
**Confiabilidade de *Software* e Qualidade da Experiência do
Usuário: Um Estudo Quanti-Qualitativo**

Lucas Miranda Assunção



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FACULDADE DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia

2019

Lucas Miranda Assunção

**Confiabilidade de Software e Qualidade da Experiência do
Usuário: Um Estudo Quanti-Qualitativo**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Rivalino Matias Júnior

Uberlândia

2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

A851c Assunção, Lucas Miranda, 1994-
2023 Confiabilidade de software e qualidade da experiência do usuário
[recurso eletrônico] : um estudo quanti-qualitativo / Lucas Miranda
Assunção. - 2023.

Orientador: Rivalino Matias Júnior.
Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2023.7117>
Inclui bibliografia.

1. Computação. I. Matias Júnior, Rivalino, 1971-, (Orient.). II.
Universidade Federal de Uberlândia. Programa de Pós-Graduação em
Ciência da Computação. III. Título.

CDU: 681.3

Glória Aparecida
Bibliotecária Documentalista - CRB-6/2047



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



Ata da defesa de DISSERTAÇÃO DE MESTRADO junto ao Programa de Pós-graduação em Ciência da Computação da Faculdade de Computação da Universidade Federal de Uberlândia.

Defesa de Dissertação de Mestrado Acadêmico: PPGCO -004/2019

Data: 28 de fevereiro de 2019

Hora de início: 15 h 00 min

Discente: Lucas Miranda Assunção Matrícula: 11712CCP008

Título do Trabalho: Confiabilidade de Software e Qualidade da Experiência do Usuário: Um Estudo Quanti-Qualitativo.

Área de concentração: Ciência da Computação

Linha de pesquisa: Engenharia de Software

Reuniu-se na sala 1B132, Bloco 1B, Campus Santa Mônica da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Computação assim composta: Professores doutores: Joaquim Carlos Rossini – IP/UFU, Luciana Aparecida Martinez Zaina – DCOMP/UFSCAR Sorocaba, Avelino Francisco Zorzo – INF/PUCRS, Catello Di Martino – BELL LABS-NOKIA e Rivalino Matias Junior, orientador do candidato.

Ressalta-se que a Prof.ª Dr.ª Luciana Aparecida Martinez Zaina participou da defesa por meio de vídeo conferência desde a cidade de São Carlos-SP, o Prof. Dr. Avelino Francisco Zorzo da cidade de Porto Alegre-RS. Os demais membros da banca e o aluno participaram *in loco*.

Iniciando os trabalhos o presidente da mesa Prof. Dr. Rivalino Matias Júnior apresentou a Banca Examinadora e o candidato, agradeceu a presença do público e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu os conceitos finais.

Em face do resultado obtido, a Banca Examinadora considerou o candidato **aprovado**.

Esta defesa de Dissertação de Mestrado Acadêmico é parte dos requisitos necessários à obtenção do título de Mestre. O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, legislação e regulamentação interna da Universidade Federal de Uberlândia.

Nada mais havendo a tratar foram encerrados os trabalhos às 17 horas e 10 minutos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora

Participou por meio de vídeo conferência

Prof.ª Dr.ª Luciana Aparecida Martinez Zaina
DCOMP/UFSCAR-Sorocaba

Dr. Catello Di Martino
BELL LABS-NOKIA

Participou por meio de vídeo conferência

Prof. Dr. Avelino Francisco Zorzo
INF/PUCRS

Prof. Dr. Joaquim Carlos Rossini
IP/UFU

Prof. Dr. Rivalino Matias Júnior
FACOM/UFU (Orientador)

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “**Confiabilidade de *Software* e Qualidade da Experiência do Usuário: Um Estudo Quanti-Qualitativo**” por Lucas Miranda Assunção como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 28 de fevereiro de 2019.

Orientador:

Prof. Dr. Rivalino Matias Júnior

Universidade Federal de Uberlândia

Banca examinadora:

Prof. Dr. Joaquim Carlos Rossini

Universidade Federal de Uberlândia

Prof. Dr. Avelino Francisco Zorzo

Pontifícia Universidade Católica do Rio Grande do Sul

Dr. Catello Di Martino

Nokia Bell Labs

Prof(a). Dr(a). Luciana Aparecida Martinez Zaina

Universidade Federal de São Carlos

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Data: 28 de fevereiro de 2019.

Autor: Lucas Miranda Assunção
Título: Confiabilidade de *Software* e Qualidade da Experiência do Usuário: Um Estudo Quanti-Qualitativo
Faculdade: Faculdade de Computação
Grau: Mestrado

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão de cópias deste documento para propósitos exclusivamente acadêmicos, desde que o autor seja devidamente informado.

Autor

O AUTOR RESERVA PARA SI QUALQUER OUTRO DIREITO DE PUBLICAÇÃO DESTE DOCUMENTO, NÃO PODENDO O MESMO SER IMPRESSO E REPRODUZIDO, SEJA NA TOTALIDADE OU EM PARTES, SEM A PERMISSÃO ESCRITA DO AUTOR.

© Todos os direitos reservados a Lucas Miranda Assunção

*Dedico este trabalho à minha mãe,
aos meus familiares e amigos.*

Agradecimentos

Ao meu orientador pela paciência, compreensão e seriedade durante os processos de elaboração deste trabalho.

Aos membros da banca que gentilmente aceitaram avaliar este trabalho, em especial ao Dr. Joaquim Carlos Rossini, que sem a ajuda eu não teria conseguido terminar este trabalho.

A toda a equipe de funcionários da UFU, em especial ao Erisvaldo e a Sônia.

A todos aqueles que dedicaram um pouco do seu tempo respondendo aos questionários, me auxiliando a convidar participantes ou me dando forças durante a pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“There is no knowledge that is not power.”

Ralph Waldo Emerson

Resumo

Estudos em confiabilidade de *software* têm predominantemente focado em avaliar apenas de forma quantitativa a confiabilidade dos sistemas, em especial aplicações de usuário. Entretanto, tal enfoque pode não representar a real *Quality of Experience* (QoE) dos usuários. Para aprimorar a QoE dos usuários é importante entender como eles percebem as falhas de sistemas. Neste trabalho foram realizadas análises de confiabilidade de *software* considerando fatores quantitativos e qualitativos, tanto relacionados com a *Quality of Service* (QoS) quanto com a QoE dos usuários. Para avaliar a QoE dos usuários em relação à confiabilidade do *software* investigado, foi elaborado um questionário com cenários descrevendo situações hipotéticas, porém realísticas, de uso do computador em diversas situações em que ocorrem falhas de *software*. Através das perguntas do questionário foi avaliado o quanto a confiabilidade de *software* estimada por métodos clássicos (baseados em QoS), aproxima-se da percepção dos usuários quanto a confiabilidade do *software* analisado (QoE). O mesmo questionário foi aplicado tanto de forma *online* quanto presencial, objetivando comparar os resultados de ambas abordagens (orientadas a QoS e a QoE). Para tal, foram investigados os fatores de influência mais relevantes na QoE dos usuários quanto à confiabilidade de *software* para os cenários de uso estudados, utilizando dez dimensões emocionais. Realizando uma análise fatorial, observou-se que as dez dimensões avaliadas estavam relacionadas com apenas um fator, sendo as dimensões Estresse, Pontualidade e Paciência as três dimensões de maior relevância. Utilizando tais dimensões, foi proposto um modelo analítico para estimar a confiabilidade de *software* baseado em QoE, bem como foi ajustado um modelo analítico utilizando as dimensões mais relevantes apontadas pelo Teste t. Os resultados mostram que as dimensões mais relevantes indicadas pelo Teste t (Satisfação e Tolerância) compuseram o modelo analítico que apresentou o resultado mais próximo à percepção dos usuários em relação à confiabilidade do *software*. Além disso, observou-se que o *Mean Time Between Failures* (MTBF) e a consequência gerada pela ocorrência de uma falha são fatores de influência de QoE que afetam negativamente a percepção dos usuários em relação à confiabilidade de *software*. Por fim, conclui-se que o resultado do cálculo da confiabilidade de um *software* considerando apenas aspectos de QoS não é compatível com a QoE dos usuários em relação à confiabilidade do *software* usado.

Palavras-chave: confiabilidade de *software*, qualidade da experiência do usuário, sistemas operacionais.

Abstract

Software reliability studies have predominantly focused on only quantitatively evaluating system reliability, especially user applications. However, such an approach may not represent the actual Quality of Experience (QoE) of users. To improve users' QoE, it is important to understand how they perceive system failures. In this work, software reliability analyzes were performed considering quantitative and qualitative factors, both related to Quality of Service (QoS) and users' QoE. In order to evaluate the users' QoE regarding the reliability of the investigated software, a questionnaire was elaborated with scenarios describing hypothetical but realistic situations of computer use in several situations in which software failures occur. Through the questionnaire questions, it was evaluated the reliability of software estimated by classical QoS-based methods, which approximates the users' perception of the reliability of the analyzed software (QoE). The same questionnaire was applied both online and in person, aiming to compare the results of both approaches (oriented to QoS and QoE). In order to do this, we investigated the most relevant influence factors in the users' QoE regarding software reliability for the use scenarios studied, using ten emotional dimensions. By performing a factorial analysis, it was observed that the ten dimensions evaluated were related to only one factor, the dimensions of stress, punctuality and patience being the three dimensions of greater relevance. Using such dimensions, an analytical model was proposed to estimate the reliability of software based on QoE, as well as an analytical model was adjusted using the most relevant dimensions pointed out by Test t. The results show that the most relevant dimensions indicated by the t-Test (Satisfaction and Tolerance) were the analytical model that presented the result closest to the users' perception regarding software reliability. In addition, it has been observed that the Mean Time Between Failures (MTBF) and the consequence generated by the occurrence of a failure are factors of influence of QoE that negatively affect the perception of the users in relation to the reliability of software. Finally, it was observed that the result of calculating the reliability of a software considering only aspects of QoS is not compatible with the QoE of the users in relation to the reliability of the software used.

Keywords: software reliability, user quality of experience, operating systems.

