questão relacionada à heurística de ajuda e documentação, pois em alguns casos, os usuários precisam de ajuda adicional para utilizar todas as funcionalidades do site. Se o site não oferece ajuda suficiente, os usuários podem ficar confusos e desistir de utilizar o site.

- **"As opções de navegação são apresentadas de forma clara e visível?"** - Questão relacionada à heurística de reconhecimento em vez de lembrança, pois é importante que as opções de navegação sejam apresentadas de forma clara e visível para que o usuário possa identificá-las facilmente. Se as opções não são claramente visíveis, o usuário pode não perceber que elas existem e pode ter dificuldade em encontrar o que está procurando.
- **"O site é responsivo e funciona bem em diferentes dispositivos?"** - Pergunta relacionada à heurística de usabilidade em dispositivos móveis, pois é importante que o site funcione bem em diferentes dispositivos, como *smartphones* e *tablets*. Caso contrário, o usuário pode ter dificuldade em utilizar o site em dispositivos móveis e isso pode prejudicar a experiência do usuário.

#### 4.1.5.2 Questões subjetivas

As questões subjetivas são importantes para avaliar a usabilidade de um sistema, pois elas permitem obter a percepção e opinião do usuário em relação à usabilidade do sistema (Nielsen 1994). Enquanto as questões objetivas avaliam aspectos concretos e mensuráveis da usabilidade, como tempo de resposta, taxa de erro e eficiência, as questões subjetivas exploram as impressões, sentimentos e opiniões dos usuários sobre o sistema (Kruger et al. 2016). Abaixo, segue alguns exemplos de avaliação subjetiva:

- **"Você acha que o site é fácil de usar?"** - Essa pergunta é subjetiva, pois depende da percepção do usuário. Através dela, é possível avaliar a heurística de facilidade de uso.
- **"Você acha que o site é agradável visualmente?"** - Essa pergunta também é subjetiva, pois depende da opinião do usuário em relação ao design do site. Através dela, é possível avaliar a heurística de estética e design minimalista.
- **"Você sentiu-se seguro ao inserir suas informações pessoais no site?"** - Essa pergunta também é subjetiva, pois depende da percepção do usuário em relação à segurança do site. Através dela, é possível avaliar a heurística de segurança.
- **"Você conseguiu encontrar rapidamente as informações que estava procurando?"** - Essa pergunta é subjetiva, pois depende da percepção do usuário em relação à rapidez de navegação. Através dela, é possível avaliar a heurística de eficiência.

Tanto as questões objetivas quanto as subjetivas são importantes na avaliação da usabilidade de um sistema. As questões objetivas avaliam aspectos específicos e mensuráveis da usabilidade, enquanto as questões subjetivas permitem entender a percepção do usuário

sobre a interface e a experiência como um todo, fornecendo informações importantes para identificar problemas e melhorias em relação à usabilidade do sistema. A combinação de ambas as abordagens permite obter uma avaliação completa e abrangente da usabilidade do sistema (Nielsen 1994).

#### 4.1.6 Questionário de Experiência do usuário

Em nosso processo ABTUX, no artefato "AT18.12 - Questionário da escala estética (UX)" inserimos informações sobre o questionário de experiência do usuário (UX), que é utilizada para avaliar a satisfação, percepções e opiniões dos usuários em relação à interface e à experiência de uso de um software. Esse tipo de questionário é utilizado em testes de software para coletar *feedback* dos usuários sobre diversos aspectos, tais como facilidade de uso, clareza de informações, estética visual, utilidade do sistema, entre outros (Rubin and Chisnell 2008).

Para avaliar a experiência do usuário, existem vários questionários utilizados para avaliação de UX. A escolha do questionário depende do objetivo da avaliação, do público-alvo e da complexidade do produto ou sistema. Abaixo, será apresentado alguns dos questionários mais utilizados (Rubin and Chisnell 2008):

- **SUS (*System Usability Scale*)**: é um questionário que avalia a usabilidade do sistema, baseado em uma escala de 0 a 100. Ele consiste em 10 afirmações relacionadas à facilidade de uso do sistema, e os participantes devem indicar em uma escala de concordância de 5 pontos, de "Discordo totalmente" a "Concordo totalmente". O SUS é um dos questionários mais utilizados na avaliação de UX e é recomendado pela ISO.
- **SEQ (*Single Ease Question*)**: é um questionário com apenas uma pergunta, que avalia a facilidade de uso do sistema. A pergunta é geralmente algo como "O quão fácil foi utilizar o sistema?" e os participantes devem responder em uma escala de 7 pontos, de "Muito difícil" a "Muito fácil". O SEQ é utilizado em pesquisas rápidas de satisfação do usuário.
- **UEQ (*User Experience Questionnaire*)**: é um questionário que avalia a experiência geral do usuário com o sistema, considerando aspectos como atração visual, eficiência, confiabilidade e usabilidade. Ele consiste em 26 perguntas relacionadas a seis dimensões da experiência do usuário, e os participantes devem responder em uma escala de 7 pontos, de "Discordo totalmente" a "Concordo totalmente".

Esses são apenas alguns dos questionários mais utilizados na avaliação de UX. É importante lembrar que cada questionário tem suas particularidades e pode ser mais adequado para diferentes contextos e objetivos ([Rubin and Chisnell 2008](#)).

## 4.2 Resumo do Capítulo

Neste capítulo apresentamos o processo ABTUX detalhando as suas atividades, seus artefatos e de onde vieram as suas referências, apresentando informações do que será realizado em cada atividade e os atores envolvidos dentro do processo, seus papéis, responsabilidades e orientações para o uso do processo.

Em relação aos questionários de Usabilidade e UX, foi demonstrado um resumo das informações contidas nos documentos gerados para auxiliar o avaliador a aplicar o teste de usabilidade e UX. Apresentamos no artefato de usabilidade as heurísticas de Nilsen, et. al, com um exemplo de um questionário em que o avaliador pode se inspirar para a criação do seu formulário de teste de usabilidade, tanto para as questões objetivas quanto as questões subjetivas.

Após realização de testes em pares com engenheiros de software do processo ABTUX, foi observado a oportunidade de melhorar a qualidade dos artefatos e reduzir o fluxo de atividades. Diante desta percepção, foi proposto um refinamento do processo ABTUX e seus artefatos com o apoio de Especialistas em Engenharia de Software e Engenheiros de Software que pudessem avaliar o processo e assim contribuir com a sua melhoria. Este refinamento iremos apresentar no capítulo a seguir.

## 5 REFINAMENTO DO PROCESSO ABTUX

Para validarmos o processo ABTUX, foi executada uma simulação do processo com engenheiros de software para identificarmos possíveis pontos de melhoria, ajustes no processo e revisão dos artefatos. Neste capítulo iremos apresentar como o processo ocorreu e os resultados obtidos com este refinamento.

### 5.1 Processo ABTUX e seus artefatos - Versão 2.0

#### 5.1.1 Revisão dos artefatos

Ao todo, 4 engenheiros de software e 1 especialista em teste de software, foram selecionados para avaliar o novo processo proposto. Para evitar viés, foi apresentado o artigo científico (Fernandes et al. 2022) publicado na *International Conference on Human-Computer Interaction, HCII 2022* para entender o conceito e a aplicação do teste e suas referências bibliográficas. Realizada esta etapa de conhecimento da proposta do artigo, foi disponibilizado todo o processo ABTUX e os seus 25 artefatos para revisão. Após a avaliação dos engenheiros de software, foram levantados os seguintes comentários na tabela 17:

Tabela 17: Questionamento dos alunos de Engenharia de software novatos sobre os artefatos do processo ABTUX v1.0

Código	Artefato	Questionamento
ARTQ01	AT01 - Definindo o caso de uso para o teste A/B	Não especifica os cenários de teste?
ARTQ02	AT02 - Definindo a intenção do teste A/B	Como seria determinar essa intenção? Poderia ser definido no cenário de testes ?
ARTQ03	AT05 - Definir as variantes, métricas e interfaces a serem avaliadas	o que esse artefato significa? As métricas já não foram definidas? O que seriam as variantes?
ARTQ04	AT06 - Definir cenário de uso e propor instruções semi-estruturadas	Os cenários já foram propostos no primeiro artefato? As tarefas desses cenários não deveriam já serem especificadas no primeiro artefato?

Continuação na próxima página

Tabela 17: Questionamento dos alunos de Engenharia de software novatos sobre os artefatos do processo ABTUX v1.0 (Continuação)

ARTQ05	AT07 - Preparar o Experimento (avaliador)	As variantes não deveriam ser definidas na AT05? Esse artefato é realmente necessário? Caso queira realmente acrescentar essa parte, não poderia ser definido na etapa 'Definir métricas ferramentas e pontuações do fluxo?
ARTQ06	AT08 - Definir data de início e fim do teste e membros do teste A/B	Esses dados já poderiam ser utilizados apenas na especificação da versão atual para teste ou até ser unido ao teste do cenários que será realizado pelos participantes?
ARTQ07	AT08.02 - Formulário de Caracterização (Para Avaliador)	Esse artefato é realmente necessário? A coleta de dados não deveria ser apenas do usuário?
ARTQ08	AT09 - Critérios de Inclusão e exclusão de usuários	Os critérios poderiam ser definidos a partir do formulário?
ARTQ09	AT12 - Instruções e Orientações para realização do teste	É necessário? No AT13 a especificação é mais favorável.
ARTQ10	AT15 e 16 - Mapear o comportamento do usuário 'durante o teste	as instruções já não foram passadas antes? qual mapeamento está sendo realizado aqui? É realmente necessário?
ARTQ11	AT18 - Comunicar a conclusão do teste	O AT17 já especifica essa atividade, seria necessário?

Sobre o ARTQ01, o questionamento levantado se refere a falta de definição dos casos de teste, pois o artefato gerado a priori apenas detalhava o objetivo da avaliação, mas não especificava quais seriam os cenários que seriam testados ao decorrer do processo.

O ARTQ02 levantou dúvidas quanto a necessidade dele no processo, tendo em vista que ele definia a intenção de usabilidade do teste sem apresentar o que era desejado que o usuário realizasse. Dessa forma, também foi comentado a possibilidade de produzir os cenários de teste juntamente com a descrição da intenção do usuário na realização da avaliação.

Já em ARTQ03 houve o questionamento quanto ao próprio significado do AT05,

pelo motivo de existir um artefato anterior, o AT04, que já definia as métricas que seriam utilizadas no processo e, portanto, não haveria necessidade de outro artefato que também mostrasse essas métricas ou definisse outras métricas além das que já tinham sido selecionadas.

Quanto ao ARTQ04, o motivo dos questionamentos abordados na tabela é que os cenários deveriam ter sido definidos logo no primeiro artefato (AT01 - Definindo o caso de uso para o teste A/B), e, ao serem apresentadas as tarefas semi-estruturadas, não é detalhado o motivo de realizar cada uma destas tarefas na avaliação, devido não ter ocorrido a definição desses cenários de teste anteriormente.

O ARTQ05 foi levado em questão devido a problemas semelhantes aos anteriores, de já existir outro artefato, nesse caso seria o AT05, que teria o objetivo de definir os pontos do fluxo das atividades do processo. Sendo assim, o AT07 não demonstrou sua real necessidade, pois os dados presentes nele já poderiam ter sido colocados em um artefato anterior.

Sobre o ARTQ06, o questionamento foi em relação a possibilidade de uni-lo com outro artefato, um que apresentasse as instruções de realização do teste aos testadores, pois dessa forma resumiria todos os dados necessários para a realização da avaliação em um único documento.

O ARTQ07, aborda o AT08.02 que seria um formulário para o avaliador do processo, mas essas informações não seriam utilizadas no processo, já que o ponto principal seria coletar dados dos usuários que iriam participar dos testes dos sistemas.

O ARTQ08 demonstra que o artefato AT09 deveria falar dos critérios da escolha e eliminação de participantes, mas estava bem incompleto e faltava mais critérios que seriam melhor detalhados na parte do processo do formulário que seria repassado para o usuário que iria testar o sistema.

O ARTQ09, retrata sobre o AT12, que não pareceu estar bem completo e bem especificado como o AT13 que detalhava melhor a aplicação dos testes que seriam realizados nos sistemas através de um script pré-estabelecido.

O ARTQ10, fala sobre o AT15 e AT16 que, na primeira versão, apenas traz texto que comentam sobre o perfil do avaliador e as instruções que poderiam ajudar no mapeamento, que já haviam sido passadas nos artefatos iniciais.

O ARTQ11, especifica sobre o A18 da primeira versão, que é semelhante a AT17 que avalia o comportamento do usuário na execução de atividades com observações, tempo gasto em cada atividade, número de erros e se houve sucesso ou não.

Após os questionamentos e discussões gerados a partir dos questionamentos feitos sobre o uso desses artefatos no processo, foi possível notar a criação de uma quantidade

exacerbada de artefatos. Então, com a ajuda da equipe de engenheiros de software novatos, foi feita uma nova avaliação do processo inicialmente proposto, assim como uma remodelagem desse processo. Nessa perspectiva, a equipe gerou um conjunto de artefatos totalmente novo. Segue abaixo os artefatos gerados pela equipe:

- ATF01 - Definir os cenários de testes A/B;
- ATF02 - Disponibilizar versão atual do sistema;
- ATF03 - Definir as métricas e ferramentas;
- ATF05.1 - Definição das datas de início e fim dos testes;
- ATF05.2 - Termo de Consentimento;
- ATF05.3 - Captura de Dados Demográficos;
- ATF06 - Recrutar participantes e dividir das interfaces de testes entre os participantes;
- ATF07 - Passos da realização dos testes;
- ATF08 - Avaliação dos esforços do usuário nas diferentes interfaces;
- ATF08 - Avaliar esforço do usuário no teste através de questionário ;
- ATF09 - Questionário de usabilidade;
- ATF10.1 - Analisar resultados;
- ATF10.2 - Planilha de resultados.

O artefato ATF01 teria como objetivo comparar as funcionalidades de dois sistemas em prol de estabelecer quais são as funções presentes em ambos os sistemas e, por fim, definir os cenários de testes baseado no benchmarking dos requisitos dos dois sistemas. Dessa forma, será possível comparar as funcionalidades em termos de concorrência com o outro sistema.