Lista de Ilustrações

Figura 1 - Contexto de estudo desta dissertação em relação aos temas: QoE, UX, HCI e Confiabilidade de <i>software</i>	36
Figura 2 - Fluxo das etapas realizadas no método	38
Figura 2 - Perguntas da primeira etapa do questionário	60
Figura 3 - Primeira pergunta dos cenários de falha (segunda etapa).....	60
Figura 4 - Segunda pergunta dos cenários de falha (segunda etapa).....	60
Figura 5 - Tela simulada da interrupção na inicialização do Windows.....	62
Figura 6 - Ícone de pasta no desktop que executa o <i>script</i>	63
Figura 7 - Mensagem de falha da aplicação <i>explorer.exe</i>	63
Figura 8 - Bloco de notas simulado antes da ocorrência da falha	64
Figura 9 - Bloco de notas simulado após a ocorrência da falha	64
Figura 10 - BSOD simulada para a falha BAD_POOL_CALLER	65
Figura 11 - Equipamentos utilizados nas sessões de coleta de dados do questionário presencial	66
Figura 14 - Gráfico de escarpa para os fatores extraídos (questionário <i>online</i>).....	78
Figura 15 - Intervalo de confiança de 95% para os resultados dos cálculos de confiabilidade (questionário <i>online</i>).....	83
Figura 16 - Médias dos escores das dimensões por cenário (questionário <i>online</i>).....	85
Figura 17 - Medianas dos escores das dimensões por cenário (questionário <i>online</i>).....	85
Figura 18 - Percentual de participantes por grau de escolaridade (questionário presencial) ...	86
Figura 19 - Gráfico de escarpa para os fatores extraídos (questionário presencial).....	89
Figura 20 - Intervalo de confiança de 95% para os resultados dos cálculos de confiabilidade (questionário presencial).....	93
Figura 21 - Médias dos escores das dimensões por cenário (questionário presencial)	95
Figura 22 - Medianas dos escores das dimensões por cenário (questionário presencial).....	96
Figura A.1 - Cenário 1	101
Figura A.2 - Cenário 2	101
Figura A.3 - Cenário 3	102
Figura A.4 - Cenário 4	102
Figura B.1 - Nuvem de palavras produzida com as respostas da categoria Consequência	104
Figura B.2 - Nuvem de palavras produzida com as respostas da categoria Contexto.....	104
Figura B.3 - Nuvem de palavras produzida com as respostas da categoria Contexto e Consequência.....	105
Figura B.4 - Nuvem de palavras produzida com as respostas da categoria Evento de Falha	105
Figura B.5 - Nuvem de palavras produzida com as respostas da categoria Expectativa de Solução	106

Figura B.6 - Nuvem de palavras produzida com as respostas da categoria Familiaridade	106
Figura B.7 - Nuvem de palavras produzida com as respostas da categoria Não Familiaridade	107
Figura B.8 - Nuvem de palavras produzida com as respostas da categoria Frequência de Falha	107
Figura B.9 - Nuvem de palavras produzida com as respostas da categoria Mensagem de Falha	108
Figura B.10 - Nuvem de palavras produzida com as respostas da categoria Recuperação	108

Lista de tabelas

Tabela 1 – Distinção dos cenários de falha do questionário.....	58
Tabela 2 - Dimensões e escalas utilizadas no questionário	60
Tabela 3 - Principais diferenças entre o questionário <i>online</i> e o questionário presencial	61
Tabela 4 - Percentual de participantes por gênero (questionário <i>online</i>)	74
Tabela 5 - Percentual de participantes por faixa etária (questionário <i>online</i>)	74
Tabela 6 – Percentual de participantes por área de atividade (questionário <i>online</i>)	75
Tabela 7 - Estatísticas descritivas da confiabilidade percebida pelos participantes (questionário <i>online</i>).....	76
Tabela 8 – Porcentagem de variância explicada por fator extraído (questionário <i>online</i>)	77
Tabela 9 – Valor de comunalidade por dimensão (questionário <i>online</i>).....	78
Tabela 10 – Resumo estatístico do Modelo 1	79
Tabela 11 - Nível de significância no Teste t das dez dimensões utilizadas na primeira etapa de ajuste do Modelo 2 (questionário <i>online</i>)	80
Tabela 12 – Resumo estatístico do Modelo 2.....	81
Tabela 13 - Comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelos Modelos 1 e 2 (questionário <i>online</i>).....	82
Tabela 14 – Cenários sem diferença estatística de acordo com o teste de Kruskal-Wallis (questionário <i>online</i>).....	84
Tabela 15 - Percentual de participantes por gênero (questionário presencial)	86
Tabela 16 - Percentual de participantes por faixa etária (questionário presencial)	86
Tabela 18 - Estatísticas descritivas da confiabilidade percebida pelos participantes (questionário presencial).....	88
Tabela 19 - Porcentagem de variância explicada por fator (questionário presencial).....	89
Tabela 20 - Valor de comunalidade por dimensão (questionário presencial)	90
Tabela 21 – Resumo estatístico do Modelo 3.....	91
Tabela 22 - Nível de significância no Teste t das dez dimensões utilizadas na primeira etapa de ajuste do Modelo 4 (questionário presencial)	91
Tabela 23 - Resumo estatístico do Modelo 4	92
Tabela 24 - Comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelos Modelos 3 e 4 (questionário presencial).....	93
Tabela 25 – Comparação entre os cenários por dimensão (questionário presencial).....	94
Tabela C.1 - Participantes descartados da amostra do questionário presencial.....	115

Lista de abreviaturas, siglas e notação

BSOD	<i>Blue Screen of Death</i>
DLLs	<i>Dynamic-link Libraries</i>
DW	Durbin-Watson
GL	Graus de Liberdade
GUI	<i>Graphic User Interface</i>
HCI	<i>Human Computer Interaction</i>
ISO	International Organization for Standardization
KMO	Kaiser-Meyer-Olkin
KQI	<i>Key Quality Indicators</i>
MOS	<i>Mean Opinion Score</i>
MTBF	<i>Mean Time Between Failures</i>
QoE	<i>Quality of Experience</i>
QoS	<i>Quality of Service</i>
$R_{(t)}$	Confiabilidade estimada por método clássico (análise quantitativa)
$R_{Per}(t)$	Confiabilidade percebida pelos usuários (análise qualitativa)
$\bar{R}_{Per}(t)$	Confiabilidade percebida pelos usuários (média)
$\tilde{R}_{Per}(t)$	Confiabilidade percebida pelos usuários (mediana)
$\check{R}_{Per}(t)$	Confiabilidade percebida pelos usuários (moda)
$R_{Per}^{dp}(t)$	Confiabilidade percebida pelos usuários (desvio padrão)
$R_{Per}^{máx}(t)$	Confiabilidade percebida pelos usuários (máximo)
$R_{Per}^{mín}(t)$	Confiabilidade percebida pelos usuários (mínimo)

$R_{Pred_i}(t)$	Confiabilidade predita pelo modelo i (análise quanti-qualitativa)
SO	Sistema Operacional
Tol	<i>Tolerance</i>
UX	<i>User Experience</i>
VIF	<i>Variance Inflation Factor</i>
VoIP	<i>Voice over Internet Protocol</i>
WFL	<i>Weber-Fechner Law</i>

Sumário

Introdução.....	34
1.1 Fluxo de Etapas do Método	36
1.2 Motivação	38
1.3 Objetivos da Pesquisa	40
1.3.1 Objetivo Geral	40
1.3.2 Objetivos Específicos	40
1.4 Perguntas de Pesquisa e Hipóteses	40
1.5 Contribuições	40
1.6 Organização da Dissertação	40
Fundamentação Teórica.....	42
2.1 Experiência do Usuário.....	42
2.2 Avaliação da Qualidade da Experiência do Usuário	44
2.4 Avaliação da Confiabilidade de <i>Software</i> Considerando Aspectos de QoE.....	51
2.5 Síntese dos Trabalhos	53
Metodologia.....	56
3.1 Introdução	56
3.2 Materiais	56
3.2.1 Terminologia	56
3.2.2 Questionário	57
3.2.3 Instrumental	61
3.2.4 Participantes	65
3.2.5 Ferramentas	65
3.3 Método	66
3.3.1 Teste Piloto	66
3.3.2 Recrutamento de Participantes	66
3.3.3 Análise dos Dados	67

Resultados.....	74
4.1 Questionário <i>online</i>	74
4.1.1 Etapa 1	74
4.1.2 Etapa 2	75
4.2 Questionário presencial.....	86
4.2.1 Etapa 1	86
4.2.2 Etapa 2	87
4.3 Validação com amostra de modelagem e amostra de teste.....	96
Conclusão	99
5.1 Síntese dos Resultados.....	99
5.2 Resposta para as Perguntas de Pesquisa	101
5.3 Ameaças à Validade e Limitações da Pesquisa	102
5.4 Dificuldades Encontradas	103
5.5 Trabalhos Futuros	103
Referências	106
Cenários do Questionário	112
Nuvens de Palavras.....	115
Participantes Descartados	121
Parecer do Comitê de Ética em Pesquisas com seres humanos da Universidade Federal de Uberlândia	125

Introdução

A relevância da computação no cotidiano da humanidade é uma unanimidade. Os sistemas de computação são utilizados para diversas finalidades, tais como: entretenimento, pesquisa, gerenciamento de processos, entre várias outras (MATIAS *et al.*, 2014). Dada tal relevância, é indispensável que esses sistemas consigam executar, com sucesso, as tarefas para as quais foram desenvolvidos. A fim de atingir tal objetivo, um atributo essencial a tais sistemas é a confiabilidade (AVIZIENIS *et al.* 2004).

Confiabilidade é “a capacidade de um sistema ou componente de executar suas funções requeridas nas condições indicadas por um período de tempo especificado”¹ (AMERICAN NATIONAL STANDARDS INSTITUTE; INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, 1991, p.64, tradução nossa). Em (LEWIS, 1996), a confiabilidade é definida matematicamente como:

$$R_{(t)} = P(T > t), t \geq 0 \quad (1)$$

em que T é uma variável aleatória denotando o tempo até a falha. Portanto, a confiabilidade $R_{(t)}$ é a probabilidade de não ocorrer falhas no sistema no período de tempo de 0 até t . Neste trabalho foi adotada a taxonomia apresentada em (AVIZIENIS *et al.* 2004) que define falha, erro e falta respectivamente como: “**falha** é o desvio do serviço (resultado) entregue pelo sistema com relação à sua especificação. **Erro** é aquela parte do estado interno do sistema que pode levar a ocorrência de uma falha. A causa de um erro é a ativação de uma **falta**”². (AVIZIENIS *et al.* 2004, p.4, tradução nossa).

A confiabilidade de *software* é um tema que possui considerável embasamento quantitativo. Os métodos de análise de confiabilidade disponíveis na literatura, em especial em confiabilidade de *software*, praticamente não consideram aspectos qualitativos em suas abordagens. Esses aspectos qualitativos referem-se à *Quality of Experience* (QoE), que é definida como:

O grau de prazer ou aborrecimento de uma pessoa cuja experiência envolve um aplicativo, serviço ou sistema. Resulta da avaliação da pessoa sobre o cumprimento de suas expectativas e necessidades com relação à utilidade e/ou prazer à luz do

¹ The ability of a system or component to perform its required functions under stated conditions for a specified period of time.

² A service failure, often abbreviated here to failure, is an event that occurs when the delivered service deviates from correct service. Since a service is a sequence of the system’s external states, a service failure means that at least one (or more) external state of the system deviates from the correct service state. The deviation is called an error. The adjudged or hypothesized cause of an error is called a fault.

contexto, personalidade e estado atual da pessoa (MOLLER; RAAKE, 2014, p. 19, tradução nossa)³.