Em ATF02, o objetivo é destacar as formas de acesso aos sistemas que serão testados, disponibilizando elas para a avaliação. Já o ATF03 define as métricas e as ferramentas que devem ser utilizadas para cada métrica proposta para o processo, de forma que haja quantificação e qualificação das respostas sobre o produto.

Quanto ao ATF04, ele foi elaborado com o intuito de estruturar as funcionalidades que serão testadas como tarefas que foram programadas de acordo com as funcionalidades que estão presentes em ambos os sistemas. Em seguida, o ATF05 deve definir a data de realização dos testes e especificar os dados dos sistemas a serem testados, de forma com que instruções para realização do teste sejam claramente estabelecidas.

Após isso, no ATF06 é proposto o recrutamento de participantes através da preenchimento de um questionário de caracterização, juntamente com assinatura de um termo de consentimento livre e esclarecido. Logo depois, o ATF07 deve estabelecer os critérios de exclusão de participantes que não atendam as especificidades do propósito da avaliação e mostrar como ocorre a divisão dos participantes entre as interfaces disponíveis para teste.

Em ATF08, o intuito estabelecido é o de permitir que o testador, em uma planilha, anote itens como tempo utilizado, se houve sucesso na realização da tarefa e observações como reclamações ou elogios espontâneos dos participantes ao decorrer da realização dos testes, o que permitirá a possibilidade de haver uma análise ainda mais detalhada ao final do processo.

Adiante, o artefato ATF09 propõe a criação questionário que deve ser utilizado para avaliar a experiência do usuário e usabilidade. Já o artefato ATF10.1 deve ser utilizados para analisar os resultados de forma qualitativa e escrita e do teste e da conclusão dos testes, enquanto no ATF10.2 é feita a análise quantitativa do teste, recomendando principalmente ser realizado em uma planilha, com o intuito de gerar gráficos de maneira mais eficaz.

Após essa proposta, na inspeção final dos artefatos, juntamente da equipe de engenheiros de software, houve um processo de comparação entre o primeiro conjunto de artefatos sugerido com o segundo, gerando uma discussão sobre a real necessidade de cada artefato proposto. Alguns pontos levantados durante essa etapa de comparação dos dois conjuntos de artefatos foram a falta de especificidade na utilização de alguns artefatos, a presença de artefatos que possuíam a mesma função e notou-se ser dispensável alguns artefatos que não estavam contribuindo para a realização do processo.

Após essa discussão, houve seleção dos artefatos, para geração da versão final. Essa seleção uniu os artefatos de melhor qualidade das duas propostas, realizando uma revisão com ajustes e junções dos mesmos. Por fim, gerou-se a versão final dos artefatos proposta para esse processo, que pode ser observada logo abaixo:

- ART01 - Definir os cenários de testes A/B;
- ART02 - Disponibilizar versão atual do sistema;
- ART03 - Definição de Métricas e ferramentas que serão utilizadas nos cenários de teste;
- ART04 - Definição das tarefas semi-estruturadas;
- ART05 - Definição das instruções e dados do teste;
- ART06 - Recrutamento dos participantes;

- ART07 - Divisão dos participantes entre as interfaces disponíveis;
- ART08 - Avaliação dos esforços do usuário nas diferentes interfaces;
- ART09 - Questionário para avaliar a experiência do usuário, usabilidade e design;
- ART10.1 - Análise dos resultados;
- ART10.2 - Planilha de resultados;
- ART11 - Relatório de decisão do software vencedor.

Como forma de detalhar o processo de criação da versão final dos artefatos do processo, a tabela abaixo foi feita para que deixe mais claro a união de alguns artefatos da primeira versão com outros artefatos sugeridos na segunda versão.

Tabela 18: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais

Versão 1	Versão 2	Versão Final	Descrição
AT03	ATF02	ART02	Houve a união da tabela proposta em AT03 com os passos para acessar o sistema de ATF02
AT04	ATF03	ART03	Houve união de métricas propostas por ambas as versões
AT13	ATF05.1	ART05	Foi feita a junção do conteúdo de AT13 com o conteúdo de ATF05.1
AT08.2, AT08.3		ART06	Houve a junção de dois artefatos propostos na versão 1
AT12	ATF06	ART07	Junção da ilustração usada em AT12 com as tabelas para organização da divisão de participantes de ATF06

Continuação na próxima página

Tabela 18: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

AT17	ATF08	ART08	Utilização da estrutura do AT17 e informações do ATF08
AT19, AT19.3		ART09	Ocorreu união dos questionários do AT19 e AT19.3

Entre as mudanças relevantes de cada versão final proposta, no ART02 foi proposta a união da tabela que apresenta os dados do sistema e a versão disponível para teste do AT03 com a descrição do passo a passo ensinando como utilizar essa versão de ATF02.

Quanto ao ART03, a união ocorreu entre a métrica de avaliação do feedback do usuário de AT04, que seria a utilização de uma escala de Likert, com as ferramentas de medição do esforço do usuário propostas em ATF03. Em seguida, no ART05 houve a junção das instruções de orientações para os testadores de AT13 com a definição de datas da realização do teste de ATF05.1, de forma que todos os dados e instruções do teste que serão necessários para os testadores ficassem disponíveis em um único documento.

No ART06 foi realizada a junção das propostas de dois artefatos da primeira versão sugerida, o AT08.2 e o AT08.3, que, em síntese, significa a união do formulário de preenchimento do usuário junto com o termo de consentimento para participação do teste em um único documento.

Em ART07, a mudança ocorreu com a junção da tabela de unificação dos dados dos participantes de ATF06 com a ilustração utilizada em AT12, que representa como ocorre o fluxo de divisão dos participantes, gerando, dessa forma, um artefato robusto que mostra como dividir os participantes juntamente com a divisão exata de usuários para cada versão do sistema.

Sobre o ART08, houve apenas algumas mudanças na estrutura da planilha que é utilizada pelo testador rastrear o comportamento do usuário durante o teste, de forma que os principais comportamentos do usuário fossem refletidos em dados para coleta nesse rastreamento, como tempo gasto na execução da tarefa, se houve a necessidade de ajuda online e outras observações que fossem identificadas pelo testador.

E por fim, em ART09 foi realizada a junção dos questionários propostos na primeira versão do conjunto de artefatos para um único questionário, de modo que simplificasse o preenchimento de dados dos usuários e facilitasse para os testadores a análise de dados, devido todos os dados de questionários estarem presentes em um único documento.

Como houveram muitas alterações ocorridas em cada um dos artefatos, acima foi apresentado apenas a formação do artefato ART02 como modo de exemplificar o processo realizado até chegar na versão final sugerida. Quanto aos artefatos ART01 e ART10.1, eles surgiram, respectivamente, dos artefatos ATF01 e ATF02, propostos no conjunto de artefatos da segunda versão, já o ART11 surgiu diretamente do artefato AT20, sugerido na primeira versão dos artefatos. No contexto dos artefatos que não refletiram na versão final, não foi apontada a real necessidade deles durante a realização do processo, portanto, para manter a organização e consistência da realização do teste, não houve utilidade deles.

A necessidade de cada um dos artefatos propostos na versão final é validada na execução prática do processo, onde foi demonstrada que a utilização desses artefatos foi benéfica e de grande valia para o bom andamento do processo de realização do teste A/B.

Em relação ao Revisão dos Questionários de Usabilidade e UX, após a unificação dos artefatos não houveram mudanças dentro dos questionários, a melhoria apenas unificou os documentos. No caso, o artefato AT19 e o AT19.3 foram unificados, se tornando um novo artefato ART09.

### 5.1.2 Ajustando o processo

Como os artefatos possuíam a maioria das informações de como executar o processo, após o seu refinamento foi possível mapear e compreender que o processo atual precisaria de ajustes, para poder ser simplificado. Para facilitar o entendimento das atividades e sua sequência, alteramos o código identificador (ID) da versão 1 para uma sigla na versão 2 do processo. Desta forma, temos as atividades padrões de siglas ACTVNN e OPCIONN, onde NN é um número inteiro que pode ir de 0 a 99, facilitando assim sua identificação dentro do processo e seu mapeamento.

Em relação as atividades, tivemos alterações nas atividades da versão 1: 8 - Recrutar participantes, 9 - Dividir os participantes e 10 Apresentar diretrizes aos participantes deixando todos estes itens agrupados na nova atividade, chamada de ACTIV07 - Definir critérios e divisão dos participantes na versão 2. A ilustração a seguir [4](#) mostra como ocorreu a formação o agrupamento das atividades no processo para gerarmos a versão 2 do processo ABTUX.

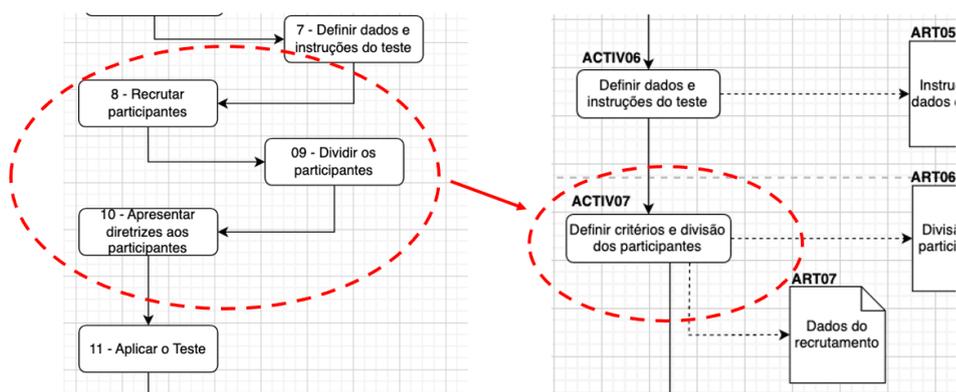


Figura 4 – Ilustração do DE-PARA do processo ABTUX (v1) e (v2 ACTIV07)

Seguindo a revisão, As atividades 11 - Aplicar o Teste, 12 - Monitorar os participantes, 13 - Criar versões de aplicações semi-estruturadas e 15 - Reportar conclusão dos testes, foram substituídas pela atividade ACTIV08 - Realizar os testes de cenário na versão 2, ilustradas abaixo.

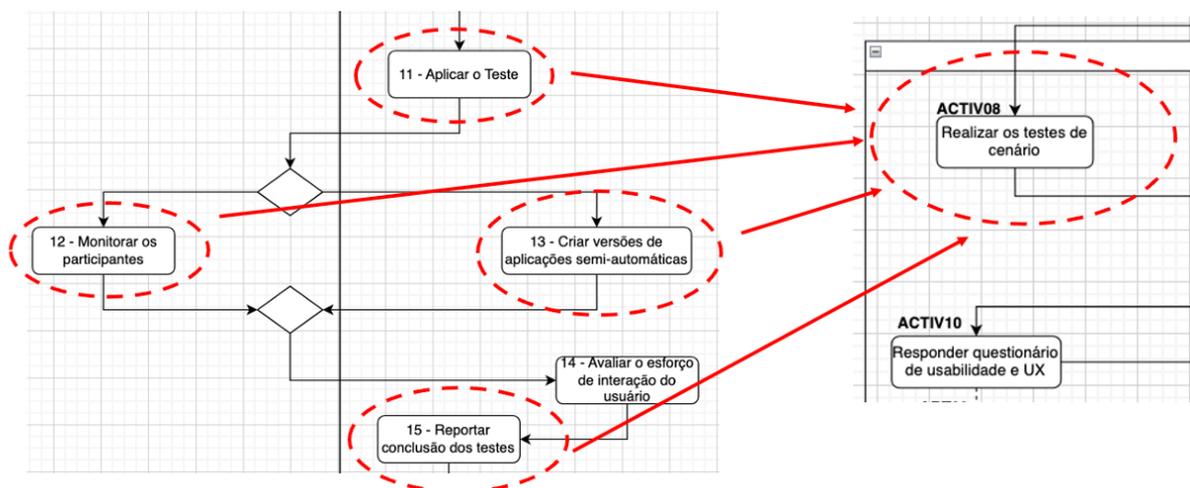


Figura 5 – Ilustração do DE-PARA do processo ABTUX (v1) e (v2 - ACTIV08)

Para melhor demonstrar esta revisão geral do fluxo do processo e suas atividades, montamos a tabela 19 para facilitar a comparação do processo na versão 1 em relação a versão 2 do processo ABTUX.

Tabela 19: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais

IDv1	Versão 1	IDv2	Versão 2	Mudança no Processo
1	Selecionar casos de testes	ACTIV01	Selecionar casos de testes	Não houve alterações
2	Determinar a Intenção do Usuário:	ACTIV02	Determinar a Intenção do Usuário:	Não houve alterações
3	Disponibilizar versão atual	ACTIV03	Disponibilizar versão atual	Não houve alterações
4	Criar versão da aplicação	OPCIO01	Criar versão da aplicação	Não houve alterações
5	Definir métricas pontuações e ferramentas	ACTIV04	Definir métricas pontuações e ferramentas	Não houve alterações
6	Definir Tarefas semi-estruturadas	ACTIV05	Definir Tarefas semi-estruturadas	Não houve alterações
7	Definir dados e instruções de teste	ACTIV06	Definir dados e instruções do teste	Não houve alterações
8	Recrutar participantes	ACTIV07	Definir critérios e divisão dos participantes	Agrupamento das atividades 8, 9 e 10
9	Dividir os participantes			
10	Apresentar diretrizes aos participantes			
11	Aplicar o Teste	ACTIV08	Realizar os testes de cenário	Agrupamento das atividades 11, 12, 13 e 15 da versão 1
12	Monitorar os participantes			

Continuação na próxima página

Tabela 19: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

13	Criar versões de aplicações semi-automáticas			
14	Avaliar o esforço de interação do usuário	ACTIV09	Avaliar o esforço de interação do usuário	Não houve alterações
15	Reportar conclusão dos testes	ACTIV08	Realizar os testes de cenário	Agrupamento das atividades 11, 12, 13 e 15 da versão 1
16	Responder questionário de usabilidade e UX	ACTIV10	Responder questionário de usabilidade e UX	Não houve alterações
17	Utilizar <i>machine learning</i> para análise de resultados	OPCIO02	Utilizar <i>machine learning</i> para análise de resultados	Não houve alterações
18	Analisar resultados	ACTIV11	Analisar resultados	Não houve alterações
19	Decidir melhor solução avaliada	ACTIV12	Decidir melhor solução avaliada	Não houve alterações

Após a conclusão da revisão do processo, obtivemos uma nova visão e agrupamento de atividades e artefatos, principalmente com a redução do número de artefatos foi possível apresentar dentro do diagrama qual o artefato é utilizado em cada etapa do processo, bem como as fases do processo: (a) Iniciação; (2) Planejamento; (3) Execução e (4) Resultado. Com esta melhoria no fluxo, temos os seguintes passos atualizados na imagem abaixo:

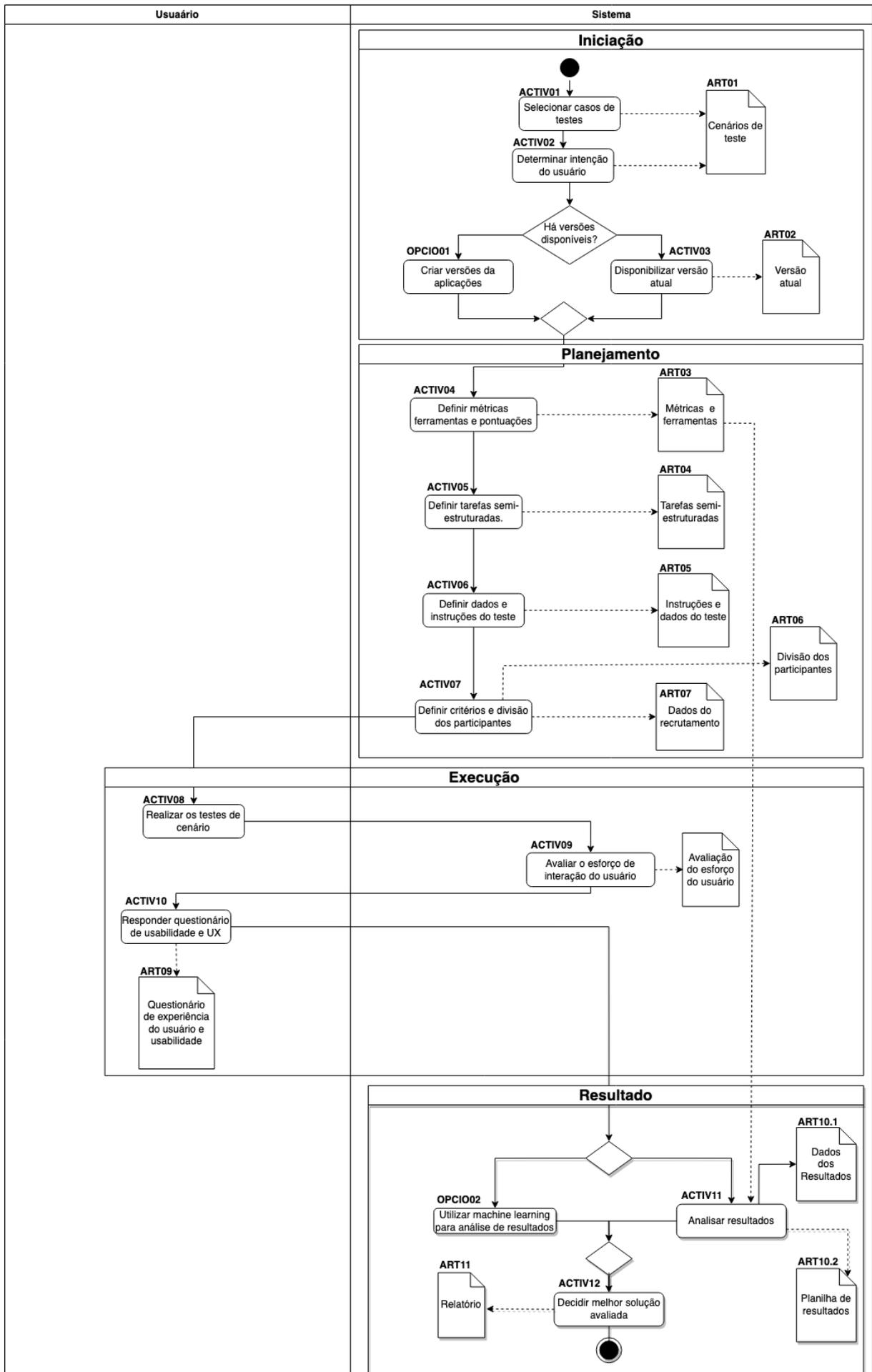


Figura 6 – Macro Fluxo da versão 2.0 do processo ABTUX

Após todas essas etapas seguidas para a estruturação e melhora da qualidade dos artefatos para a realização eficaz do teste, é apresentada a tabela que mostra todas as etapas do fluxo de atividades, assim como se há a necessidade de artefatos e quais são os artefatos necessários para cada uma dessas etapas. A tabela também conta uma descrição objetivo do motivo de existir essa etapa do teste, de forma que torne claro para qualquer testador que aderir esse processo o que deve ser realizado em cada uma das fases do processo ABTUX.

Tabela 20: ABTUX 2.0: Lista de atividades e artefatos do novo processo

ID	Atividade	Descrição da atividade	Artefato vinculado
ACTIV01	Selecionar casos de testes	Essa atividade definir os sistemas computacionais que queremos realizar o teste	ART01 - Definir os cenários de testes A/B
ACTIV02	Determinar a Intenção do Usuário:	Definir as funcionalidades que iremos testar nos sistemas computacionais selecionados	
ACTIV03	Disponibilizar versão atual	Neste item os sistemas a serem comparados são existentes, logo você deve informar qual a versão dos sistemas computacionais farão parte do teste	ART02 - Disponibilizar versão atual do sistema
OPCIO01	Criar versão da aplicação	Cria uma nova versão da aplicação que iremos comparar (variação) que fará parte do teste	Não de aplica

Continuação na próxima página

Tabela 20: ABTUX 2.0: Lista de atividades e artefatos do novo processo (Continuação)

ACTIV04	Definir métricas pontuações e ferramentas	Atividade em que iremos definir como iremos realizar as ferramentas e as medições de tarefas a serem realizadas no teste (tempo, pontuações, etc)	ART03 - Definição de Métricas e ferramentas que serão utilizadas nos cenários de teste
ACTIV05	Definir Tarefas semi-estruturadas	Nesta atividade iremos descrever a execução das tarefas programadas de acordo com as funcionalidades que estão presentes em ambos os sistemas	ART04 - Definição das tarefas semi-estruturadas
ACTIV06	Definir dados e instruções de teste	Define a data, local a forma de realização dos testes, e define as instruções para realização do teste	ART05 - Definição das instruções e dados do teste

Continuação na próxima página

Tabela 20: ABTUX 2.0: Lista de atividades e artefatos do novo processo (Continuação)

ACTIV07	Definir critérios e divisão dos participantes	Nesta atividade deverá ser necessário avaliar se os grupos e pessoas possuem algum critério que possa viciar ou prejudicar a avaliação da funcionalidade do sistema computacional. Separa os participantes, conforme premissas e estabelecida a validação do Teste A/B, para que a avaliação esteja balanceada. Apresentar o resumo de como o teste será aplicado e o que os participantes deverão executar	ART06 - Recrutamento dos participantes ART07 - Divisão dos participantes entre as interfaces disponíveis
ACTIV08	Realizar os testes de cenário	Durante a execução do teste é necessário avaliar e acompanhar as ações realizadas pelos participantes, suas dificuldades e dúvidas durante o teste	AT16 - Avaliar os esforços das diferentes interfaces

Continuação na próxima página

Tabela 20: ABTUX 2.0: Lista de atividades e artefatos do novo processo (Continuação)

ACTIV09	Avaliar o esforço de interação do usuário	Acompanhar os tempos de execução dos testes, por tarefa	ART08 - Avaliação dos esforços do usuário nas diferentes interfaces
ACTIV10	Responder questionário de usabilidade e UX	Após a execução das tarefas, os testadores precisarão responder ao questionário para coleta dos dados do teste aplicado. Este ponto é muito importante para o nosso processo	ART09 - Questionário para avaliar a experiência do usuário, usabilidade e design
OPCIO02	Utilizar <i>machine learning</i> para análise de resultados	Em nosso processo, deixamos como opcional o uso de ferramentas automatizadas para análise da usabilidade e experiência do usuário (UX), pois alguns testes já possuem ferramentas próprias para acelerar e facilitar a coleta e análise dos dados	N/A

Continuação na próxima página

Tabela 20: ABTUX 2.0: Lista de atividades e artefatos do novo processo (Continuação)

ACTIV11	Analisar resultados	Apresentar os resultados e divulgar qual foi o melhor entre as funcionalidades e nos sistemas computacionais em geral	ART10.1 - Análise dos resultados ART10.2 - Planilha de resultados
ACTIV12	Decidir melhor solução avaliada	Apresentar os resultados e divulgar qual foi o melhor entre as funcionalidades e nos sistemas computacionais em geral	ART11 - Relatório de decisão do software vencedor

## 5.2 Resumo do Capítulo

Após a participação dos engenheiros de software novatos, a qualidade do controle e orientação do processo e seus artefatos ficaram bem enxutos e mais simples de serem aplicados. A revisão dos artefatos com a substituição por novos ou unificação e alterações em artefatos existentes fez com que o processo ABTUX esteja pronto para ser aplicado e validado em campo para confirmar se o processo montado irá atender a questão da pesquisa. No capítulo a seguir, iremos apresentar como realizamos este teste, após todas as melhorias aplicadas.

## 6 APLICAÇÃO PRÁTICA DO PROCESSO ABTUX V2.0

O processo ABTUX foi disponibilizado para uma equipe de quatro engenheiros de software aplicarem o teste. Esta equipe assumiu o papel de avaliador para poder realizar todas as atividades do processo.

### 6.1 Objetivo da Aplicação

A aplicação tem como objetivo validar se o novo processo possui os requisitos necessários para responder a questão de pesquisa.

#### 6.1.1 Executando o processo ABTUX versão 2.0

O teste se inicia com a seleção dos casos de testes (ACTIV01), a determinação da intenção do usuário (ACTIV02) e a disponibilização da versão atual do sistema para o teste (ACTIV03). Para responder as atividades selecionamos duas plataformas educacionais: (1) Udemy; e (2) Coursera. Estas plataformas foram escolhidas aleatoriamente, porém, observando que são plataformas concorrentes, que possuem as mesmas funcionalidades, mas tratadas de acordo com sua plataforma, de modo a podermos comparar a usabilidade e UX de cada uma delas.

Após essa fase inicial, ocorre o planejamento do teste, que começa com a definição de métricas e ferramentas que serão utilizadas ao decorrer do processo (ACTIV04). Para atender esta atividade iremos utilizar estratégias de pesquisa quantitativa e qualitativa, através de questionários com atributos que avaliam a usabilidade e a UX, já a métrica para extração do dados será através da escala de Likert e medição do tempo para realizar cada tarefa. Os participantes responderão os questionários os valores da escala Likert para cada tarefa realizada no Coursera e Udemy, conforme tabela abaixo.

Tabela 21: Escala de Likert, adaptado pelo autor

Código	Componente da Escala	Valor Associado
DT	Discordo Totalmente	1
DD	Discordo	2
DP	Discordo Parcialmente	3

Continuação na próxima página

Tabela 21: Escala de Liket, adaptado pelo autor (Continuação)

NN	Nem concordo e nem discordo (Indiferente ou Neutro)	4
CP	Concordo Parcialmente	5
CC	Concordo	6
CT	Concordo totalmente	7

Os resultados obtidos com os respondentes serão utilizados para medir os níveis de satisfação.

A seguir, devemos realizar a atividade de definição das tarefas semi-estruturadas (ACTIV05) que serão realizadas pelos participantes do teste. Para este teste, selecionamos as tarefas abaixo para serem executada pelos testadores:

Tabela 22: Lista das tarefas semi-estruturadas e o tempo de execução, criado pelo autor

<b>Código</b>	<b>Atividade</b>	<b>Tempo</b>
T01	Acessar o sistema pelo navegador e realizar seu cadastro/login na plataforma (udemy/coursera)	5 minutos
T02	Após realizar o acesso ao sistema, clicar no campo de busca, para procurar por cursos de diferentes formas como autor, descrição, nome do conteúdo e etc	8 minutos
T03	O usuário logado deve adquirir um conteúdo gratuito e clicar em um dos seus vídeo do curso, e utilizar as ferramentas de legenda e aceleração do vídeo	8 minutos
T04	O usuário deve localizar um conteúdo recomendado pelo sistema e acessá-lo	5 minuto
T05	O usuário deve identificar o seu progresso no consumo de determinado conteúdo	5 minutos
T06	O usuário logado deve entrar no conteúdo que deseja avaliar ou comentar, digitar sua avaliação e enviar	8 minutos
T07	O usuário deve realizar o compartilhamento de um curso	8 minutos

Continuação na próxima página

Tabela 22: Lista das tarefas semi-estruturadas e o tempo de execução, criado pelo autor (Continuação)

T08	O usuário deve clicar no perfil, editar seus dados e salvá-los	8 minutos
T09	O usuário deve acessar o campo de central de ajuda e buscar por uma dúvida	8 minutos

O avaliador deverá repassar a definição dos dados e instruções do teste (ACTIV06) aos participantes e concluir com a definição de critérios para recrutamento dos participantes e divisão dos participantes entre as duas interfaces disponíveis para realização do teste (ACTIV07).

Considerando que os participantes podem ter origens diferentes que podem causar efeitos indesejados nos resultados, fornecemos um questionário pré-teste para caracterizar cada participante e se todos estão atendendo os critérios de participação. Este questionário contém perguntas sobre: (i) familiaridade com os softwares educacionais, (ii) Idade (iii) sexo (iv) escolaridade (v) nível de conhecimento sobre usabilidade/UX (vi) profissão. Essa informações foram usadas para determinar o nível de habilidade e conhecimento dos usuários. Dos entrevistados, 63% eram do sexo masculino e 38% do sexo feminino. As idades dos participantes estavam na faixa de 18 a 28 anos. Dos entrevistados seis são estudantes de graduação (Ensino Superior incompleto) e dois já possuem graduação (Ensino superior completo).