Um dos principais fatores que influenciam a QoE do usuário é a *Quality of Service* (QoS). No âmbito de telecomunicações, QoS é definida como “totalidade das características de um serviço de telecomunicações que tem sua capacidade de satisfazer as necessidades declaradas e implícitas do usuário do serviço”⁴ (INTERNATIONAL TELECOMMUNICATION UNION, 2008, p.3, tradução nossa). A QoS, em telecomunicações, é avaliada por meio de métricas tais como: atraso na transmissão de dados (*delay*), variação no atraso (*jitter*), taxa de perda de pacotes, largura de banda, tempo de configuração da conexão, entre outras. Diferente da QoS, que é orientada ao serviço, a QoE é a percepção do usuário em relação ao serviço prestado. Por exemplo, um usuário contrata uma rede de comunicação de dados com largura de banda de 100 Mbps e, apesar da QoS entregue se apresentar alta (ex. 98 Mbps), ou seja, próxima ou igual ao contratado, o usuário pode não ter uma boa experiência ao usar esta rede (ex. ao realizar *downloads* em seu *browser*), por razões variadas. A QoS é estimada por meio de indicadores de sistema, enquanto a QoE considera fatores não técnicos, tais como fatores humanos, de contexto, entre outros. Portanto, nesse trabalho, a análise quantitativa de confiabilidade é referente ao cálculo de confiabilidade considerando apenas fatores de QoS, enquanto que a análise qualitativa de confiabilidade utiliza não apenas fatores de QoS, mas também leva em consideração os aspectos de QoE. A QoE dos usuários neste trabalho foi avaliada através da análise da percepção deles perante falhas de *software*, no qual o *software* avaliado foi o SO do computador. Portanto, a análise foi realizada com foco em aspectos **hedônicos** do sistema, ou seja, considerando o bem-estar psicológico do usuário ao utilizar o sistema (MOLLER; RAAKE, 2014). Na análise também foram levados em consideração conceitos de *Human Computer Interaction* (HCI) e *User Experience* (UX).

A HCI é um campo de estudo que se concentra na interação entre seres humanos e sistemas de computador, especialmente em como as pessoas interagem com interfaces de *software* e *hardware*. Abrange o *design*, a avaliação e a implementação de sistemas que sejam eficazes, eficientes e agradáveis de usar. A HCI se preocupa principalmente com a usabilidade, a acessibilidade e a experiência do usuário ao interagir com um sistema (MOLLER; RAAKE, 2014). A UX é uma disciplina dentro da HCI que se concentra especificamente na experiência geral do usuário ao interagir com um produto ou sistema, levando em consideração aspectos emocionais, cognitivos e funcionais. Inclui elementos como *design* de interfaces, arquitetura da informação, usabilidade e a sensação geral que um usuário tem ao usar um produto ou serviço. O objetivo da UX é criar produtos que sejam agradáveis, úteis e que atendam às necessidades e expectativas dos usuários (MOLLER; RAAKE, 2014).

³ Quality of experience is the degree of delight or annoyance of a person whose experiencing involves an application, service, or system. It results from the person’s evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person’s context, personality and current state.

⁴ Totality of characteristics of a telecommunication service that bear on its ability to satisfy stated and implied need of the user of the service.

Nesta dissertação a *Human Computer Interaction* (HCI) foi utilizada como um instrumento que auxiliou na avaliação da QoE, diferente de trabalhos na área de HCI em que o foco principal é verificar a qualidade da HCI em si. A *User Experience* (UX) é avaliada considerando não somente a qualidade da HCI, mas também o *design* das aplicações, sua usabilidade⁵ e o quão agradável é para o usuário utilizar o sistema. Portanto, para realizar a análise da confiabilidade de *software* apresentada neste trabalho, foram utilizados conhecimentos das áreas HCI, UX, QoE e confiabilidade de *software*, como mostra a Figura 1.

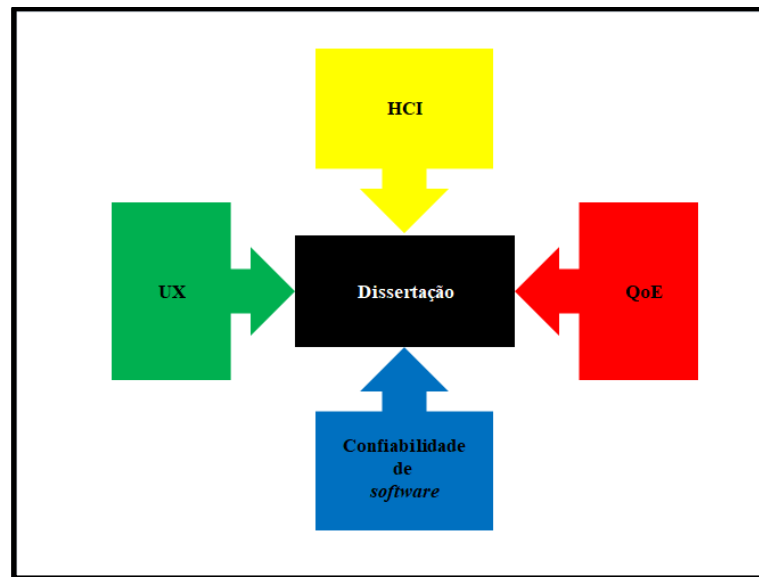


Figura 1 - Contexto de estudo desta dissertação em relação aos temas: QoE, UX, HCI e Confiabilidade de *software*

Por fim, ressalta-se que este estudo apresenta abordagens quanti-qualitativas por considerar dados numéricos e mensuráveis, empregando métodos estatísticos para identificar padrões, relações de causa e efeito e tendências (análise quantitativa). Já a análise qualitativa utiliza dados descritivos e observacionais, buscando compreender significados, interpretações e contextos subjacentes aos dados coletados. Enquanto a análise quantitativa se concentra em números, estatísticas e generalizações, a análise qualitativa explora contextos, busca compreensão profunda e contextual dos dados e busca por novas interpretações (FÁVERO *et al.*, 2009). No contexto deste trabalho, a análise quantitativa refere-se ao uso de análise estatística para quantificar a percepção dos usuários com relação à confiabilidade do *software*. A análise qualitativa realizada aborda como e quais fatores podem influenciar tal percepção (fatores como Tolerância, Satisfação, entre outros).

1.1 Fluxo de Etapas do Método

A Figura 2 mostra o fluxo de etapas desta dissertação de mestrado. Primeiramente, criou-se uma versão adaptada do dicionário de termos utilizado em (YABE *et al.*, 2018). Para

⁵ Neste trabalho, considerou-se a seguinte definição de usabilidade “a medida em que um produto pode ser usado por usuários especificados para atingir metas especificadas com eficácia, eficiência e satisfação em um contexto especificado de uso” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2010, p.9, tradução nossa)

tal, foram elaboradas nuvens de palavras com as respostas dos participantes do estudo prévio realizado em (YABE *et al.*, 2018). Conforme é relatado na Seção 3.2.2., tais respostas foram classificadas em 10 categorias: Consequência, Contexto, Contexto e Consequência, Evento de Falha, Expectativa de Solução, Familiaridade, Não Familiaridade, Frequência da Falha, Mensagem de Falha e Recuperação. Essa classificação criada usando as nuvens de palavras auxiliou na elaboração das perguntas de pesquisa e do questionário.

Após essa etapa, foram elaborados os cenários de falha que foram utilizados no questionário. Foi criado um questionário com quatro cenários descrevendo situações hipotéticas, porém realísticas, de uso do computador em que os usuários assumiram estar vivenciando as falhas da maneira como elas foram apresentadas em cada cenário. Para analisar o impacto causado pelas diferentes falhas de SO na QoE do usuário, os cenários foram divididos em “baixo impacto” e “alto impacto”. Além disso, cada cenário apresentou um valor curto ou longo de MTBF (*Mean Time Between Failures*), para que seja possível avaliar como a frequência de falhas afeta a QoE do usuário. Os quatro cenários de falha apresentam as seguintes características: a descrição da tarefa (interação usuário-computador) realizada no cenário, como a falha acontece durante a realização da tarefa, a consequência gerada pela ocorrência da falha (que determina o impacto baixo ou alto), uma imagem mostrando a mensagem de falha gerada e uma frase relatando o MTBF. Os participantes responderam a duas perguntas para cada cenário do questionário. Para a primeira pergunta, eles responderam em intervalo de 0 (sem confiança alguma) até 100 (confiança total), qual o seu grau de confiança no SO descrito no cenário apresentado (Figura 3). A segunda pergunta foi elaborada com o intuito de medir as dimensões emocionais do participante em cada cenário, através de uma adaptação de escalas de diferencial semântico (PASQUALI, 2010).

Com os cenários elaborados, foi realizado um teste piloto no qual aplicou-se o questionário de forma presencial para cinco pessoas diferentes para aprimorar a execução dessa etapa da pesquisa (ver Seção 3.3.1). Após a realização do teste piloto, a aplicação do questionário *online* foi realizada. Enquanto o questionário *online* ficou disponível para ser respondido, foi feito o ajuste do instrumental que simulou as falhas descritas nos cenários do questionário *online* para que esse mesmo questionário também fosse aplicado de forma presencial, com os participantes vivenciando as falhas durante o uso do computador (ver Seção 3.2.3). Terminado o ajuste do instrumental e da coleta de dados *online*, foi realizada a aplicação do questionário de forma presencial.

Finalizada a coleta de dados *online* e presencial, foi possível avaliar a percepção dos usuários em relação à confiabilidade do *software* (ver Figura 3) calculando os intervalos de confiança para cada cenário. O Teste de Esfericidade de Bartlett e o Teste de *KMO* foram aplicados para verificar a adequação dos dados à análise fatorial. Com a adequação dos dados à análise fatorial verificada, foi realizada uma análise de componentes principais para avaliar a relação entre as dimensões utilizadas (ver Tabela 2) e os fatores retidos, resumindo tais dimensões no menor número de fatores possível. A retenção de fatores foi avaliada através das técnicas: o critério da raiz latente, o critério de porcentagem de variância e o critério do gráfico de escarpa. Também foi feita a análise fatorial dos dados, o ajuste e avaliação dos modelos. Para ajustar os modelos foi utilizada a técnica de regressão linear múltipla (ver

Seção 3.3.2.2) e para fazer sua avaliação foram utilizadas as técnicas: R^2 e R^2 ajustado, Teste t, Teste de Durbin-Watson, *VIF* e *Tolerance*. Por fim, para verificar se houve ou não diferença significativa nas respostas das dimensões entre os cenários foi utilizado o Teste de Kruskal-Wallis (ver Seção 3.3.2.3).

Depois da análise dos dados, concluiu-se que o resultado do cálculo da confiabilidade de um *software* considerando apenas aspectos de QoS não é compatível com a QoE dos usuários em relação à confiabilidade do *software* usado. Também foi observado que o MTBF é um fator de influência de QoE, afetando negativamente a percepção dos usuários em relação à confiabilidade de *software*, sendo que um valor de MTBF longo afeta mais que um valor de MTBF curto. Outro ponto observado é que os modelos ajustados que apresentaram maior precisão na predição de confiabilidade tiveram em comum a dimensão Satisfação, o que mostra que para elaborar modelos de predição de confiabilidade de *software* que considerem a percepção dos usuários deve-se levar em consideração a satisfação deles ao utilizar o sistema. Demais conclusões encontram-se no Capítulo 5 desta dissertação.

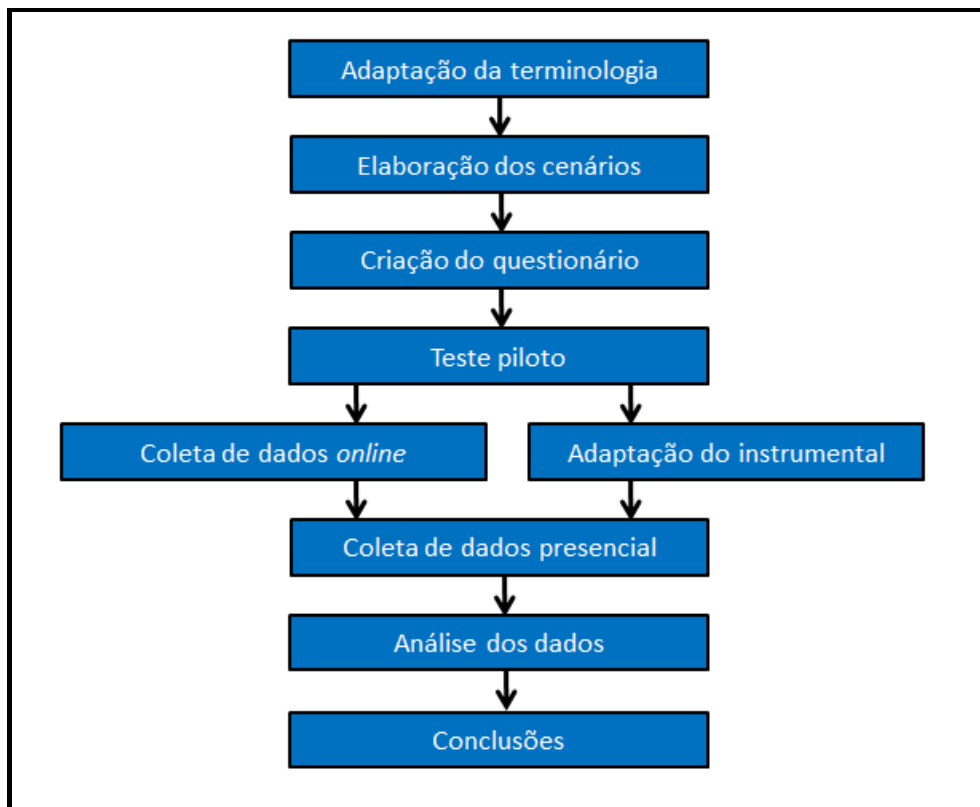


Figura 2 - Fluxo das etapas realizadas no método

No restante deste capítulo são apresentadas a motivação, os objetivos, as perguntas e hipóteses de pesquisa e as contribuições do trabalho.

1.2 Motivação

Dentre os diversos atributos de qualidade de *software* existentes (funcionalidade, usabilidade, entre outros), a confiabilidade é considerada um dos mais importantes por analisar as falhas de *software* que são responsáveis pela maior parte da inatividade dos

sistemas computacionais (Lyu, 2007). Relacionar a confiabilidade do *software* com a QoE do usuário é importante para que se consiga observar o impacto que a confiabilidade do sistema tem sobre a *User Experience* (UX). Alguns trabalhos acadêmicos relatam o impacto de falhas no usuário, bem como o que é relevante para ele em relação a *software*. Em (JALOTE *et al.* 2004) é mencionado que muitas vezes há múltiplos grupos de usuários que usam o mesmo produto de *software* de maneiras diferentes e, portanto, o impacto das falhas varia entre esses diferentes grupos. Atualmente, pesquisas sobre qualidade de *software* estão focadas na satisfação dos usuários e suas emoções durante o uso de um sistema (AL-BADAREEN *et al.* 2011), o que reforça que é relevante considerar a opinião dos usuários nos diversos tipos de análises de *software*.

Uma das perspectivas que vem ganhando destaque em pesquisas qualitativas em diferentes áreas da computação é a análise da QoE, não se restringindo apenas aos fatores de QoS (REICHL; TUFFIN; SCHATZ, 2013). Considerando esses dois aspectos, QoS e QoE, no tocante à confiabilidade de *software*, em (MATIAS; OLIVEIRA; ARAÚJO, 2013) os autores introduzem a problemática de que modelos de análise de confiabilidade de *software* tradicionais podem não representar a real percepção dos usuários em termos de confiabilidade. Desse modo, incorporar fatores de QoE no processo de estimação da confiabilidade de *software* pode ser uma forma de estimar a confiabilidade do *software* tendo como base não apenas fatores de QoS, mas também de QoE representados pela percepção do usuário em relação à confiabilidade do produto de *software* sendo analisado.

Na maioria dos sistemas computacionais, a camada de *software* subdivide-se em programas de aplicação e o *software* de sistema operacional (SO). Considerando que a execução das aplicações depende do SO, mesmo que um sistema computacional tenha um *hardware* e aplicações altamente confiáveis, sua confiabilidade será prejudicada se o SO não apresentar a confiabilidade necessária (MATIAS; ANTUNES, 2014). Portanto, como a confiabilidade dos sistemas operacionais afeta diretamente a confiabilidade dos sistemas computacionais como um todo, as hipóteses deste trabalho são avaliadas por meio de um estudo de caso que envolve dados de falha do SO Microsoft Windows 7. Optou-se por usar esse *software* por ele ter sido o SO de propósito geral mais utilizado no mundo durante os anos de 2017 e 2018 (NETMARKETSHARE, 2018), período em que esse estudo foi realizado. Contudo, é importante salientar que a abordagem de modelagem proposta neste trabalho é genérica suficiente para ser utilizada para outras categorias de *software*.

Ao revisar o estado da arte em confiabilidade de *software*, notou-se uma ausência de estudos voltados para a análise da confiabilidade de *software* considerando a QoE do usuário. Apenas um trabalho diretamente relacionado foi encontrado, o qual faz uma análise de confiabilidade considerando fatores relacionados à UX (VEERARAGAVAN *et al.* 2016). Entretanto, o modelo desenvolvido pelos autores é específico para avaliar a confiabilidade de *softwares* tele imersivos. Além disso, UX e QoE possuem diferenças que são discutidas na Seção 2.1. As diferenças do trabalho supracitado com os propósitos desta pesquisa são discutidas na Seção 2.4.

1.3 Objetivos da Pesquisa

1.3.1 Objetivo Geral

- Análise da confiabilidade de *software* estimada com fatores quantitativos e qualitativos relacionados à QoE do usuário.

1.3.2 Objetivos Específicos

- Investigar as relações entre confiabilidade de *software* e QoE do usuário.
- Propor um modelo para estimar a confiabilidade de um *software* seguindo uma abordagem quanti-qualitativa considerando fatores de QoE.
- Comparar o desempenho do modelo proposto com métodos clássicos de estimar a confiabilidade de *software*, os quais são baseados em fatores de QoS.

1.4 Perguntas de Pesquisa e Hipóteses

Pergunta 01: Fatores que influenciam a QoE do usuário afetam a percepção do usuário em relação à confiabilidade de *software*?

Hipótese: Conjectura-se que tais fatores afetam a percepção do usuário em relação à confiabilidade do *software*, da mesma maneira que a confiabilidade do ponto de vista do usuário pode ser maior caso a falha aconteça, mas ele não a percebe; em termos de QoE do usuário, a confiabilidade percebida pode ser menor do que a confiabilidade calculada.

Pergunta 02: A confiabilidade de *software* calculada seguindo a abordagem clássica (estado da arte), baseada em QoS, é compatível com a real percepção dos usuários sobre a confiabilidade do *software* usado?

Hipótese: A hipótese de trabalho é que a confiabilidade do *software* calculada com base apenas em parâmetros de QoS nem sempre é representativa da real percepção de confiabilidade que os usuários têm em relação ao *software*, pois tais modelos não consideram aspectos qualitativos em suas abordagens (QoE).

1.5 Contribuições

Por meio dos resultados deste trabalho foi desenvolvida uma abordagem de análise de confiabilidade que considera aspectos qualitativos (QoE). Os achados com esta abordagem sugerem que é possível estimar melhor a confiabilidade de um sistema incorporando elementos da percepção dos usuários (QoE) quanto à sua confiança no funcionamento do sistema. Até o momento, não foram encontrados trabalhos na literatura que propusessem procedimentos de análise de confiabilidade de *software* considerando a percepção dos usuários.

1.6 Organização da Dissertação

Os demais capítulos desta dissertação estão organizados da seguinte forma:

- **Capítulo 2:** apresenta a fundamentação teórica utilizada neste trabalho, descrevendo conceitos de UX, QoE, confiabilidade de *software* e confiabilidade de sistemas operacionais.
- **Capítulo 3:** descreve a metodologia aplicada neste estudo, explicando os materiais e métodos utilizados.
- **Capítulo 4:** discute os resultados dos experimentos e análises realizadas neste trabalho.
- **Capítulo 5:** expõe as conclusões resumindo os principais resultados, respondendo as perguntas de pesquisa, as ameaças à validade e limitações do trabalho, dificuldades encontradas durante a pesquisa e trabalhos futuros.

Fundamentação Teórica

2.1 Experiência do Usuário

No trabalho desenvolvido em (HASSENZAHN; TRACTINSKY, 2006), os autores discutem três abordagens sobre a experiência do usuário. Na primeira abordagem eles afirmam que o objeto interativo tem uma função além da instrumental, pois estudos referentes à HCI se concentram em como o usuário é capaz de realizar determinada tarefa. Entretanto, não apenas de usabilidade se compõe a experiência do usuário. A segunda abordagem mostra a importância da emoção e do afeto em alguns processos centrais do usuário. A UX deve prestar atenção no afeto e nas emoções a partir da perspectiva do usuário, pois, dessa maneira, pode-se entender seu papel ao utilizar o sistema. Para os autores, o que é novo no termo UX é o foco em resultados emocionais positivos como, por exemplo, os sentimentos de diversão e orgulho. A terceira abordagem considera que uma experiência é uma combinação de vários elementos, como o produto e o estado interno do usuário (temperamento, expectativas, objetivos, entre outros). No artigo, os autores descrevem UX como “uma consequência do estado interno de um usuário (predisposições, expectativas, necessidades, motivação, humor, etc.), as características do sistema projetado (por exemplo, complexidade, propósito, facilidade de uso, funcionalidade, etc.) e o contexto (ou o ambiente), dentro do qual a interação ocorre (por exemplo, configuração organizacional/social, significado da atividade, voluntariedade de uso, entre outros)”. A principal contribuição do trabalho foi apresentar uma agenda para estudos em UX. Além disso, os autores sugerem que as pesquisas nessa área deveriam buscar aprimorar a qualidade de vida dos usuários ao invés de apenas reduzir os problemas existentes em UX.