Tabela 23: Informações dos participantes do Teste ABTUX v2.0

Plataforma	Participantes	Idade
Udemy	(3) Ensino superior Incompleto (1) Ensino superior Completo	21-28
Coursera	(3) Ensino superior Incompleto (1) Ensino superior Completo	21-24

Após a escolha dos testadores dos sistemas, apresentamos a estes testadores um resumo do estudo que estão participando, em seguida foram assinados os termos de consentimento para que em seguida elas sejam divididas de forma equilibrada, sendo que para esta aplicação de teste, quatro pessoas em dois grupos para testarem o Coursera e quatro pessoas para testarem o Udemy.

Depois de planejado, o teste pode ser executado no âmbito das interfaces escolhidas. Como se trata de um teste A/B, foram escolhidas duas interfaces concorrentes para serem

colocadas para teste e ser decidido, ao final da análise dos resultados, qual interface apresenta o melhor *feedback* do usuário. Durante a realização dos testes, é realizada a atividade de realização dos cenários de teste pelos participantes (ACTIV08), em que foi solicitado aos participantes que preenchessem um formulário com o intuito de avaliar as plataformas usadas, através das seguintes perguntas abaixo:

Tabela 24: Questões de usabilidade e UX, criado pelo autor

Código	Pergunta
QT01	Na plataforma, foi fácil localizar o curso?
QT02	Você teve facilidade de se matricular no curso?
QT03	Você encontrou alguma facilidade em controlar o vídeo , como interrupções do vídeo, facilidade em visualizar a legenda na plataforma?
QT04	Ao carregar ou atualizar os dados do seu perfil, a plataforma é rápida e eficiente?
QT05	A plataforma proporcionou uma melhor experiência de compartilhamento enquanto compartilhava o curso com outras pessoas?
QT06	A plataforma tem uma seção de Ajuda com a melhor documentação?
QT07	A plataforma mostrou uma orientação clara enquanto você está trabalhando, como exibir mensagens de erro enquanto um passo errado é dado?
QT08	Ao considerar a experiência geral durante a execução das tarefas, a plataforma proporciona a melhor experiência do usuário?
QT09	O site é original
QT10	O site possui poluição visual?
QT11	O site é sofisticado?
QT12	O site é confuso?
QT13	O site é criativo?
QT14	O site faz bom uso de efeitos especiais?
QT15	O site é agradável?
QT16	O site é simétrico?
QT17	O site é bonito?
QT18	O site é fácil de navegar?

Enquanto o teste está sendo executado, o avaliador precisa acompanhar o participante para a avaliação do esforço de interação do usuário pelo testador (ACTIV09).

Aqui o avaliador irá obter informações como: (a) tempo de execução da tarefa, (b) se houve necessidade de ajuda por parte do testador, (c) observações adicionais durante o teste, seguido pela resposta dos usuários ao questionário de usabilidade e UX (ACTIV10).

Sobre o questionário iremos avaliar cada pergunta sobre as heurísticas de usabilidade de Nilsen e assim identificar erros ou problemas durante o teste sobre cada uma das questões. A seguir, é detalhado a relação de cada heurística com as questões da tabela 25, conforme (Nielsen 1994):

Tabela 25: Heurísticas de usabilidade x Questões associadas

Heurísticas de Usabilidade	Questões
a. Visibilidade do status do sistema	QT05
b. Concordância do sistema com o mundo real	QT01;
c. Controle do usuário e liberdade	QT04
d. Consistência e padrões	QT01;QT03
e. Prevenção de erros	QT12
f. Reconhecer ao invés de lembrar	QT03
g. Flexibilidade e eficiência de uso	QT02;QT08
h. Estética e design minimalista	QT03
i. Ajudar os usuários a reconhecer, diagnosticar e corrigir erros	QT07
j. Ajuda e documentação	QT6;QT7

- **a. Visibilidade do status do sistema:** A questão **QT05** refere-se à necessidade de fornecer *feedback* aos usuários sobre o status do sistema, por exemplo, informar quando uma operação está sendo executada ou quando uma tarefa foi concluída. Essa questão está relacionada à heurística de visibilidade do status do sistema, que diz que o sistema deve manter os usuários informados sobre o que está acontecendo.
- **b. Concordância do sistema com o mundo real:** A questão **QT01** refere-se à necessidade de tornar o sistema coerente com o mundo real, ou seja, utilizar linguagem e convenções familiares aos usuários. Essa questão está relacionada à heurística de concordância do sistema com o mundo real, que diz que o sistema deve falar a linguagem dos usuários.
- **c. Controle do usuário e liberdade:** A questão **QT04** refere-se à necessidade de fornecer opções de desfazer e refazer para permitir que os usuários possam corrigir seus erros. Essa questão está relacionada à heurística de controle do usuário e liberdade, que diz que o sistema deve permitir que os usuários controlem o que acontece e tenham liberdade para desfazer ações indesejadas.

- **d. Consistência e padrões:** As questões **QT01** e **QT03** referem-se à necessidade de utilizar padrões de interface e manter a consistência entre as diferentes partes do sistema. Essas questões estão relacionadas à heurística de consistência e padrões, que diz que o sistema deve ser consistente e seguir padrões de interface reconhecidos pelos usuários.
- **e. Prevenção de erros:** A questão **QT12** refere-se à necessidade de prevenir erros por meio do design de interfaces seguras e da inclusão de avisos de confirmação. Essa questão está relacionada à heurística de prevenção de erros, que diz que o sistema deve ajudar os usuários a evitar erros e a recuperar-se deles.
- **f. Reconhecer ao invés de lembrar:** A questão **QT03** refere-se à necessidade de tornar a interface intuitiva e fácil de usar, sem exigir que os usuários memorizem informações. Essa questão está relacionada à heurística de reconhecer ao invés de lembrar, que diz que o sistema deve ser fácil de usar e não exigir que os usuários memorizem informações.
- **g. Flexibilidade e eficiência de uso:** As questões **QT02** e **QT08** referem-se à necessidade de permitir que usuários experientes possam utilizar atalhos e aceleradores, além de tornar o sistema flexível o suficiente para ser utilizado de diferentes maneiras. Essas questões estão relacionadas à heurística de flexibilidade e eficiência de uso, que diz que o sistema deve ser flexível e permitir que os usuários realizem tarefas de diferentes maneiras.
- **h. Estética e design minimalista:** A questão **QT03** refere-se à necessidade de tornar a interface visualmente agradável e fácil de usar, sem sobrecarregar os usuários com informações.
- **i. Ajudar os usuários a reconhecer, diagnosticar e corrigir erros:** A questão **QT07** refere-se à necessidade de ajudar os usuários a lidar com erros por meio de mensagens claras e informativas. Essa questão está relacionada à heurística de ajudar os usuários a reconhecer, diagnosticar e corrigir erros, que diz que o sistema deve fornecer informações úteis e orientações claras para ajudar os usuários a lidar com erros.
- **j. Ajuda e documentação:** As questões **QT6** e **QT7** referem-se à necessidade de fornecer ajuda e documentação para orientar os usuários a utilizar o sistema. Essas questões estão relacionadas à heurística de ajuda e documentação, que diz que o sistema deve fornecer informações claras e úteis para orientar os usuários a utilizar o sistema e resolver problemas.

Seguindo o mesmo padrão, criamos a tabela 26 com a relação dos atributos de UX existentes em cada questão pra podermos avaliar a experiência do usuário. a explicação de cada item será apresentada em seguida:

Tabela 26: Atributos de UX x Questões associadas

Item	Atributo de Qualidade	Questões
a	Atratividade	QT10;QT14;QT15;QT17
b	Perspividade	QT12
c	Eficiência	QT17
d	Confiabilidade	QT16
e	Estimulação	QT18
f	Novidade	QT09;QT13

- **a. Atratividade:** As questões **QT10**, **QT14**, **QT15** e **QT17** estão relacionadas à atração que o sistema exerce sobre os usuários, como sua aparência visual e a capacidade de despertar emoções positivas. A atratividade é um dos atributos mais importantes de UX, pois o usuário deve ser atraído pelo sistema para se engajar e utilizá-lo com frequência.
- **b. Perspicuidade:** A questão **QT12** refere-se à clareza e simplicidade do sistema, para que o usuário possa utilizá-lo sem confusão ou dificuldade. A perspicuidade está relacionada ao atributo de qualidade de UX de facilitação do uso, que envolve tornar a utilização do sistema fácil e intuitiva.
- **c. Eficiência:** A questão **QT17** refere-se à rapidez e eficiência com que o usuário pode executar as tarefas no sistema. A eficiência é um atributo importante da UX, pois os usuários desejam concluir as tarefas de forma rápida e sem esforço excessivo.
- **d. Confiabilidade:** A questão **QT16** refere-se à consistência e confiabilidade do sistema, garantindo que o sistema funcione corretamente em todas as situações. A confiabilidade é um atributo importante da UX, pois os usuários desejam que o sistema funcione corretamente e sem erros.
- **e. Estimulação:** A questão **QT18** refere-se ao estímulo emocional que o sistema proporciona ao usuário, como entretenimento ou diversão. A estimulação é um atributo importante de UX, pois pode motivar os usuários a continuar utilizando o sistema.
- **f. Novidade:** As questões **QT09** e **QT13** referem-se à novidade e originalidade do sistema, que pode ser um fator de atração para os usuários. A novidade é um

atributo importante de *UX*, pois os usuários desejam utilizar sistemas inovadores e que ofereçam algo diferente do que já estão acostumados.

Ao final do processo ABTUX é feita a análise dos resultados obtidos pelos questionários e observações do testador durante o teste (ACTIV11), assim como é gerado um relatório informando a conclusão e apresentando os resultados do teste juntamente com a melhor interface segundo a avaliação (ACTIV12). Para estas duas últimas atividades do processo ABTUX, iremos apresentar no próximo capítulo.

## 6.2 Análise dos Resultados

Esta seção contém os resultados da pesquisa realizada com oito usuários que executaram o processo de teste de usabilidade e UX através de um Teste A/B. Através da análise destes resultados coletamos informações sobre os atributos de qualidade eficácia e eficiência, o tempo que cada usuário leva para concluir uma tarefa, o nível de satisfação dos usuários e a divulgação do software que teve a melhor avaliação, baseados nos atributos de qualidade.

### 6.2.1 Perguntas objetivas de Usabilidade e UX:

Após a conclusão dos testes, as respostas dos usuários foram distribuídas na tabela abaixo, onde é possível observar a quantidade de pessoas que marcou cada uma das alternativas dentro da escala de Likert assim como o total de marcações que cada uma das componentes da escala recebeu no âmbito geral do questionário.

Tabela 27: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais

Questionário de Usabilidade														
	Coursera							Udemy						
QQ	DT	DD	DP	NN	CP	CC	CT	DT	DD	DP	NN	CP	CC	CT
QT01	-	-	-	-	1	-	3	-	-	-	-	1	1	2
QT02	-	-	-	-	-	-	4	-	-	-	-	-	-	4
QT03	-	-	-	1	-	1	2	-	-	-	-	-	-	4
QT04	-	-	-	-	1	-	3	-	-	-	-	1	2	1
QT05	-	-	-	2	-	1	1	-	-	-	1	1	1	1
QT06	-	1	-	-	1	2	-	-	-	-	-	2	-	2

Continuação na próxima página

Tabela 27: Proposta A/B testando processo por Usabilidade e avaliação em UX - Atividades opcionais (Continuação)

QT07	-	-	-	2	-	2	-	1	-	-	2	-	-	1
QT08	-	-	-	-	1	2	1	-	-	-	-	1	2	1
QT09	-	1	-	1	-	2	-	-	-	1	-	1	1	1
QT10	-	2	1	1	-	-	-	-	-	-	1	-	1	2
QT11	-	-	-	-	1	1	2	1	-	-	-	1	2	-
QT12	1	1	-	-	1	1	-	1	2	-	-	-	-	1
QT13	-	2	-	-	1	1	-	-	1	-	2	1	-	-
QT14	2	1	-	-	-	1	-	1	1	1	-	-	1	-
QT15	-	-	-	-	2	1	1	1	-	1	-	1	-	1
QT16	-	-	-	1	-	3	-	1	-	-	-	1	1	1
QT17	-	-	-	1	-	1	2	-	1	-	1	1	-	1
QT18	-	-	-	-	-	4	-	-	-	-	1	-	1	2
Total	3	8	1	9	9	23	19	6	5	3	8	12	13	25

Legenda: QQ - Questões; DT - Discordo Totalmente; DD - Discordo DP - Discordo Parcialmente NN - Nem concordo nem Discordo (indiferente ou neutro); CC - Concordo Parcialmente CC - Concordo CT - Concordo Totalmente.

### 6.2.2 Análise dos resultados das Questões objetivas

A pergunta QT01 da pesquisa foi mapeada para reconhecimento em vez de recordação, consistência e padrão. O reconhecimento e a recuperação são definidos como se o usuário desejasse pesquisar um curso na barra de pesquisa; o usuário não deve ter que se lembrar de cada palavra da pesquisa que faz e o Coursera e Udemy têm a funcionalidade de lembrar os resultados pesquisados anteriormente. O resultado das respostas demonstram que do grupo Coursera, 75% não tiveram problemas ao utilizar esta funcionalidade, porém 25% encontrou alguma dificuldade para para o uso desta facilidade. Já na Udemy 50% tiveram sucesso na tarefa, quanto 25% encontrou alguma dificuldade e outros 25% sentiram a ausência de algo na funcionalidade.

A pergunta QT02 da pesquisa foi mapeada para corresponder entre o sistema e o mundo real. O usuário é solicitado a se inscrever no curso especificado e assistir aos vídeos introdutórios. A interface das plataformas Coursera e Udemy deve corresponder entre seu sistema e o mundo real, pois a parte introdutória deve ser separada do restante dos vídeos. A Udemy e o Coursera acompanham o esta heurísticas entre o sistema e o mundo real

por meio de vídeos de introdução durante a inscrição no curso. O resultado foi que todos os testadores do grupo Udemy e Coursera conseguiram observar esta funcionalidade atendeu plenamente as expectativas, com 100% de resultado positivo.