Em (HASSENZAHN, 2007), o autor apresentou um modelo que presume que as pessoas compreendem a UX em dois eixos independentes: hedônico e pragmático. O primeiro eixo tem como foco a pessoa, seu estado interno e vai além do instrumental. Para o autor, esse eixo distingue três fatores: estimulação, identificação e evocação. A estimulação está ligada a novidade e ao crescimento pessoal, enquanto a identificação está conectada com a comunicação da própria identidade e a evocação está relacionada à provocação de memórias e simbolismos. Ou seja, o eixo hedônico está conectado aos objetivos de “ser”, ser divertido, ser relacionado a outros, ser especial. O segundo eixo se relaciona aos objetivos de “fazer”, “encontrar uma imagem”, “escrever um texto”, “criar uma planilha”. Esse eixo coloca foco no produto, utilidade, na eficiência e eficácia. A importância do trabalho foi em definir um modelo que apresenta como a UX é compreendida. Assim, pode-se concluir que a UX leva em consideração o tempo em que o usuário está envolvido com uma experiência, buscando compreender os seus fatores emocionais antes ou depois da interação. Como o usuário decide se irá utilizar uma aplicação ou não, deve-se atentar para sua satisfação.

Buscando compreender a relação entre a satisfação das necessidades dos usuários e sua percepção positiva dos produtos, em (HASSENZAHL; DIEFENBACH; GORITZ, 2010) os autores avaliaram 548 respondentes através de um questionário *online*. Os respondentes eram solicitados a relatarem uma experiência recente e extremamente positiva envolvendo o uso de tecnologia. Após essa etapa, eles respondiam uma série de escalas MOS (*Mean Opinion Score*) que avaliavam as seguintes características da experiência relatada: atendimento da necessidade, como a experiência foi afetada, atribuição, percepção do produto e avaliação. As respostas foram avaliadas através da análise de componentes principais (análise fatorial), utilizando a rotação *varimax* quando necessário. Os resultados encontrados sugerem que as experiências podem ser categorizadas pelas necessidades primárias dos usuários que elas atendem. Outro achado importante do estudo é relatar a possibilidade de explicar a relação entre a satisfação das necessidades dos usuários e a percepção positiva dos produtos através dos fatores: estimulação, *relatedness* (sentimento de se ter contato íntimo regular com pessoas que se importam com você), competência, e popularidade. Os autores abordaram diversos aspectos como ameaças à validade do estudo, tais como: a dependência da amostra para a obtenção dos resultados, a generalização dos fatores encontrados para todos os usuários e o fato da descrição da experiência dos respondentes ter sido textual, o que gerou uma amostra heterogênea das descrições em tamanho, estilo e detalhamento.

No trabalho realizado em (AL-BADAREEN *et al.* 2011), os autores apresentaram a qualidade de *software* do ponto de vista do usuário. Nesse estudo os autores citaram a funcionalidade, confiabilidade, desempenho, usabilidade e portabilidade como características consideradas pelo usuário e que afetam suas emoções. A confiabilidade apresenta a capacidade do *software* de executar as funções corretamente (sem falha). O desempenho é definido pelos autores como quantidade de trabalho útil obtido por um sistema em comparação com o tempo e recursos utilizados. O fator de desempenho do sistema é destinado a avaliar se a aplicação funciona de forma eficiente sob os recursos computacionais disponíveis. Além disso, eles classificaram a usabilidade dentro de três principais propriedades: características de interface, treinamento e capacidade de suporte operacional. A portabilidade foi definida como a capacidade de o *software* operar corretamente em diferentes tipos de plataformas. Com esse estudo os autores concluíram que a satisfação do usuário é fundamental para o sucesso de qualquer produto no mercado. Portanto, a principal contribuição desse trabalho foi apresentar os requisitos que precisam de mais atenção ao desenvolver um sistema que será aceito pelo usuário. Considerando esses conceitos encontrados na literatura, pode-se observar que a QoE do usuário com respeito ao uso do *software* repercute na sua percepção de qualidade do *software*.

Os trabalhos descritos até aqui são focados em UX. Apesar de esse tema possuir forte relação com QoE, as duas áreas possuem diferenças (MOLLER; RAAKE, 2014). Os estudos em UX geralmente são focados na interação humano-computador, enquanto que os trabalhos realizados em QoE são centrados nas tecnologias com que os usuários interagem, porém os trabalhos em UX também podem avaliar tecnologias. Outra diferença importante é que o tema UX possui um embasamento teórico mais consolidado que a área de QoE, que atualmente ainda é mais estudada empiricamente. Os objetos de estudo em UX geralmente são artefatos

em que o usuário interage através de uma interface ou serviço. Em QoE, usualmente são estudados sistemas de comunicação multimídia. Por fim, no âmbito mercadológico, os produtos e serviços prestados com foco em UX possuem como objetivo direto o lucro com a venda do produto/serviço em si. Por outro lado, do ponto de vista de QoE o lucro é obtido através do quanto o cliente está disposto a pagar pela melhora de sua QoE ao utilizar produtos e serviços. Por exemplo, um cliente que faz assinatura de determinado serviço de *streaming* precisa pagar mais para que a empresa provedora de conteúdo forneça mais recursos com o objetivo de melhorar a QoE percebida pelo cliente, através do aumento da variedade de conteúdo disponível a ser assistido, melhorando a transmissão de dados, disponibilizando mais funcionalidades para o cliente utilizar dentro da plataforma de interação, entre outros meios.

Em (HASSENZAHL, TRACTINSKY, 2006), os autores apresentaram uma agenda de estudos em UX, sem validar esse procedimento com um estudo de caso. Em (HASSENZAHL, 2007) os autores apresentaram um modelo que apresenta como a UX é compreendida, também sem mostrar uma aplicação prática desse modelo. Em (HASSENZAHL, DIEFENBACH, GORITZ, 2010), os autores avaliaram a relação entre a satisfação dos usuários e sua percepção positiva dos produtos, mas sem propor um modelo que relacionasse os fatores encontrados com a satisfação dos usuários. No estudo (AL-BADAREEN *et al.*, 2011) os autores apresentaram os requisitos mais importantes para desenvolver um sistema com maior aceitação entre os usuários, mas sem apresentar uma maneira prática de aplicação desses requisitos. Nesta dissertação de mestrado é mostrado na prática, através de um estudo de caso, como fatores de QoE podem influenciar a percepção de confiabilidade de *software* dos usuários, diferente dos estudos (HASSENZAHL, TRACTINSKY, 2006), (HASSENZAHL, 2007), (HASSENZAHL, DIEFENBACH, GORITZ, 2010) e (AL-BADAREEN *et al.*, 2011) que não apresentam avaliações práticas de suas propostas.

2.2 Avaliação da Qualidade da Experiência do Usuário

No trabalho desenvolvido em (CHEN; TU; XIAO, 2009) os autores propuseram um *framework* para mensurar a QoE de redes. Para tal, os autores utilizaram uma amostra de três voluntários que usaram em um ambiente de teste de aplicações de conversação (Skype, MSN Messenger e AIM Messenger) e jogos (Unreal Tournament e Halo). O ambiente de teste era composto por um computador com um *key logging* instalado conectado a uma LAN controlada. O voluntário tinha à sua disposição um botão que deveria ser pressionado sempre que se sentisse insatisfeito com a conexão. O ambiente de teste foi projetado para que a qualidade da conexão fosse controlada pelos autores. Após a coleta de dados, que nesse estudo foi a série de pressionamentos de botão que cada voluntário produziu, os autores fizeram um ajuste entre essa série, a largura de banda da rede e a taxa de perda de pacotes para que os cliques efetuados se aproximassem ainda mais do exato momento em que o voluntário percebeu perda de qualidade na rede. Os resultados foram comparados com outras abordagens encontradas na literatura (PESQ (INTERNATIONAL TELECOMMUNICATION UNION, 2001) e VQM (INTERNATIONAL TELECOMMUNICATION UNION, 2004)) e os resultados obtidos foram semelhantes com o que foi observado nas outras abordagens. Os

autores concluíram o estudo indicando que as avaliações de QoE devem considerar mais de uma medida de QoS, pois considerar apenas uma medida pode levar o pesquisador a conclusões equivocadas.

Em (FIEDLER; HOSSFELD; TRAN-GIA, 2010), os autores propuseram uma fórmula genérica relacionando QoE e QoS através de uma função exponencial, chamada hipótese IQX. De acordo com os autores, para derivar o relacionamento entre QoE e QoS é necessário comparar a qualidade percebida de um conteúdo (uma imagem, ou vídeo, por exemplo) sem distorções, denominado referência, com o conteúdo recebido pelos usuários na transmissão do conteúdo, chamado resultado. As métricas de avaliação de QoE podem ser classificadas em três tipos: *full reference*, *no reference* e *reduced reference*. *Full reference* são avaliações em que a referência e o resultado estão disponíveis para a avaliação de QoE ser realizada, comparando a percepção do usuário utilizando a referência e o resultado. Essa abordagem é utilizada em testes laboratoriais. *No reference* são avaliações em que a percepção de qualidade é extraída somente com o resultado, devido a ausência de referência. No tipo *reduced reference*, a qualidade dos parâmetros de QoS do conteúdo é avaliada através de mensuração indireta e comparada com o resultado. Nesse estudo, os autores compararam modelos de QoE ajustados em outros trabalhos (mensuração de qualidade de *web browsing* e *codecs* de áudio) com a hipótese IQX. Foi observado melhor ajuste dos modelos que consideraram a hipótese IQX em todos os casos. Todos os modelos avaliados consideravam a variação na QoE em função de um único parâmetro de QoS. O resultado alcançado também permite elaborar mecanismos de controle de QoE em função do monitoramento de QoS.

No estudo realizado em (ICKIN *et al.* 2012), os autores investigaram fatores que influenciam a QoE dos usuários quanto ao uso de aplicativos em dispositivos móveis. Através de um questionário *online*, os autores coletaram dados de 29 participantes dentre 430 que se disponibilizaram a participar do estudo. Eles deram preferência para usuários de dispositivos com SO *Android* e que relataram utilizar com alta frequência seus aparelhos. Para avaliar a QoE dos participantes, foi instalado em cada aparelho um programa que captura informações de uso, sem violar a privacidade dos participantes. Além disso, um formulário foi utilizado para coletar a QoE do participante após o uso das aplicações que o participante relatou utilizar com maior frequência no questionário *online*. Esse formulário exibia uma escala MOS para o usuário avaliar sua QoE em momentos aleatórios do dia, com um número máximo de 12 avaliações por dia. Por fim, os participantes eram entrevistados pessoalmente, uma vez por semana, para relatar sua QoE com seus aparelhos para os pesquisadores. Essa etapa foi incluída no estudo para investigar as possíveis inconsistências entre as informações coletadas do aparelho e as avaliações feitas pelo formulário. A coleta de dados foi realizada durante quatro semanas. Analisando os dados, foram encontrados os seguintes fatores de influência: *design* e desempenho das aplicações, capacidade da bateria, características específicas dos aparelhos (GPS, modo silencioso, despertador, entre outras), custo de conectividade com a *Internet*, a rotina do participante e seu estilo de vida.

Em (REICHL; TUFFIN; SCHATZ, 2013), os autores mostraram que para cenários que contenham o uso de *Voice over Internet Protocol* (VoIP) e de banda larga móvel, há evidências crescentes de que a experiência e a satisfação do usuário seguem leis logarítmicas.