A pergunta QT03 da pesquisa foi mapeada para Controle e liberdade do usuário. Às vezes, os usuários podem cometer erros e navegar para uma página desnecessária e precisam voltar à página desejada, onde o controle e a liberdade do usuário desempenham um papel crucial. Um site deve ter sinalizações claras de onde o usuário está e por onde ele pode navegar. Pelas respostas apresentadas, demonstra-se que na Udemy 100% dos testadores conseguiram realizar a tarefa plenamente e não se perderam na navegação, porém a Coursera 50% dos testadores conseguiram realizar plenamente a tarefa, enquanto 25% alguma dificuldade de navegação e outros 25% permaneceram com uma opinião neutra sobre esta navegação.

A pergunta QT04 da pesquisa foi mapeada para a Visibilidade do status do sistema. A tarefa do usuário é carregar/atualizar a foto do perfil dentro de suas respectivas plataformas. Ao final de uma tarefa, o usuário precisa ser notificado sobre a atualização da foto do perfil. Neste caso há feedback em ambas as plataformas (Coursera e Udemy), pois a foto de perfil dessas plataformas está incluída na seção de perfil, e para salvar a foto o usuário precisa salvar toda a seção de perfil. Nas respostas, reflete que a Coursera tivemos 75% dos testadores sem qualquer dificuldade para realizar a tarefa e 25% teve alguma dificuldade. Já a Udemy, apenas 25% dos testadores conseguiram executar a tarefa sem barreiras, porém 50% teve um pouco de dificuldade de realizar a tarefa e outros 25% tiveram uma maior dificuldade.

A pergunta QT05 da pesquisa foi mapeada para Flexibilidade e eficiência de uso. Aqui a tarefa do usuário é compartilhar o curso nas duas plataformas. Conhecer a interação do usuário durante o compartilhamento do curso permite ao usuário um uso eficiente com outras pessoas em redes sociais (WhatsApp, Twitter, Facebook). Para ambas as plataformas (Udemy e Coursera), apenas 25% dos testadores conseguiram realizar a tarefa sem dificuldades, já para os demais testadores temos que na Udemy 25% ficaram neutros na resposta, 25% tiveram um pouco de dificuldade em realizar a tarefa. Já na Coursera 50% ficaram neutros e 25% tiveram um pouco de dificuldade para executar a tarefa.

A pergunta QT06 da pesquisa está relacionada a ajuda e documentação. Na maioria das vezes, os usuários não precisam de ajuda para executar as suas tarefas. No entanto, a documentação é necessária, caso algum problema ocorra e o usuário precisar de orientação para resolver. Pode ser necessário fornecer uma documentação mais extensa para resolver alguma dúvida do usuário, logo é muito importante que esta ajuda tenha uma fácil navegação para que o usuário tenha segurança em como realizar a tarefa. Na plataforma Udemy 50% conseguiram realizar a tarefa sem dificuldades, porém outros 50% tiveram uma dificuldade maior em realizar a tarefa. Já a Coursera 25% não conseguiram localizar o

acesso a documentação, 50% conseguiu localizar a documentação com pouca dificuldade e 25% com dificuldade um pouco maior.

A pergunta QT07 da pesquisa foi mapeada para ajudar os usuários a reconhecer, diagnosticar e corrigir erros. Para qualquer usuário, se ocorrer algum erro, uma mensagem de erro deve aparecer para que o usuário possa corrigir o erro e continuar executando as tarefas com a indicação de como resolver a questão. Este item é bastante utilizado no preenchimento de formulários, como por exemplo validação de e-mail e seleção de País ou Estado. Ambas as plataformas possuem esta funcionalidade. Nas plataformas Coursera e Udemy o grupo de 50% não percebeu erros na plataforma e ficaram em posição de neutralidade, já na Coursera, outros 50% identificaram erros na plataforma mas tiveram um pouco de dificuldade para identificar e fazer a correção ou sair do erro. Já na Udemy, 25% conseguiram tratar o erro e seguir com a tarefa e outros 25% identificaram um erro mas não conseguiram perceber pela plataforma como realizar o tratamento.

A pergunta QT08 da pesquisa é a experiência geral do usuário no Coursera e Udemy. Neste quesito tivemos um resultado espelho entre ambas as plataformas Udemy e Coursera, sendo que em ambas 25% tiveram uma excelente experiência, 50% foi uma experiência muito boa e 25% tiveram uma experiência boa.

Em relação a perspectiva de avaliação dos atributos de qualidade, as questões QT09 a QT08 estão mais voltadas para o sentimento e percepção do testadores. Então observando a atratividade (AQ1) as questões QT10, QT14, QT15 e QT17 foram respondidas, obtendo as seguintes informações:

Os testadores do site Coursera acharam o site bem Atrativo, contra 58% que acharam o site pouco atrativo. Já na Udemy 50% deles acharam o site atrativo, ficando uma avaliação bem dividida. Sobre o atributo Perspicuidade, na Coursera 50% dos testadores acham a plataforma claro e compreensível, já na Udemy 75% dos usuários acharam o site muito pouco compreensível e apenas 25% achou o site claro e compreensível.

No quesito Eficiência (AQ3), na plataforma Coursera, 75% dos testadores entendem que o site é eficiente contra 25% que não possuem a mesma opinião, já a Udemy temos 50% dos resultados como um site eficiente. Para o atributo Confiabilidade (AQ04), obtivemos 75% de avaliação positiva no Coursera e na Udemy. O atributo Estimulação (AQ05) a coursera em unanimidade ficaram animados e acharam o site divertido de usar, contra 75% da Udemy.

Por fim, o item Novidade (AQ06) que avalia se o design do produto é criativo e atrai o interesse dos usuários, o qual tivemos o resultado de 75% na Udemy, contra 50% na Coursera.

Em geral, as avaliações indicam que há espaço para melhorias nas plataformas Coursera e Udemy em relação à usabilidade, especialmente nas heurísticas relacionadas

a ajuda ao usuário, reconhecimento de informações importantes e correção de erros. As heurísticas "Consistência e padrões" e "Controle do usuário e liberdade" foram as mais bem avaliadas, o que sugere que são áreas em que as plataformas estão melhorando e devem continuar a investir para melhorar a experiência do usuário.

Apresentamos a seguir os resultados quantitativos, baseado em todas as respostas coletadas pelos usuários na tabela 27.

As questões de QT01 até QT08 fazem parte do questionário de usabilidade das heurísticas de Nielsen, enquanto as questões de QT09 até QT18 estão mais voltadas para avaliações subjetivas, vinculadas a experiência de uso. Baseado nisso, fizemos uma análise para apontar o vencedor dentro de cada uma dessas áreas, usabilidade e UX. Para calcular o vencedor em cada área, nos baseamos nas métricas (Escala Likert) aplicadas no questionário e no valor associado a cada uma das componentes da escala, então fizemos da seguinte forma:

S = Soma da quantidade de vezes que as métricas foram escolhidas pelos participantes nas questões. P = Peso associado a cada componente da escala (1 a 7)  
V = Valor Final  $V = S * P$

Ex: Concordo Parcialmente (Coursera)  $V = (S1+S2+S3+S4+S5+S6+S7+S8) * P$   
 $V = (1+0+0+1+0+1+0+1) * 5 = 20$

Após a realização dos cálculos, em relação ao questionário de usabilidade, a interface vencedora foi a Udemy com 50,4%. Através do gráfico mostrado abaixo e pela porcentagem apertada, podemos notar que as interfaces demonstraram uma concorrência acirrada. Entre os pontos de maior destaque da Udemy relatados no questionário estão: facilidade em se matricular em um curso e no controle dos vídeos (visualização, velocidade, legenda) dos cursos.

Já em relação ao questionário de design, a interface vencedora foi o Coursera com 51%. Nesse quesito também foi possível observar a concorrência acirrada entre as interfaces, como mostra o gráfico abaixo. Entre os pontos em que o Coursera recebeu maior destaque estão a sofisticação da interface e a facilidade em conseguir navegar nela, além de também ter sido considerada uma interface agradável pelos participantes e com pouca poluição visual, sendo bastante interativa e eficiente, ter um designer sofisticado, criativo, agradável, simétrico e bonito.

### 6.2.3 Análise das perguntas subjetivas

Durante a aplicação dos questionários durante o teste, extraímos os seguintes relatos:

- Aponte situações em que você achou fácil utilizar no sistema: Em relação às situações de facilidade na utilização da Coursera, foi possível observar a navegação nos cursos

e telas como tarefas mais simples no sistema. Segue abaixo os relatos. Na plataforma Udemty, as situações de facilidade na utilização foram nas tarefas de busca por cursos e edição de perfil.

– Coursera

1. A1 - Navegação do curso; Configurações da conta; Edição do perfil.
2. A2 - cadastro, pesquisa dos cursos
3. A3 - Em todos os campos
4. A4 - Navegação entre as telas

– Udemty

- \* A1 - O sistema de inscrições é bem simples e didático.
- \* A2 - Forma de cadastro; pesquisa de curso; alteração de dados pessoais; progresso do curso.
- \* A3 - Encontrar cursos e editar perfil.
- \* A4 - pesquisar outros cursos.

Em ambas as plataformas, a busca por cursos e a edição de perfil foram consideradas tarefas fáceis pelos participantes da avaliação. Na Coursera, a navegação nos cursos e entre as telas também foram destacadas como pontos de facilidade. Já na Udemty, o sistema de inscrições e o acompanhamento do progresso do curso foram mencionados como aspectos simples e didáticos. Além disso, a forma de cadastro e a alteração do perfil também foram apontadas como situações fáceis de utilizar pelos usuários.

- No âmbito das áreas em que os participantes tiveram dificuldades, alguns problemas foram relatados no Coursera, como filtragem dos cursos e seção de dúvidas. Já na Udemty, as dificuldades destacadas foram em relação à localização das perguntas frequentes e como realizar a avaliação e compartilhamento do curso.

– Coursera

- \* A1 - Filtragem de cursos.
- \* A2 - Dúvidas.
- \* A3 - Na hora de escolher forma de pagamento e comprar conteúdo por boleto, apenas por paypal
- \* A4 - Editar perfil(o botão salvar está no fim da página. A pessoa muda a foto e tem que ir até o fim pra poder salvar), Seção de ajuda está carregada demais.

– Udemty

- \* A1 - No compartilhamento, principalmente pelo fato de não ter um link direto pro WhatsApp.
- \* A2 - Dar avaliação do curso.
- \* A3 - Encontrar barra de progresso e de compartilhar curso.
- \* A4 - Localizar perguntas frequentes e avaliar.

Analisando os resultados da questão 1, onde os participantes A1, A2, A3 e A4 avaliaram a usabilidade e UX de duas plataformas de ensino, Coursera e Udemy, foi possível identificar situações de facilidade e dificuldades em ambas as plataformas.

Em relação às situações de facilidade na utilização da Coursera, os participantes destacaram a navegação nos cursos e telas, bem como a configuração da conta e edição do perfil. Já na Udemy, as situações de facilidade foram encontradas nas tarefas de busca por cursos e edição de perfil, além do sistema de inscrições ser simples e didático.

No entanto, os participantes também relataram dificuldades nas plataformas. No Coursera, as dificuldades destacadas foram a filtragem dos cursos e a seção de dúvidas. Além disso, um dos participantes (A3) mencionou que teve dificuldades na hora de escolher a forma de pagamento e comprar conteúdo por boleto, já que essa opção só estava disponível por PayPal. Outra dificuldade relatada pelos participantes do Coursera foram a edição do perfil, onde o botão "salvar" está no final da página, e a seção de ajuda, que estava carregada demais.

Já na Udemy, as dificuldades destacadas foram em relação à localização das perguntas frequentes e como realizar a avaliação e compartilhamento do curso. Além disso, os participantes tiveram dificuldades em encontrar a barra de progresso e compartilhamento do curso, avaliar o curso e localizar as perguntas frequentes. Um dos participantes da Udemy (A1) também mencionou que teve dificuldades no compartilhamento do curso, principalmente pelo fato de não ter um link direto para o WhatsApp.

Portanto, podemos concluir que ambos os participantes tiveram facilidades e dificuldades nas duas plataformas de ensino. A Coursera se destacou na facilidade de navegação nos cursos e telas, enquanto a Udemy teve um sistema de inscrições mais simples e didático. Por outro lado, ambos os participantes relataram dificuldades específicas em cada plataforma, como a filtragem dos cursos e a seção de dúvidas no Coursera e a localização das perguntas frequentes e como realizar a avaliação e compartilhamento do curso na Udemy.

- Você utilizou a Ajuda on-line do sistema em alguma momento? Em caso afirmativo, descreva em quais situações você utilizou a Ajuda on-line do sistema (comente também se as informações da Ajuda on-line foram de pouca ou grande valia). Nessa

questão, foi consenso aos utilizadores da Udemy a necessidade do uso de ajuda, mas no Coursera houve um caso de uso de ajuda online.

– Coursera

- \* A1 - Não foi necessário.
- \* A2 - Não usei.
- \* A3 - Não.
- \* A4 - Só para procurar a parte de certificados.

– Udemy

- \* A1 - Não foi necessário.
- \* A2 - Não.
- \* A3 - Não.
- \* A4 - Não.

Considerando as respostas dos usuários, pode-se inferir que a usabilidade da plataforma em relação à ajuda online é satisfatória tanto no Coursera quanto na Udemy. Embora nenhum dos usuários da Udemy tenha precisado usar a ajuda online, é positivo que eles tenham expressado a necessidade do recurso. Já no Coursera, um dos usuários utilizou a ajuda online para buscar informações específicas sobre certificados, o que demonstra que a opção de ajuda online foi útil nesse caso.

- - Diante do teste realizado, você acha que o software atingiu o objetivo para o qual foi desenvolvido? Explique. Na perspectiva do objetivo do software, todos os participantes relataram que ambos os sistemas atingiram o objetivo de oferecer um modelo de ensino/aprendizagem online apoiado no uso de Tecnologia de Informação e Comunicação para o oferecimento de cursos. Segue abaixo as opiniões relatadas:

– Coursera

- \* A1 - Sim. Encontrei cursos de professores de universidades renomadas e cursos de empresas de nível internacional.
- \* A2 - Sim, visto que tem o intuito de oferecer cursos e é feito isso.
- \* A3 - Sim. Usabilidade bem simples.
- \* A4 - Quase todo. Tiveram seus problemas, mas no geral foi ótimo.

– Udemy

- \* A1 - Sim, para uma plataforma de curso de aprendizado a Udemy cumpre bem o seu papel, torna as coisas mais fáceis para a realização.