Também foi demonstrado no trabalho que é possível fazer a conversão de fatores de QoS para fatores de QoE. Uma queda na QoS é convertida em um estímulo, que por sua vez é convertido em uma métrica de QoE através do uso da *Weber-Fechner Law* (WFL) (PASQUALI, 2010). A WFL descreve o efeito de um estímulo físico sobre a percepção humana de acordo com os princípios da psicofísica. Esta lei afirma que a relação entre estímulo e percepção é logarítmica, o que leva à conclusão de que, para que uma mudança de estímulo seja detectada de forma confiável por um observador, isso deve diferir do seu valor original por uma fração constante. A partir desta lei, emerge a noção de "diferenças apenas visíveis", que descreve a menor diferença detectável entre dois níveis sequenciais de um estímulo particular. Os autores também demonstraram a elaboração de modelos de QoE através dos chamados *Key Quality Indicators* (KQI's), considerando quais KQIs podem influenciar na análise de QoE, atribuindo pesos para cada KQI escolhido. Para atestar a validade desse método, os autores relatam a avaliação de QoE de *download* de arquivos e de navegação na web. Os dois experimentos foram realizados em um ambiente controlado no qual um operador controlava a rede e a interface dos usuários através de um *desktop* e os voluntários interagiam com os sistemas através de um *laptop*. Em ambos os experimentos, os autores observaram que a percepção de queda de qualidade dos usuários é maior quando a qualidade é reduzida em momentos em que o serviço prestado já está com alta qualidade do que quando o serviço está com qualidade ruim e a mesma cai ainda mais.

No trabalho desenvolvido em (RAINER; TIMMERER, 2014), os autores investigaram três efeitos sensoriais (luz ambiente, vento e vibração) para entender como esses efeitos influenciam a QoE percebida durante o uso de mídias audiovisuais. Para tal, os autores expuseram os 32 participantes da pesquisa a quatro diferentes vídeos, todos com duração de dez segundos. Cada vídeo foi exibido a todos os participantes sem nenhum efeito sensorial e com combinações desses efeitos, em um total de oito etapas de exibição. Nesse estudo, os participantes respondiam a três questionários: o primeiro foi feito para coleta de dados demográficos e era respondido antes do experimento, o segundo para avaliar a QoE percebida após a exibição de cada vídeo e o terceiro, respondido ao final da coleta de dados, perguntava se o sujeito humano já havia participado de pesquisas desse tipo antes e abria espaço para o participante dar seu *feedback* sobre a pesquisa. Os participantes avaliavam a QoE percebida após cada exibição através de uma barra deslizante com valores de 0 a 100, na qual 0 indicava QoE muito baixa e 100 indicava QoE muito alta. Como resultado, os autores observaram que o efeito sensorial que mais influenciou a QoE percebida foi a vibração. Também foi observado uma relação linear entre a QoE percebida sem efeitos sensoriais e a QoE percebida com tais efeitos. Por fim, foi ajustado um modelo de predição de QuaSE que considera os efeitos sensoriais de forma individual ou coletiva.

Em (LIOTOU *et al.* 2015), os autores abordaram os problemas de *design* e questões importantes para o gerenciamento de QoE em redes de dispositivos móveis. Também foi proposto um *framework* para a obtenção de provisionamento de QoE de ponta a ponta. No trabalho, os autores destacam os pilares de influência de QoE e seus fatores, que são: a rede e seus aspectos (*wireless*, mobilidade, entre outros), contexto (ambiente, conteúdo, equipamento, entre outros) e fatores humanos (psicofísicos, psicológicos, ciência cognitiva,

entre outros). Considerando os fatores de influência, os autores desenvolveram o *framework* proposto, a central de gerenciamento de QoE. O *framework* é composto por quatro componentes: monitor de QoE, gerenciador de QoE, controlador de QoE e a rede. O *framework* funciona da seguinte maneira: o controlador envia instruções de controle de dados de QoE para a rede para coletar os dados a serem analisados. Os dados relacionados ao estado atual da rede são enviados ao monitor que envia as pontuações de QoE calculadas para o gerenciador. No gerenciador, os processos de gerenciamento da experiência do usuário são realizados e ações corretivas são enviadas ao controlador para serem implementadas caso necessário. Para avaliar o *framework*, os autores elaboraram uma simulação de uma rede utilizando dados reais coletados previamente. Foram executadas diversas tarefas na rede e a QoE foi avaliada com base no MOS. Os testes foram realizados sem o *framework* e com o *framework*. Nos testes sem o *framework*, a QoE caiu com o passar do tempo chegando a atingir valores inferiores a 1,5. Com a implementação do *framework*, assim que a QoE percebida atingia o limite de 3,5, o *framework* iniciou o processo para elevar a QoE, que atingiu o mínimo de 3,3 e máximo de 4,0. Nesse estudo, os autores conseguiram elaborar uma ferramenta para melhorar a QoE em redes de celulares, porém eles destacam a importância de melhorar a QoE em outros contextos de uso. Outro ponto importante desse trabalho é a exibição de fatores de influência de QoE e a demonstração de que considerar esses fatores é relevante para melhorar a QoE dos usuários.

Segundo os autores em (TSOLKAS *et al.* 2017), a QoE pode ser modelada de três formas: subjetiva, objetiva e híbrida. Os testes subjetivos geralmente são baseados em experimentos controlados com sujeitos que avaliam diretamente sua experiência ao utilizar um aplicativo e/ou serviço. Esses usuários podem estar envolvidos na experiência de forma passiva (apenas visualizando e/ou ouvindo) ou de forma ativa/interativa (participando de uma conversa) e julgam a qualidade em relação à apresentação de algum estímulo. Por exemplo, os participantes podem ser chamados para avaliar a qualidade de audição ou conversação de um serviço de telefone, a qualidade de um vídeo, entre outros. Várias técnicas podem ser usadas para avaliação subjetiva. Por exemplo, os usuários podem avaliar a qualidade usando uma escala MOS, que basicamente é um *rank* de 1 a 5 (1 – muito ruim, 2 – ruim, 3 – regular, 4 – bom, 5 - excelente) que é utilizado para avaliar a QoE do usuário. Também é possível comparar imagens sequenciais/vídeos/sons indicando qual é melhor. Os resultados são baseados nas opiniões dos usuários, experiências passadas, expectativas, percepção do usuário, entre outros. O principal objetivo é quantificar a eficácia, eficiência e satisfação geral do usuário ao usufruir de um serviço.

Testes subjetivos são considerados os mais confiáveis, uma vez que incorporam aspectos conscientes e inconscientes da avaliação da qualidade humana, aspectos que de outra forma não podem ser capturados (Moller & Raake, 2014). Somente os testes de qualidade percebida podem expressar de forma válida e confiável o estado interno do sujeito humano. Uma desvantagem desse método é que os resultados de tais experimentos são válidos apenas para testes laboratoriais de algum serviço e não para suporte de QoE em tempo real. Além disso, testes subjetivos são dispendiosos, demorados, não reproduzíveis sob demanda e, inevitavelmente, não são executados em tempo real (Moller & Raake, 2014). Portanto, não

podem ser usados para monitoramento de qualidade em serviço. Essas restrições levantaram a necessidade de desenvolvimento de modelos objetivos que buscam medir ou prever a qualidade percebida pelos usuários finais, sem sua intervenção.

Os modelos objetivos podem ser classificados de acordo com vários critérios, tais como: utilização do sinal de referência, modo do modelo, calendário do modelo, uso final e método de avaliação (Moller & Raake, 2014). Por fim, existe a avaliação de QoE híbrida, que se situa entre as avaliações subjetivas e objetivas. Ela opera de forma híbrida, ou seja, funciona como um estimador de qualidade automático e objetivo, confiando, no entanto, nas pontuações subjetivas disponíveis anteriormente. Métodos híbridos são baseados em ferramentas de aprendizagem de máquina e eles estão usando pontuações subjetivas de teste como entrada para treinar um modelo QoE.

Nesta dissertação, a abordagem adotada foi a subjetiva, que apesar de suas desvantagens, é a abordagem mais confiável das três apresentadas anteriormente.

2.3. Avaliação da Confiabilidade de Sistemas Operacionais

Em (XU; KALBARCZYK; IYER, 1999), os autores investigaram a confiabilidade de servidores com o SO Windows NT. Eles coletaram dados de falha de 503 servidores ao longo de quatro meses, considerando como falhas todos os eventos inesperados que levaram a uma reinicialização do sistema, travamento ou parada. Assim, o artigo deu ênfase à disponibilidade do sistema, investigando falhas relacionadas ao tempo de inatividade do sistema. Os autores descobriram que os principais fatores que contribuem para o tempo de inatividade total do sistema são falhas de *hardware* e *software*. Também foram encontrados indícios de dependência de erros ou propagação de erros na rede. Por fim, eles observaram muitos casos que exigiam mais de uma reinicialização para o sistema se recuperar de uma falha. Os autores concluem o trabalho recomendando que se tenha cautela ao analisar os *logs* de um sistema, evitando superestimar os valores presentes nos dados.

Da mesma forma, (SWIFT; BERSHAD; LEVY, 2003) apresenta um subsistema de confiabilidade que melhora a confiabilidade do sistema, isolando o SO de falhas de *driver* dos dispositivos. Esse subsistema, chamado Nooks, impede a maioria das falhas de *driver* dos dispositivos com pouca alteração no código de tais *drivers*. Os autores relataram que as extensões do *kernel* do SO (por exemplo, *drivers* de dispositivos) representavam mais de 70% do código do *kernel* do Linux, enquanto o Windows XP apresentava mais de 35.000 *drivers* de dispositivo diferentes. Os autores concluíram que as extensões de terceiros eram a principal causa de falhas no Windows XP, em que os *drivers* eram responsáveis por 85% das falhas relatadas.

Os autores de (GANAPATHI; GANAPATHI; PATTERSON, 2006) analisaram 2.528 eventos de falhas no *kernel* do Windows XP de 617 computadores. Eles descobriram que o SO não foi responsável pela maioria das falhas no sistema. A principal causa de falhas do sistema eram *drivers* de dispositivos mal escritos de terceiros. Portanto, este trabalho corrobora o estudo apresentado em (SWIFT; BERSHAD; LEVY, 2003), enfatizando que os *drivers* de dispositivo contribuem para a maioria das falhas do sistema. Os autores concluem

o estudo alertando os usuários para que evitem a instalação de programas e *drivers* inseguros, bem como indicam a divulgação dos dados de falha como forma de melhorar a concorrência entre as desenvolvedoras de sistemas, pois os consumidores irão procurar adquirir os *softwares* com maior confiabilidade.

Nota-se que em (XU; KALBARCZYK; IYER, 1999), (SWIFT; BERSHAD; LEVY, 2003) e (GANAPATHI; GANAPATHI; PATTERSON, 2006), apenas falhas em espaço de *kernel* foram consideradas. No entanto, de acordo com (MATIAS; OLIVEIRA; ARAÚJO, 2013), essa abordagem não representa bem a confiabilidade do SO, uma vez que os sistemas operacionais modernos têm várias partes (por exemplo, aplicativos gerenciadores de janela) além dos componentes do *kernel* e não considerar tais componentes na análise de confiabilidade pode interferir, significativamente, na avaliação da confiabilidade.