- \* A2 - Sim, uma vez que a plataforma permitiu fácil acesso, desde o cadastro até a inscrição e consumo dos cursos.
- \* A3 - Para um site com plataforma de vários cursos é considerável objetivo e facilita a experiência para o usuário.
- \* A4 - sim. os cursos são de rápido acesso e as recomendações são de acordo com os cursos que demonstro interesse.

Coursera e Udemy, atingiram o objetivo para o qual foram desenvolvidos, que é oferecer uma plataforma de ensino/aprendizagem online apoiada em Tecnologia de Informação e Comunicação. Todos os participantes relataram que conseguiram encontrar cursos relevantes e se inscreverem facilmente nos cursos desejados.

Na Coursera, os participantes destacaram a qualidade dos cursos oferecidos por professores de universidades renomadas e empresas de nível internacional, além da usabilidade simples da plataforma. Houve apenas um relato de que houve problemas pontuais.

Na Udemy, os participantes enfatizaram a facilidade de acesso e consumo dos cursos, desde o cadastro até as recomendações personalizadas de acordo com seus interesses.

Por fim, foi aberto um espaço para que o participante sugerisse melhorias no sistema. Entre as melhorias destacadas para o Coursera, foram relatadas a adição de um filtro para cursos gratuitos/pagos e a recomendação de colocar o botão de salvar as alterações do perfil mais visível para o usuário. Já para a Udemy, as melhorias recomendadas foram em relação ao design do site, principalmente no quesito da quantidade de informação na tela para tornar a interface menos poluída, e adição de filtro e recomendação para cursos grátis oferecidos pelo sistema. Segue abaixo a opinião completa dos usuários.

- O espaço abaixo é reservado para que você exponha sua opinião e sugira melhorias no sistema:
  - Coursera
    - \* A1 - Plataforma simples e relativamente intuitiva. Sugiro a adição de um novo filtro na busca por cursos: Filtrar por cursos gratuitos e pagos.
    - \* A2 - Recomendo melhorias no sistema, por exemplo na barra de progressão do curso.
    - \* A3 - Nada a comentar.
    - \* A4 - Botar um botão em cima para salvar as alterações no perfil.
  - Udemy
    - \* A1 - Padronizar as imagens de capa nos cursos, de certa forma existe uma pequena poluição visual em alguns casos.

- \* Senti falta de tópicos/categorias de "cursos longos", "cursos curtos", "cursos grátis".
- \* A3 - O site deveria fazer recomendações de cursos gratuitos e bem avaliados na sua homepage com base nos cursos que o usuário já fez antes.
- \* A4 - achei visualmente poluído, design ruim/sem atrativo e um pouco desorganizado.

Analisando as opiniões dos usuários sobre o Coursera, é possível perceber que a maioria dos participantes considera a plataforma simples e intuitiva, mas ainda apontam algumas melhorias, como a adição de um filtro para cursos gratuitos/pagos e tornar o botão de salvar as alterações do perfil mais visível para o usuário. Essas são sugestões relevantes, pois podem melhorar a experiência do usuário e facilitar a busca por cursos mais adequados ao seu interesse.

Em relação à Udemy, alguns usuários relataram que o design do site é poluído e pouco atrativo, sugerindo a necessidade de melhorias no quesito de quantidade de informação na tela para tornar a interface menos poluída. Outra sugestão importante é a adição de filtros e recomendações para cursos grátis oferecidos pelo sistema, além de tópicos/categorias que permitam uma melhor organização dos cursos. Além disso, a padronização das imagens de capa dos cursos pode ajudar a melhorar a estética visual do site.

### 6.3 Questionário de Avaliação da Experiência do usuário

Para este teste optamos por utilizar o questionário UEQ (*User Experience Questionnaire*), que avalia 6 fatores de UX, são eles: (1) A **Perspicuidade** que se refere a capacidade de executar tarefas com eficiência e eficácia; (2) A **Eficácia** que trata da capacidade de executar a tarefa; (3) A **Confiabilidade** avalia se o usuário está no controle do software; (4) A **estimulação** avalia o desejo do usuário de continuar usando o software; (5) A **Novidade** avalia o grau em que o software captura a atenção dos usuários; e (6) A **Atratividade** avalia a impressão do usuário (Hassenzahl and Tractinsky 2006).

Para cada um dos fatores acima, temos uma lista de adjetivos associados a ela que a UEQ utiliza, tendo em sua totalidade 26 pares de adjetivos opostos. Cada um desses adjetivos é importante para avaliar a qualidade de um software. A tabela 28 abaixo demonstra esses adjetivos para medir a satisfação do usuário com relação aos aspectos do produto ou serviço, permitindo que os desenvolvedores possam identificar pontos de melhoria e aprimorar a experiência do usuário (Hassenzahl and Tractinsky 2006).

O preenchimento das respostas do questionário por cada um dos participantes deve ser realizada de forma imediata e da forma mais espontânea possível. Neste questionário

não há respostas "certas" ou respostas "erradas". A opinião pessoal do participante é o mais importante (Hassenzahl and Tractinsky 2006).

A UEQ utiliza uma escala similar a Escala de Likert, que varia a pontuação na escala de 1 a 7 pontos, onde os usuários marcam o ponto que está mais próximo do adjetivo que melhor descreve sua UX, onde o grau 1 da escala pode, por exemplo, representar um extremo oposto (muito feio) e 7 representa o outro extremo oposto (muito atraente) (Hassenzahl and Tractinsky 2006).

Tabela 28: Listas de adjetivos da UEQ

Adjetivo negativo	Escala							Adjetivo positivo
	1	2	3	4	5	6	7	
Desagradável	<input type="radio"/>	Agradável						
Incompreensível	<input type="radio"/>	Compreensível						
Criativo	<input type="radio"/>	Sem criatividade						
De Fácil aprendizagem	<input type="radio"/>	De difícil aprendizagem						
Valioso	<input type="radio"/>	Sem valor						
Aborrecido	<input type="radio"/>	Excitante						
Desinteressante	<input type="radio"/>	Interessante						
Imprevisível	<input type="radio"/>	Previsível						
Rápido	<input type="radio"/>	Lento						
Original	<input type="radio"/>	Convencional						
Obstrutivo	<input type="radio"/>	Condutor						
Bom	<input type="radio"/>	Mau						
Complicado	<input type="radio"/>	Fácil						
Desinteressante	<input type="radio"/>	Atrativo						
Comum	<input type="radio"/>	Vanguardista						
Incómodo	<input type="radio"/>	Cómodo						
Seguro	<input type="radio"/>	Inseguro						
Motivante	<input type="radio"/>	Desmotivante						
Atende as expectativas	<input type="radio"/>	Não atende as expectativas						
Ineficiente	<input type="radio"/>	Eficiente						
Evidente	<input type="radio"/>	Confuso						

Continuação na próxima página

Tabela 28: Listas de adjetivos da UEQ (Continuação)

Impraticável	o	o	o	o	o	o	o	Prático
Organizado	o	o	o	o	o	o	o	Desorganizado
Atraente	o	o	o	o	o	o	o	Feio
Simpático	o	o	o	o	o	o	o	Antipático
Conservador	o	o	o	o	o	o	o	Inovador

Logo abaixo explicamos o significado de cada um dos adjetivos, segundo (Hassenzahl and Tractinsky 2006), que são utilizados na UEQ demonstrados na tabela acima:

- **Desagradável:** Que causa desprazer ou incômodo, em oposição a **Agradável:** Que agrada, que é agradável.
- **Incompreensível:** Que não pode ser compreendido, em oposição a **Compreensível:** Que pode ser compreendido.
- **Criativo:** Que tem a capacidade de criar, de inventar coisas novas, em oposição a **Sem criatividade:** Que não tem a capacidade de criar, que não é criativo.
- **De Fácil aprendizagem:** Que é fácil de ser aprendido, em oposição a **De difícil aprendizagem:** Que é difícil de ser aprendido.
- **Valioso:** Que tem valor, importância, em oposição a **Sem valor:** Que não tem valor, sem importância.
- **Aborrecido:** Que causa tédio, enfado, em oposição a **Excitante:** Que é emocionante, empolgante.
- **Desinteressante:** Que não desperta interesse, em oposição a **Interessante:** Que desperta interesse, chama a atenção.
- **Imprevisível:** Que não pode ser previsto, em oposição a **Previsível:** Que pode ser previsto.
- **Rápido:** Que é veloz, ágil, em oposição a **Lento:** Que é devagar, vagaroso.
- **Original:** Que é novo, único, diferente, em oposição a **Convencional:** Que é tradicional, comum, corriqueiro.
- **Obstrutivo:** Que impede ou dificulta a passagem, em oposição a **Condutor:** Que conduz, que facilita a passagem.

- **Bom:** Que é de boa qualidade, útil, agradável, em oposição a **Mau:** Que é de má qualidade, prejudicial, desagradável.
- **Complicado:** Que é difícil de ser entendido, realizado, em oposição a **Fácil:** Que é simples, fácil de ser entendido ou realizado.
- **Desinteressante:** Que não desperta interesse, em oposição a **Atrativo:** Que atrai, que é interessante.
- **Comum:** Que é normal, corriqueiro, em oposição a **Vanguardista:** Que é inovador, diferente, fora do comum.
- **Incômodo:** Que causa desconforto, desagrado, em oposição a **Cômodo:** Que é confortável, agradável.
- **Seguro:** Que é livre de riscos, perigos, em oposição a **Inseguro:** Que é perigoso, arriscado.
- **Motivante:** Que incentiva, anima, em oposição a **Desmotivante:** Que desanima, desestimula.
- **Atende as expectativas:** Que corresponde ao esperado, em oposição a **Não atende as expectativas:** Que não corresponde ao esperado.
- **Ineficiente:** Que não funciona bem, que não é eficaz, em oposição a **Eficiente:** Que funciona bem, que é eficaz.
- **Evidente:** Que é claro, óbvio, que não deixa dúvidas, em oposição a **Confuso:** Que é obscuro, que causa dúvidas.
- **Impraticável:** Que não pode ser posto em prática, em oposição a **Prático:** Que pode ser posto em prática, é viável.
- **Organizado:** Que está bem arrumado, em ordem, em oposição a **Desorganizado:** Que está desarrumado, bagunçado.
- **Atraente:** que causa atração, que é bonito, em oposição a **Feio:** que não é bonito, que causa repulsa ou desagrado.
- **Simpático:** que tem simpatia, que é agradável, em oposição a **Antipático:** que não tem simpatia, que é desagradável.
- **Conservador:** que se apegua às tradições, à ordem estabelecida, em oposição a **Inovador:** que busca a mudança, que traz novidades e inovações.

O questionário acima foi aplicado sobre cada um dos participantes do teste de forma a podermos coletar os resultados e obter assim uma visão dos resultados.

### 6.3.1 Análise do questionário UEQ na Coursera

Para realizar a análise dos resultados dos questionários respondidos pelos 4 participantes que testaram o software Coursera, inserimos as informações no artefatos da UEQ para calcular e obter as informações consolidadas sobre os 6 fatores já mencionados anteriormente. Como resultado, a figura 7 mostra que atualmente, o Coursera precisa melhorar nos requisitos de Transparência, Eficiência e Estimulação.

Em seguida extraímos do relatório de análise da UEQ a média gerada em cada um dos fatores estudados para podemos entender mais sobre as respostas dos participantes. Explicaremos os resultados após a apresentarmos tabela 29.

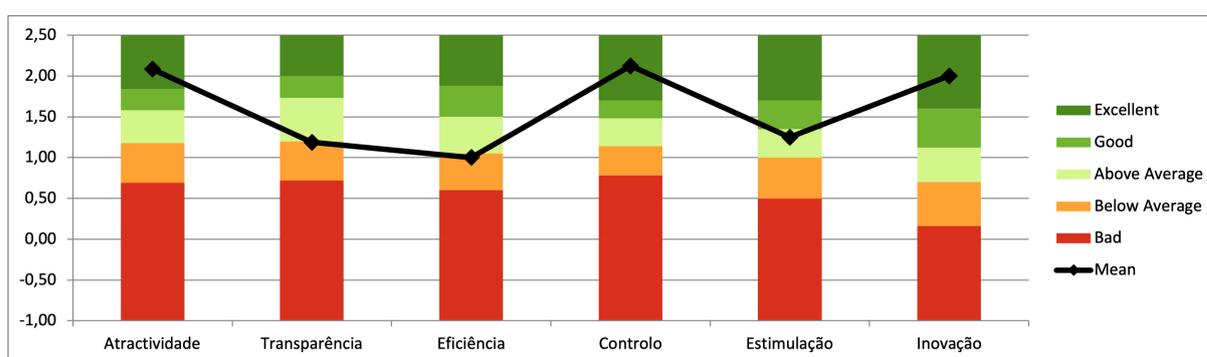


Figura 7 – UEQ: Gráfico com dos resultados da Coursera

A tabela 29 a seguir vemos os resultados para a média obtida na avaliação.

Tabela 29: UEQ: Resultado da Escala do Coursera

Escala	Média
Atractividade	2,08
Transparência	1,19
Eficiência	1,00
Controlo	2,13
Estimulação	1,25
Inovação	2,00

A escala de Atractividade apresentou uma média de 2,08, o que indica que os usuários avaliaram positivamente a aparência e o design do software. Isso pode indicar que a interface do software é visualmente atraente e agradável para os usuários.

A escala de Transparência apresentou a média mais baixa, de 1,19, o que indica que os usuários não se sentem confortáveis com a forma como as informações são apresentadas e a forma como as tarefas são executadas. Isso pode indicar que a interface do software é

confusa ou difícil de usar, tornando difícil para os usuários saberem o que está acontecendo no software e como realizar tarefas específicas.

A escala de Eficiência apresentou uma média de 1,00, o que indica que os usuários acham que o software não é eficiente na realização de tarefas e na entrega de resultados. Isso pode indicar que o software é lento ou que tem muitos erros e falhas, o que impede os usuários de trabalhar de forma produtiva e eficiente.

A escala de Controle apresentou uma média de 2,13, o que indica que os usuários se sentem capazes de controlar o software e realizar tarefas de forma eficaz. Isso pode indicar que o software oferece recursos e funcionalidades que permitem aos usuários executar tarefas com facilidade e de forma intuitiva.

A escala de Estimulação apresentou uma média de 1,25, o que indica que os usuários não sentem que o software os desafia ou os motiva a continuar usando-o. Isso pode indicar que o software não é interessante o suficiente ou não oferece recursos que incentivem os usuários a explorar suas funcionalidades e recursos.

Por fim, a escala de Inovação apresentou uma média de 2,00, o que indica que os usuários acham que o software apresenta recursos e funcionalidades inovadoras. Isso pode ser um ponto positivo para o software, pois pode atrair usuários interessados em utilizar ferramentas tecnológicas mais avançadas.

Com base nas opiniões dos usuários, é possível concluir que tanto o *Coursera* quanto a *Udemy* oferecem uma plataforma de ensino/aprendizagem online eficiente e de fácil acesso. No entanto, ainda existem algumas melhorias que podem ser feitas para aprimorar a experiência do usuário e apesar disso, avaliando as respostas subjetivas, tivemos um maior número de respostas positivas para o *Coursera* em relação a *Udemy*.

### 6.3.2 Análise do questionário UEQ na Udemy

Para realizar a análise dos resultados da Udemy, foi realizada a extração dos dados dos 4 questionários respondidos pelo grupo de participantes. Inserimos as informações no artefatos da UEQ para calcular e obter as informações consolidadas sobre os 6 fatores já mencionados anteriormente e como resultado, a figura 8 mostra que atualmente, a Udemy precisa melhorar nos requisitos de Eficiência e Estimulação.

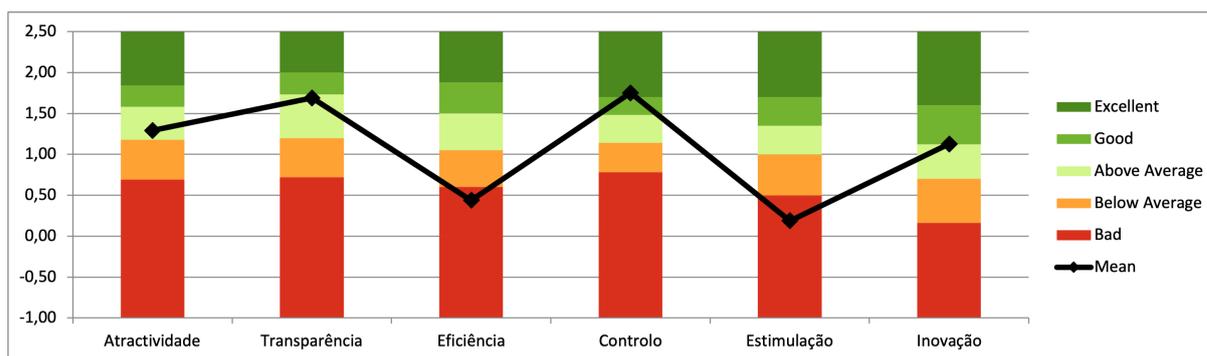


Figura 8 – Udemq: Gráfico com dos resultados da Udemq

Complementando o estudo extraímos as informações da Udemq para entender as médias geradas sobre os 6 fatores estudados. A tabela 30 a seguir temos os resultados para a média obtida na avaliação, em que faremos a análise dos dados após sua demonstração.

Tabela 30: Udemq: Resultado da Escala da Udemq

Escala	Média
Atractividade	1,29
Transparência	1,69
Eficiência	0,44
Controlo	1,75
Estimulação	0,19
Inovação	1,13

Com base nos resultados da escala média da Udemq, podemos observar que a avaliação da usabilidade do produto em questão apresentou resultados abaixo da média esperada em várias dimensões avaliadas.

A dimensão de Atractividade, que se refere à qualidade estética e de design do produto, apresentou uma média de 1,29, o que indica que o produto não foi avaliado como atraente ou visualmente agradável.

A dimensão de Transparência, que se refere à clareza e compreensibilidade das informações apresentadas pelo produto, apresentou uma média de 1,69, o que indica que o produto apresenta dificuldades em transmitir as informações de forma clara e compreensível para o usuário.

A dimensão de Eficiência, que se refere à facilidade e rapidez com que as tarefas podem ser realizadas pelo usuário, apresentou uma média de 0,44, indicando que o produto apresenta dificuldades em proporcionar uma experiência de uso ágil e eficiente.

A dimensão de Controlo, que se refere ao grau de controle que o usuário tem sobre

o produto, apresentou uma média de 1,75, indicando que o produto permite um grau razoável de controle por parte do usuário.

A dimensão de Estimulação, que se refere ao grau em que o produto pode estimular e motivar o usuário, apresentou uma média muito baixa de 0,19, o que indica que o produto não é capaz de motivar ou envolver o usuário.

Por fim, a dimensão de Inovação, que se refere à percepção do usuário sobre a originalidade e inovação do produto, apresentou uma média de 1,13, indicando que o produto não foi avaliado como muito inovador.

Dessa forma, pode-se concluir que o produto em questão apresenta alguns problemas de usabilidade e design que precisam ser melhorados para proporcionar uma experiência mais satisfatória e eficiente para o usuário.

### 6.3.3 Análise comparativa da Udemy e Coursera, através da UEQ

Comparando O Coursera (azul), com a Udemy (vermelha), chegamos ao gráfico onde mostra que na parte de UX, para este teste, a Coursera apresentou as melhorias pontuações, conforme imagem abaixo na figura 9:

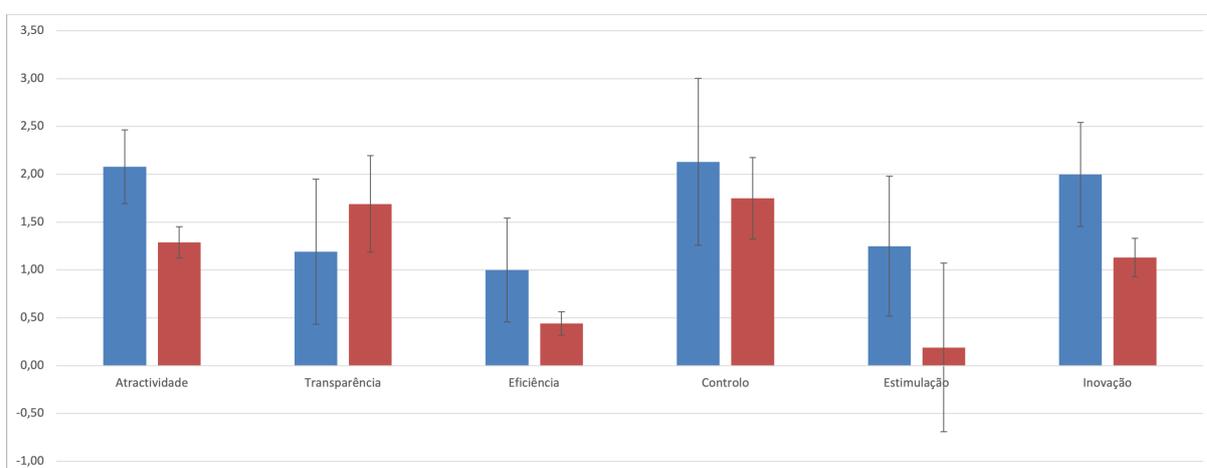


Figura 9 – Comparativo entre Coursera e Udemy

Em relação aos fatores, pelo gráfico acima temos os seguintes resultados:

- **Atratividade:** a Coursera teve uma pontuação média mais alta (2,08) do que a Udemy (1,29). Isso sugere que a Coursera é mais atraente para os usuários em termos de design, layout e experiência visual geral.
- **Transparência:** a Udemy obteve uma pontuação média mais alta (1,69) do que a Coursera (1,19). Isso sugere que a Udemy é mais transparente em suas políticas, termos e condições e comunicação geral do que a Coursera.

- **Eficiência:** a Coursera obteve uma pontuação média mais alta (1,00) do que a Udemy (0,44). Isso sugere que a Coursera é mais eficiente em termos de velocidade de carregamento de conteúdo, facilidade de uso e desempenho geral do que a Udemy.
- **Controle:** a Coursera teve uma pontuação média mais alta (2,13) do que a Udemy (1,75). Isso sugere que a Coursera oferece mais controle e flexibilidade para os usuários personalizarem sua experiência de aprendizagem do que a Udemy.
- **Estimulação:** a Coursera obteve uma pontuação média mais alta (1,25) do que a Udemy (0,19). Isso sugere que a Coursera é mais estimulante e envolvente para os usuários do que a Udemy.
- **Inovação:** a Coursera teve uma pontuação média mais alta (2,00) do que a Udemy (1,13). Isso sugere que a Coursera é mais inovadora em termos de apresentação de conteúdo, recursos e funcionalidades do que a Udemy.

No geral, a plataforma Coursera teve uma pontuação média de UEQ mais alta do que a Udemy. A pontuação média da Coursera foi de 1,68, enquanto a Udemy obteve uma pontuação média de 1,04. A análise das pontuações UEQ sugere que a Coursera oferece uma experiência de usuário mais atraente, eficiente, flexível, estimulante e inovadora do que a Udemy. No entanto, a Udemy parece ser mais transparente em suas políticas e comunicação geral com os usuários.

## 6.4 Avaliação do processo ABTUX através do Feedback dos Avaliadores

Após a finalização do teste, foi aplicado um questionário de avaliação para os avaliadores que utilizaram o teste ABTUX. Entre os pontos levantados no questionário constam tanto questionamentos quantitativos quanto qualitativos. O objetivo para a coleta do feedback desses avaliadores é a de entender como foi a experiência de utilizar o processo ABTUX e ainda encontrar os pontos onde o processo pode ter deixado a desejar. Dessa forma, será possível ter um parâmetro fiel de quem aplicou o teste na prática, de modo que seja possível verificar se o intuito de facilidade da aplicação e da identificação tanto de defeitos quanto do melhor software ficaram nítidos na visão dos avaliadores.

No total, a equipe que realizou o papel de avaliador contou com 4 integrantes, então foram coletadas o *feedback* de todos eles a partir desse questionário. Para as perguntas de quantitativas, foi utilizada uma escala que conta com as opções concordo totalmente (CT), concordo parcialmente (CP), discordo parcialmente (DP) e discordo totalmente (DT). A tabela 30 abaixo mostra tanto os questionamentos levantados quanto a resposta da equipe seguindo a escala mencionada.

Tabela 31: Questionário dos avaliadores

<b>Questionário dos Avaliadores</b>				
<b>Questão</b>	<b>CT</b>	<b>CP</b>	<b>DP</b>	<b>DT</b>
Para mim ficou claro como utilizar o ABTUX para avaliar dois softwares	<b>3</b>	<b>1</b>	-	-
A interação com o processo ABTUX foi uma experiência satisfatória.	-	<b>4</b>	-	-
O ABTUX utilizado demandou pouco esforço mental.	-	<b>3</b>	<b>1</b>	-
Eu acho que o ABTUX utilizado é fácil de usar.	<b>2</b>	<b>2</b>	-	-
O ABTUX utilizado foi claro e fácil de entender.	<b>1</b>	<b>3</b>	-	-
É fácil para mim lembrar como utilizar o ABTUX para avaliar qualquer software	<b>1</b>	<b>3</b>	-	-
De maneira geral, eu acho fácil encontrar defeitos em software com o apoio do ABTUX utilizado.	<b>2</b>	<b>2</b>	-	-
O ABTUX utilizado melhora o meu desempenho ao encontrar defeitos em softwares.	<b>2</b>	<b>2</b>	-	-
O ABTUX utilizado me permite economizar tempo.	<b>3</b>	<b>1</b>	-	-
O ABTUX utilizando me permite realizar a avaliação de usabilidade e UX de qualquer tipo de software	<b>2</b>	<b>2</b>	-	-
O ABTUX utilizando aumenta minha eficácia ao identificar defeitos em softwares	<b>1</b>	<b>3</b>	-	-
O ABTUX utilizado melhora a qualidade na atividade de identificação de defeitos em qualquer tipo softwares	<b>1</b>	<b>3</b>	-	-
O ABTUX utilizado melhora a minha produtividade ao identificar defeitos em qualquer tipo de software	<b>1</b>	<b>3</b>	-	-
Eu acho que o ABTUX utilizado é útil para encontrar defeitos em qualquer tipo de software	<b>2</b>	<b>2</b>	-	-

Continuação na próxima página

Tabela 31: Questionário dos avaliadores (Continuação)

Assumindo que eu tenha acesso ao processo, eu pretendo utilizar ele para avaliar versões de um software ou softwares concorrentes.	<b>2</b>	<b>2</b>	-	-
--	----------	----------	---	---

Como é possível observar pelas respostas majoritariamente positivas, no geral o processo foi compreendido e realizado com sucesso. Entre os pontos levantados no questionário pode-se destacar a clareza de como utilizar o ABTUX para avaliar dois softwares e a economia de tempo com a realização do processo, ambos com três respostas de concordo totalmente e um com concordo parcialmente. Além desses, outros pontos como satisfação ao interagir com o processo, facilidade de uso e melhora na produtividade e qualidade ao identificar defeitos.

Quanto as respostas de aspecto negativo, apenas uma foi encontrada ao ser questionado se o processo demandou pouco esforço mental. Embora um dos avaliadores da equipe tenha respondido com discordo parcialmente, os outros três concordaram parcialmente com a questão, o que continua demonstrando que, de modo geral, o processo cumpre seu objetivo de facilitar a comparação e identificação de defeitos de softwares concorrentes ou versões do mesmo software.

No âmbito das perguntas qualitativas, o primeiro questionamento levantado abordou o que é fácil de utilizar no processo ABTUX. De modo geral, a avaliação dos testadores informou que a facilidade ao comparar os resultados foi notável, principalmente pela simplicidade em recolher o feedback dos usuários sobre a facilidade em realizar as tarefas, também destacada como uma etapa fundamental para garantir a qualidade de um produto e melhor experiência do usuário. Nesse sentido, foi destacado que ao comparar os diferentes sistemas, é possível identificar pontos fortes e fracos em cada um deles e avaliar quais recursos e funcionalidades devem ser aprimoradas, assim como estabelecer as semelhanças e diferenças ao relacionar os sistemas testados.