Seguindo a mesma abordagem apresentada em (MATIAS; OLIVEIRA; ARAÚJO, 2013), nos próximos trabalhos apresentados as seguintes categorias de falhas do SO foram levadas em consideração nas análises: falhas de aplicação, *kernel* e serviço do SO. A primeira categoria contém as falhas causadas pelo funcionamento incorreto das aplicações de SO. Tais aplicações geralmente executam em primeiro plano e realizam funções específicas que são requisitadas de acordo com a necessidade do usuário do sistema. Um exemplo de aplicação de SO é o programa *explorer.exe* no sistema operacional Microsoft Windows 7, responsável pelo gerenciamento de janelas e manutenção/aceso aos arquivos nesse SO. A segunda categoria possui falhas causadas pelos subsistemas do *kernel* do SO, as quais ocorrem no espaço do *kernel* e, geralmente, exigem a reinicialização do sistema. As falhas nesta categoria estão fortemente relacionadas às bibliotecas do sistema (ex. *ntdll.dll* e *Kernelbase.dll*) e *drivers* de dispositivos. Um exemplo de falha nesta categoria é a VIDEO_TDR_ERROR. Essa falha ocorre quando o controlador do dispositivo de vídeo detecta que já não pode controlar a unidade de processamento gráfico. Isso indica uma falha durante a tentativa de recarregar o *driver* de vídeo (RUSSINOVICH; SOLOMON; IONESCU, 2012). Por fim, a terceira categoria contém as falhas de serviços do SO. A maioria desses serviços é iniciada durante a carga do SO e executa em segundo plano (*background*). Normalmente, os serviços de SO são utilizados por outros componentes do SO e, portanto, suas funcionalidades não estão diretamente ligadas à necessidade do usuário do sistema. Como exemplo pode-se citar o *svchost.exe*, que possui a função de hospedar outros serviços individuais do Microsoft Windows 7 que são implementados por *Dynamic-link Libraries* (Dlls), fazendo a interface entre as Dlls de serviços do Microsoft Windows 7 e os demais componentes do sistema (ANSON *et al.*, 2012).

Em (MATIAS; OLIVEIRA; ARAÚJO, 2013) foi realizado um estudo exploratório sobre confiabilidade de sistemas operacionais com o foco na perspectiva da experiência do usuário. Para isso, os autores criaram uma abordagem que considera não apenas as falhas do *kernel* como falhas de SO, mas também as falhas causadas nas aplicações do SO e serviços do SO. Nesse estudo foi realizada uma análise de 2.634 falhas provenientes de 106 computadores que utilizavam o *Microsoft Windows 7* como SO. Essas amostras de falhas foram coletadas e organizadas em dois grupos G1 e G2. O grupo G1 é composto por dados de falha coletados de 50 computadores em um mesmo ambiente acadêmico, enquanto que o grupo G2 é composto

por 56 computadores de diferentes ambientes empresariais. Após a análise dos resultados, os autores chegaram à conclusão de que o grupo G1 apresentava maior tempo médio até a falha (MTTF) do que o grupo G2 devido ao processo sistemático de manutenção do SO adotado no ambiente de produção, especialmente no sistema de atualização. Além disso, esse grupo apresentava maior número de falhas no nível de *kernel*. Enquanto o grupo G2 apresentava maior número de falhas nas aplicações do SO. Os autores acreditam que essa categoria apresenta um impacto significativo na experiência do usuário quando comparada com as outras categorias (*kernel* e serviço).

Em (MATIAS *et al.* 2014), o objetivo da pesquisa foi realizar um estudo empírico e exploratório sobre confiabilidade de sistemas operacionais. Os autores analisaram 735 computadores de diferentes ambientes de trabalho que utilizavam o Microsoft Windows 7 como SO. Essa análise resultou em 30.815 falhas que foram coletadas a partir do RAC (*Reliability Analysis Component*), componente do SO que contém todos os eventos de falhas registrados desde a sua instalação. Essas falhas foram categorizadas em falhas de aplicações, *kernel* e serviços do SO. Os autores chegaram à conclusão que a maioria das falhas no SO é da categoria serviços do SO, sendo que a ocorrência de falhas das outras duas categorias é menos frequente e muito similar. A partir de análises estatísticas dos tempos de falhas, descobriram que as distribuições de probabilidade *Gamma* e *Weibull* foram as distribuições que melhor aderiram aos dados. Também foi observado que as falhas de *kernel* do SO ocorrem mais em ambientes corporativos do que em ambientes acadêmicos.

Os autores de (MATIAS; ANTUNES, 2014) tiveram como propósito apresentar um estudo experimental sobre a confiabilidade de um SO e, por essa razão, analisaram uma amostra com 3.235 registros reais de falhas de SO obtidos de diferentes computadores. Os autores avaliaram quantitativamente a importância de cada uma das causas identificadas sobre confiabilidade do Microsoft Windows 7. Por meio das causas de falhas encontradas, eles criaram modelos estocásticos de confiabilidade que permitiram avaliar a sensibilidade da confiabilidade do SO com respeito a cada uma das causas de falhas observadas. Uma das contribuições do trabalho foi apresentar que as evidências da amostra analisada sugerem que as falhas nas categorias serviços e aplicações de SO têm impacto importante na redução da confiabilidade do sistema. Os autores concluíram que a análise das principais causas de falhas de SO observadas mostra que a atualização de *software* é o serviço de SO com maior índice de importância de confiabilidade, ou seja, o componente cuja melhoria deve ser priorizada.

No trabalho (DOS SANTOS, 2016) é apresentado um estudo exploratório sobre o comportamento das falhas de SO, identificando os padrões de falha do Microsoft Windows 7. A amostra foi composta por 7.007 registros de falhas de SO obtidos de 566 computadores de diferentes ambientes de trabalho executando o sistema operacional Microsoft Windows 7. A partir desses dados, o autor realizou uma análise quantitativa para investigar diferentes propriedades das falhas de SO analisadas. Além disso, foi desenvolvido um protocolo para detectar e caracterizar padrões de falhas dos diferentes computadores investigados. Adaptando o protocolo, o autor conseguiu também caracterizar padrões de causalidade entre as falhas de SO. O autor concluiu que as falhas de serviços de SO, principalmente falhas de atualização de *software*, foram mais frequentes do que qualquer outra categoria de falha. Esse resultado

confirmou que os estudos de confiabilidade de sistemas operacionais não devem concentrar apenas em falhas de *kernel*. O autor também detectou 45 padrões de falhas com 153.511 ocorrências, encontrando evidências empíricas de correlação entre falhas de componentes específicos de SO, o que ajudou a melhor compreender o comportamento dessas falhas de *software*.

Em (DOS SANTOS *et al.*, 2018) é apresentado um estudo empírico sobre a confiabilidade do SO Microsoft Windows 7, em que os autores avaliaram as três falhas do SO que mais ocorreram na amostra das três categorias: aplicação, *kernel* e serviço do SO. Também foi avaliado como as falhas de cada categoria afeta a confiabilidade geral do SO. A amostra investigada no estudo foi composta por 113 computadores com 5.351 falhas do SO. Após a categorização das falhas, os autores utilizaram o Critério de Informação de Akaike para verificar as distribuições de probabilidade de falha que melhor se ajustavam ao tempo entre falhas de cada categoria. Após o ajuste dessas distribuições, foram criados modelos estocásticos com diagramas de blocos de confiabilidade em dois níveis para avaliar a confiabilidade do SO. O primeiro nível continha as três categorias de falhas do SO e o segundo as três falhas que mais ocorreram em cada categoria. Os resultados apontaram as distribuições exponencial e lognormal como as de melhor ajuste para a categoria *kernel*, levy e exponencial para a categoria serviço e as distribuições levy, exponencial e log-logística para a categoria aplicação do SO. Foi observado que a confiabilidade do SO calculada considerando as três categorias de falhas atinge menos de 15% nas primeiras 4 horas de uso, enquanto que ao considerar apenas a categoria *kernel* esse valor é maior que 90%, cerca de 30% ao considerar somente a categoria serviço e quase 40% ao considerar apenas a categoria aplicação do SO. A categoria que mais impactou na confiabilidade do SO foi a categoria serviço do SO para as primeiras 8 horas de funcionamento do sistema e a categoria aplicação do SO entre 12 e 24 horas de funcionamento. Os autores concluem o estudo sugerindo a redução do escopo de funcionamento das aplicações e serviços do SO que mais falharam, bem como a desinstalação dos componentes que não são utilizados pelos usuários mas que continuam sendo atualizados pelo sistema como forma de melhorar a confiabilidade do SO.

2.4 Avaliação da Confiabilidade de *Software* Considerando

Aspectos de QoE

Em (YABE *et al.*, 2018), os autores investigaram efeitos de diferentes tipos de falhas, além de outros fatores de influência, na percepção do usuário sobre a confiabilidade do SO. Para avaliar tal percepção, foram utilizados cenários que descreviam situações hipotéticas, mas realísticas em que ocorriam falhas de *software*. Nesses cenários foram consideradas falhas do SO Microsoft Windows 7 das categorias aplicação, *kernel* e serviço do SO. Também foi considerada uma falha de uma aplicação que não era do SO para verificar se falhas que não são do SO são percebidas de forma diferente. Além da descrição da ocorrência de uma falha, cada cenário descrevia um contexto (profissional ou lazer) em que o cenário ocorria e uma consequência para a ocorrência da falha (por exemplo, o usuário perde o documento que estava editando). Foi aplicado um questionário *online* para 121 sujeitos utilizando 8 cenários para avaliar se o nível de conhecimento em computação dos respondentes (leigos e intermediários) impactava em sua percepção, bem como investigar quais foram os cenários de

maior e menor impacto na percepção dos usuários. Também foi aplicado um questionário presencial para 120 participantes, no qual cada participante pôde vivenciar um cenário de ocorrência de falha através de um instrumental que simulava as falhas relatadas nos cenários utilizados. Os cenários do questionário presencial eram diferentes daqueles utilizados no questionário *online*. Com o questionário presencial, foi possível avaliar os principais fatores de influência na QoE dos participantes. Os resultados mostraram que o nível de conhecimento dos participantes não teve influência significativa na percepção das falhas. Foram encontradas evidências estatísticas de que a QoE do usuário, no que diz respeito à confiabilidade do SO, é afetada muito menos pelos tipos de falha do SO do que por outros fatores de influência, como a consequência da falha e o contexto (prazo de trabalho). Também foram observados efeitos significativos de fatores humanos na QoE dos usuários, como expectativa de solução e familiaridade com eventos de falha. Esse estudo contribuiu para uma melhor compreensão do impacto causado pelas falhas do SO na QoE do usuário, o que pode melhorar a modelagem de confiabilidade de futuros sistemas operacionais adicionando fatores de percepção de confiabilidade dos usuários em vez de apenas elementos de QoS.