Em seguida foi questionado quais foram as dificuldade na execução do processo ABTUX. Os avaliadores destacaram a dificuldade em encontrar participantes para a aplicação dos testes, principalmente por conta de um dos critérios de inclusão, que seria nunca ter usado uma das plataformas (Udemy /Coursera), o que, por conta da popularidade das plataformas, demandou um esforço maior.

A respeito do feedback dos avaliadores, é possível concluir que o ABTUX aplicado é bastante completo e satisfatório. Os avaliadores consideram o processo detalhado, pois ele cumpre com seu objetivo de avaliar a usabilidade do usuário e identificar os problemas relacionados aos sistemas. Ademais, o processo pode ser aplicado em testes futuros sem grande dificuldades, tendo em vista a sua clareza e flexibilidade para se adaptar a diferentes

projetos, sem requerer grandes mudanças.

A respeito do completude do processo ABTUX, os avaliadores concluíram que o processo aplicado é bastante completo e satisfatório. Os avaliadores consideram o processo detalhado, pois ele cumpre com seu objetivo de avaliar a usabilidade do usuário e identificar os problemas relacionados aos sistemas. Ademais, o processo pode ser aplicado em testes futuros sem grandes dificuldades, tendo em vista a sua clareza e flexibilidade para se adaptar a diferentes projetos, sem requerer grandes mudanças.

Quando perguntados sobre o que poderia ser feito para diminuir o tempo e esforço da aplicação do teste, os avaliadores salientaram as etapas criadas para o processo foram simples e já foram rápidas o bastante, destacando a praticidade na execução dessas tarefas.

Ao final, o último questionamento foi sobre o que poderia ser feito para melhorar a identificação de defeitos ao decorrer do processo. Os avaliadores opinaram que não seria necessário realizar mudanças no processo e também sugeriram que, futuramente, a aplicação de testes em outras plataformas para que se possa continuar a validação da eficiência dos artefatos criados do teste A/B e do processo ABTUX por completo.

Em síntese, a praticidade e facilidade do processo se demonstrou notável para toda a equipe. Dessa forma, fica comprovado que o objetivo do processo foi atingido e também que o ABTUX pode ser replicado para testar mais softwares, de modo que a validação de sua eficiência continue sendo validada a cada aplicação.

## 7 CONCLUSÃO E TRABALHOS FUTUROS

Neste capítulo são apresentadas nossas considerações finais sobre o processo ABTUX v2.0. Também são apresentadas as expectativas de trabalhos futuros, orientando novas pesquisas para teste de software através de usabilidade e UX sobre um teste A/B.

### 7.1 Epílogo

Esta dissertação de mestrado apresentou um novo processo de avaliação de experiência de usuário com UX de forma genérica, o qual podemos realizar o comparativo de dois softwares distintos ou não, com a mesma funcionalidade e aplicar de forma similar, com a possibilidade de incluir ou alterar itens do processo para se adequar ao caso de uso que será avaliado.

Durante a nossa avaliação, aplicando o este estudo sobre as plataformas *Coursera* e *Udemy*, onde conseguimos identificar os pontos de falha de usabilidade e experiência do usuário através de um teste A/B tradicional.

Coletamos os dados durante essas avaliações em contextos reais e conseguimos verificar em que medida o novo processo permite coletar dados relevantes de usabilidade e UX do ponto de vista dos usuários. Além disso, avaliamos os artefatos que desenvolvemos para a execução das atividades obrigatórias dentro do nosso processo.

Ao fornecer um processo e um conjunto de artefatos para avaliação de usabilidade e UX, conseguimos apoiar a os usuários e a indústria de desenvolvimento de software a identificação de problemas de usabilidade e UX, bem como apoiar a correção desses problemas, melhorando a qualidade do software futuro e a experiência dos usuários ao realizar tarefas e atingir metas em sistemas e aplicativos de informação.

Em nosso estudo de caso aplicado, através do ABTUX v2.0, foi possível avaliar a interface Udemy no âmbito de usabilidade e da interface e o Coursera no quesito design. Ao ser realizado na prática o processo, além de enriquecer o passo a passo de intenção do usuário em cada tarefa semi-estruturada para o participante.

Dessa forma, concluímos que o fluxo de realização do teste proposto é prático e eficiente, sendo possível, também, fazer uma análise de resultados detalhada com base na escala utilizada e na divisão do questionário de respostas dos participantes em diferentes áreas.

Além disso, iremos disponibilizar para a sociedade todos os artefatos produzidos por este trabalho para facilitar a replicação do teste e tornar o ABTUX um critério essencial

na tomada de decisão para a escolha de um software.

## 7.2 Contribuições

Nesta pesquisa foi apresentada uma série de contribuições importantes para o campo da usabilidade e experiência do usuário (UX). Abaixo, descrevo em detalhes os principais pontos que foram abordados:

- **Proposta de um novo processo de teste A/B para avaliar a usabilidade e UX**

O propósito desta pesquisa é de criar um novo processo de avaliação de usabilidade e UX através de um Teste A/B em que além de criarmos e detalhamos todas as atividades de um processo de avaliação genérico, foi criado dentro do fluxo atividades opcionais para substituição de cenários de forma ao processo genérico se adaptar a testes mais específicos, assim como atividades adicionais para incluir atividades de forma a deixar o processo mais robusto e flexível, atendendo além de cenários de testes genéricos os testes especializados. Neste processo também foi elaborado uma seleção de métricas relevantes, a definição de grupos de teste e controle, a criação de protótipos, a coleta de dados e a análise estatística dos resultados. A proposta desse processo contribuiu para aprimorar a avaliação da usabilidade e UX em projetos de desenvolvimento de software.

- **Repositório de artefatos para aplicação do processo**

Como complemento à proposta do novo processo de teste A/B, foi desenvolvido um repositório de artefatos. Esse repositório contém modelos, diretrizes e exemplos práticos para a aplicação do processo em diferentes contextos. Com isso, é possível facilitar a adoção e implementação do processo por outros pesquisadores e profissionais da área. O repositório de artefatos se mostra como uma contribuição valiosa para aqueles que desejam utilizar o processo de teste A/B para avaliar a usabilidade e UX de suas próprias aplicações.

- **Exemplo de estudo experimental para avaliação de processos de teste A/B**

Além da proposta e do repositório de artefatos, foi apresentado um exemplo de estudo experimental detalhado para avaliar a eficácia e a aplicabilidade do processo de teste A/B proposto. Esse estudo experimental foi realizado em um ambiente controlado, seguindo todas as etapas do processo. Os resultados obtidos demonstraram a efetividade do processo na identificação de melhorias na usabilidade e UX de uma aplicação específica. Ao fornecer um exemplo concreto, contribuí para a validação e a replicação do processo proposto em outros estudos e projetos.

- **Comparativo entre dois sistemas concorrentes**

Por fim, outro ponto de contribuição deste trabalho foi o comparativo entre dois sistemas concorrentes. Foi realizada uma análise detalhada desses sistemas, levando em consideração critérios como desempenho, funcionalidades, usabilidade e UX. Através desse comparativo, foi possível identificar as principais diferenças e semelhanças entre os sistemas, oferecendo *insights* valiosos para os desenvolvedores e usuários dessas aplicações. Essa comparação contribuiu para uma compreensão mais aprofundada do cenário atual e auxiliou na tomada de decisões informadas.

- **Publicação de artigo com os resultados desta pesquisa**

A partir deste trabalho tivemos a aprovação e publicação na HCI International [Fernandes et al. 2022] Fernandes, Í., Rocha, S., Portela, C., Braz Junior, G., Almeida, J., Silva, A., Viana, D., Rabelo, J., Paiva, A., and Rivero, L. (2022). *Defining an a/b testing process for usability and user experience evaluation through the analysis of the results of a literature review*. In *HCI International 2022-Late Breaking Papers. Design, User Experience and Interaction: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, pages 204–213. Springer.

### 7.3 Perspectivas futuras

Como trabalho próximos passos, pretendemos aplicar o processo ABTUX v2.0 em mais softwares e assim obter novas percepções de adequações de fluxo no experimento e a possibilidade de utilizar o teste A/B/n para podemos comparar de forma dinâmica mais cenários de teste. O número de participantes também é necessário incrementar para que tenhamos informações estatísticas suficientes para aumentar a confiabilidade dos resultados.

Incluir no processo as tarefas adicionais, de forma a adaptar o processo também a receber não apenas tarefas opcionais. Aplicar o processo na prática na indústria; Realizar um novo refinamento do processo para reduzir esforço e tempo na execução das tarefas.

# Referências

- [Albert and Tullis 2013] Albert, B. and Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes.
- [Anderson 2016] Anderson, C. (2016). How to design for a/b testing. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2773–2776. ACM.
- [Bargas-Avila and Hornbæk 2011] Bargas-Avila, J. A. and Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2689–2698.
- [Borsci et al. 2013] Borsci, S., Kurosu, M., Federici, S., and Mele, M. L. (2013). *Computer systems experiences of users with and without disabilities: an evaluation guide for professionals*. CRC Press.
- [Fernandes et al. 2022] Fernandes, Í., Rocha, S., Portela, C., Braz Junior, G., Almeida, J., Silva, A., Viana, D., Rabelo, J., Paiva, A., and Rivero, L. (2022). Defining an a/b testing process for usability and user experience evaluation through the analysis of the results of a literature review. In *HCI International 2022-Late Breaking Papers. Design, User Experience and Interaction: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, pages 204–213. Springer.
- [Fernandez et al. 2011] Fernandez, A., Insfran, E., and Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study. *Information and Software Technology*, 53(8):789–817.
- [Firmenich et al. 2019] Firmenich, S., Garrido, A., Grigera, J., Rivero, J. M., and Rossi, G. (2019). Usability improvement through a/b testing and refactoring. *Software Quality Journal*, 27(1):203–240.
- [Gardey and Garrido 2020] Gardey, J. C. and Garrido, A. (2020). User experience evaluation through automatic a/b testing. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, pages 25–26.
- [Garousi et al. 2020] Garousi, V., Rainer, A., Lauvås Jr, P., and Arcuri, A. (2020). Software-testing education: A systematic literature mapping. *Journal of Systems and Software*, 165:110570.

- [Garrett 2011] Garrett, J. (2011). The elements of user experience: User-centered design for the web and beyond 2 (california: New riders pearson education).
- [Gegner and Runonen 2012] Gegner, L. and Runonen, M. (2012). For what it is worth: Anticipated experience evaluation. In *Proceedings of the 8th International Conference on Design and Emotions*. Central Saint Martins University of the Arts London and the Design and . . . .
- [Hassenzahl and Tractinsky 2006] Hassenzahl, M. and Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, 25(2):91–97.
- [ISO and STANDARD 2010] ISO, B. and STANDARD, B. (2010). Ergonomics of human-system interaction. *British Standatds Institution*.
- [Jabbar et al. ] Jabbar, A., Memon, R. N., Memon, I., Arain, A. A., and Sodhar, I. N. Web design trends and their usability by a| b testing method.
- [Kohavi and Longbotham 2017] Kohavi, R. and Longbotham, R. (2017). Online controlled experiments and a/b testing. *Encyclopedia of machine learning and data mining*, 7(8):922–929.
- [Krug 2014] Krug, S. (2014). *Don't Make Me Think: A Common Sense Approach to Web Usability*. New Riders.
- [Kruger et al. 2016] Kruger, R. M., Gelderblom, H., and Beukes, W. (2016). The value of comparative usability and ux evaluation for e-commerce organisations.
- [Kujala et al. 2011] Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., and Sinnelä, A. (2011). Ux curve: A method for evaluating long-term user experience. *Interacting with computers*, 23(5):473–483.
- [Lund 2001] Lund, A. M. (2001). Measuring usability with the use questionnaire. *Usability Interface*, 8(2):3–6.
- [Magazine 2012] Magazine, S. (2012). *A Field Guide to Usability Testing*. Smashing Magazine.
- [Nakamura et al. 2021] Nakamura, W. T., Ahmed, I., Redmiles, D., Oliveira, E., Fernandes, D., de Oliveira, E. H., and Conte, T. (2021). Are ux evaluation methods providing the same big picture? *Sensors*, 21(10):3480.
- [Nielsen 1994] Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann Publishers, San Francisco, Calif.
- [Nielsen 2012] Nielsen, J. (2012). *Usability engineering*. Morgan Kaufmann.

- [Niranjanamurthy et al. 2014] Niranjanamurthy, M., Nagaraj, A., Gattu, H., and Shetty, P. K. (2014). Research study on importance of usability testing/user experience (ux) testing. *International Journal of Computer Science and Mobile Computing*, 3(10):78–85.
- [Norman 2013] Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- [Prasetyo and Soliman 2021] Prasetyo, Y. T. and Soliman, K. O. S. (2021). Usability evaluation of erp systems: A comparison between sap s/4 hana & oracle cloud. In *2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 120–125. IEEE.
- [Preece et al. 2015] Preece, J., Sharp, H., and Rogers, Y. (2015). *Interaction design: beyond human-computer interaction*. John Wiley & Sons.
- [Rivero and Conte 2017] Rivero, L. and Conte, T. (2017). A systematic mapping study on research contributions on ux evaluation technologies. In *Proceedings of the XVI Brazilian symposium on human factors in computing systems*, pages 1–10.
- [Rubin and Chisnell 2008] Rubin, J. and Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley & Sons.
- [Saffer 2006] Saffer, D. (2006). Design for interaction.
- [Scroferneker 2021] Scroferneker, R. F. (2021). Aplicando testes a/b em aplicativos.
- [Souza and Silva 2020] Souza, R. F. and Silva, D. R. (2020). Usability test metrics: A systematic literature review. *International Journal of Computer Science and Information Security (IJCSIS)*, 18(2):1–16.
- [Speicher et al. 2014] Speicher, M., Both, A., and Gaedke, M. (2014). Wappu: Usability-based a/b testing. In *International Conference on Web Engineering*, pages 545–549. Springer.
- [Vanderdonckt et al. 2019] Vanderdonckt, J., Zen, M., and Vataavu, R.-D. (2019). Ab4web: An on-line a/b tester for comparing user interface design alternatives. *Proceedings of the ACM on Human-Computer Interaction*, 3(EICS):1–28.
- [Vermeeren et al. 2010] Vermeeren, A. P., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*, pages 521–530.
- [Young 2014] Young, S. W. (2014). Improving library user experience with a/b testing: Principles and process. *Weave: Journal of Library User Experience*, 1(1).