No estudo relatado em (JALOTE *et al.* 2004), os autores definiram a confiabilidade como uma medida capaz de representar uma operação sem falhas e, para fins práticos, é denominada como taxa de falha. Dessa forma, eles consideraram a confiabilidade de *software* como um vetor composto por taxas de diferentes tipos de falhas. Segundo os autores, as diferentes percepções de confiabilidade de vários grupos de usuários podem ser refletidas através da atribuição de pesos diferentes para diferentes tipos de falhas. Eles afirmam que a classificação da falha, tamanho da população, dados da falha, tempos de uso, configuração de *hardware* e *software* são requisitos importantes que precisam ser coletados para medir a confiabilidade de *software*. No estudo é relatado que para capturar a UX quanto à confiabilidade do sistema a classificação da falha tem de ser realizada a partir da perspectiva do usuário. Eles afirmam que como as falhas encontradas por um usuário dependem claramente da quantidade de uso do produto, para ter uma ideia precisa de confiabilidade do *software* também é necessário capturar o tempo de uso do sistema. Os autores acreditam que medir a confiabilidade dessa forma, levando em consideração as necessidades dos usuários, pode ajudar a determinar os módulos do *software* que precisam ser reforçados para melhorar a UX quanto à confiabilidade do sistema. Esse estudo não apresenta um método para medir a confiabilidade de sistemas, apenas apresenta uma alternativa sem avançar neste sentido.

Os autores em (VEERARAGAVAN *et al.* 2016) propuseram um modelo para avaliar a QoE aplicada à confiabilidade de *software*, denominada *confiabilidade percebida*, com o objetivo de explicar a percepção subjetiva de diferentes usuários em aplicações tele imersivas. Os autores defendem que a noção clássica de confiabilidade não pode avaliar precisamente tais aplicações, porque falhas parciais e intermitentes podem não necessariamente prejudicar o desempenho do sistema. O modelo proposto foi avaliado em um estudo de caso envolvendo um ensaio de ópera *online*, considerando os níveis de qualidade de áudio e vídeo percebidos pela orquestra, o cantor, o maestro, o cravo e o público (papéis do usuário desempenhado na ópera), em diferentes locais e etapas. A avaliação foi realizada executando simulações utilizando um *framework* denominado Mobius. Os resultados sugerem que os valores de

confiabilidade percebidos não são os mesmos em locais diferentes e para diferentes funções de usuário, pois a UX é dependente de contexto. No estudo, os autores relataram que a adição de peças sobressalentes para todos os componentes não fornece uma melhoria significativa em comparação com a adição de peças sobressalentes apenas para os componentes menos confiáveis. Nesse caso, o uso de componentes confiáveis não melhora necessariamente o desempenho de todas as funções do usuário em todos os estágios.

O estudo relatado em (VEERARAGAVAN *et al.*, 2016) possui algumas diferenças em relação ao trabalho realizado nesta dissertação. A avaliação de confiabilidade realizada no estudo levou em consideração que uma falha na QoS que não é percebida pelos usuários pode não afetar sua QoE como um todo. Por exemplo, uma falha no áudio durante uma apresentação de uma orquestra pode não ser percebida pela audiência, desde que o vídeo não tenha falhado no momento em que o áudio falhou. Desse modo, nas avaliações realizadas no estudo os autores não consideraram a definição clássica de confiabilidade (ver Seção 1.1), pois tal definição não se adequou aos requisitos por eles considerados. Nesta dissertação, a definição clássica de confiabilidade foi considerada, pois a percepção dos usuários foi avaliada após a interrupção da entrega contínua e correta do serviço prestado pelo sistema. Outra diferença entre ambos os trabalhos é que em (VEERARAGAVAN *et al.*, 2016) os autores dividiram os usuários em grupos (orquestra, cantor, audiência e cravo), para avaliar a confiabilidade percebida por cada grupo. Nesta pesquisa não foi feito agrupamento dos usuários, pois em pesquisa anterior (YABE *et al.*, 2018), em que os usuários foram agrupados por nível de conhecimento em computação (leigos e intermediários), não foi observada diferença estatística significativa entre esses tipos de grupos.

2.5 Síntese dos Trabalhos

Os trabalhos relatados na Seção 2.1 mostram conceitos e descobertas relacionadas à experiência dos usuários. Nesses trabalhos pôde-se analisar as diversas abordagens ao avaliar a UX dos usuários, bem como observar a importância de se considerar os sentimentos positivos vivenciados pelos usuários ao utilizarem um sistema. Nesses estudos foram apresentados os aspectos hedônicos e pragmáticos da UX. Nesta dissertação, o foco está em avaliar os aspectos hedônicos da UX dos usuários. Tais estudos relatam que as experiências vivenciadas pelos usuários podem ser categorizadas pelas necessidades primárias dos usuários atendidas por tais necessidades. Nessa dissertação, avaliou-se como as falhas de *software* afetam as necessidades dos usuários (a necessidade do usuário aqui seria completar as atividades descritas nos cenários com sucesso). Outro ponto importante citado nesses estudos que a satisfação do usuário é fundamental para o sucesso de qualquer produto no mercado, apresentando os requisitos que precisam de mais atenção ao desenvolver um sistema que será aceito pelo usuário. Entre esses requisitos é citada a confiabilidade do sistema, que é avaliada nesta dissertação.

Na Seção 2.2 são apresentados estudos que avaliam a QoE dos usuários utilizando a QoS como referência para as avaliações. Esses trabalhos avaliaram a QoE dos usuários realizando coletas de dados *online* e presencial, nos quais em cada trabalho foram levantados diferentes fatores de influência na percepção dos usuários quanto à QoE do sistema avaliado. Exceto em (FIEDLER; HOSSFELD; TRAN-GIA, 2010), no qual os autores propuseram uma

fórmula genérica relacionando QoE e QoS através de uma função exponencial. Como pode ser observada uma alta variedade de fatores de influência nos trabalhos analisados, nesta dissertação os fatores de influência utilizados foram baseados em um estudo prévio (ver Capítulo 3) com características semelhantes ao realizado nesta dissertação.

Os estudos analisados que avaliaram a confiabilidade através da QoS (ver Seção 2.3) calculavam a confiabilidade do sistema utilizando os tempos de falhas desses sistemas. Esses estudos são importantes por evidenciarem os componentes que apresentam mais falhas em um sistema, entretanto, nesse trabalhos a percepção dos usuários não é levada em consideração. Na Seção 2.4, são relatados trabalhos que avaliaram a confiabilidade de *software* considerando aspectos de QoE. Nesses estudos, são mencionados possíveis fatores de influência na percepção dos usuários em relação à confiabilidade do *software* (YABE *et al.*, 2018), (JALOTE *et al.* 2004), bem como um possível método para calcular a confiabilidade do sistema percebida pelos usuários (VEERARAGAVAN *et al.*, 2016).

Metodologia

3.1 Introdução

Para verificar se o resultado do cálculo de confiabilidade seguindo a abordagem clássica, é representativo do ponto de vista dos usuários, ou seja, reflete a confiança do usuário no sistema com base em sua experiência ao usá-lo (QoE), foi utilizado um questionário que foi aplicado de forma *online* e presencial. Primeiramente, optou-se por aplicar o questionário *online* devido principalmente ao custo reduzido para alcançar mais respondentes em menos tempo (GOLDENBERG, 2004; VIEIRA, 2009).

Após a aplicação do questionário *online* também foram realizadas sessões de coleta de dados de forma presencial para mitigar uma das ameaças à validade deste estudo, relacionada com a possibilidade de os participantes do questionário *online* não terem entendido as perguntas do questionário (ver Seção 5.2). Durante a coleta de dados *online* não houve interferência do pesquisador quanto à opinião dos participantes, o que caracteriza este estudo como de observação transversal. Pesquisas de observação têm como objetivo compreender comportamentos e acontecimentos sem interferência do pesquisador, de forma que seja possível compreender a complexidade do estudo, analisar o comportamento e percepção das atitudes dos participantes (ZANELLI, 2002).

Buscando minimizar o tempo e o custo para aplicar o questionário, bem como evitar ambiguidades nas respostas dos participantes, optou-se por utilizar questões fechadas no questionário sempre que possível. Além disso, foram consideradas as recomendações apresentadas em (MARSDEN; WRIGHT, 2010; INTERNATIONAL TELECOMMUNICATION UNION, 1996) na criação e aplicação do questionário.

3.2 Materiais

3.2.1 Terminologia

Alguns termos foram utilizados neste estudo para a elaboração das perguntas de pesquisa e do questionário. Para detalhar tais termos no âmbito deste trabalho, criou-se uma versão adaptada do dicionário de termos utilizado em (YABE *et al.*, 2018):

- **Cenário:** um cenário é uma descrição idealizada, mas detalhada, de uma instância específica de uma interação humano-computador⁶ (YOUNG; BARNARD, p.1, 1987, tradução nossa). Exemplos de cenário podem ser vistos no Apêndice A.

⁶ Scenario is an idealized but detailed description of a specific instance of human-computer interaction.

- **Consequência:** efeito que um evento de falha provoca no usuário. Exemplo: devido à ocorrência do evento de falha, o usuário perde um documento que estava editando e, conseqüentemente, tem retrabalho.
- **Contexto:** qualquer informação que pode ser usada para caracterizar a situação da entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerado relevante para a interação entre o usuário e uma aplicação, incluindo tal usuário e aplicação⁷ (ABOWD *et al.*, p. 304, 1999, tradução nossa). Devido a essa definição ser ampla, neste trabalho o contexto foi restringido ao ambiente hipotético no qual ocorrem os cenários, que foi definido em todos os cenários como de cunho profissional. Exemplo: você está no seu trabalho editando um documento e devido à ocorrência de uma falha você perde o documento editado. O ambiente de trabalho relatado nesse exemplo é o contexto.
- **Impacto:** o quanto a QoE é afetada negativamente pela consequência de uma falha associada ao contexto que o usuário está inserido. Exemplo: após a ocorrência do evento de falha, o usuário precisou cancelar uma reunião. O usuário fica insatisfeito e avalia como ruim a sua experiência quanto ao uso do computador.
- **Percepção:** é o processamento consciente da informação sensorial a qual o sujeito humano está exposto⁸ (MOLLER; RAAKE, p.13, 2014, tradução nossa). Exemplo: um sentimento, conceito ou percepção sensorial que o usuário processou conscientemente ao vivenciar uma ou mais ocorrências de falha.
- **Vivenciar:** é o fluxo individual de percepções (de sentimentos, percepções sensoriais e conceitos) que ocorre em uma situação particular de referência, ou seja, é um conjunto de percepções⁹ (MOLLER; RAAKE, p.13, 2014, tradução nossa). Neste trabalho, é o fluxo de percepções que ocorre quando o usuário vivencia a ocorrência de uma ou mais falhas. Exemplo: o usuário vivencia uma falha e se sente frustrado ao ver a mensagem de falha e se recordar de um prejuízo financeiro que ele teve no passado devido a essa mesma falha.

3.2.2 Questionário

Foi elaborado um questionário com quatro cenários descrevendo situações hipotéticas, mas realísticas, de uso do computador em que os usuários assumiram estar vivenciando as falhas da maneira como elas foram apresentadas em cada cenário. Cada cenário apresenta um evento de falha e uma consequência próprios, porém todos os cenários são definidos para um contexto profissional. Os cenários em questão podem ser visualizados no Apêndice A. A Tabela 1 mostra a distinção de cada cenário de falha do SO utilizado no questionário. Para analisar o impacto causado pelas diferentes falhas de SO na QoE do usuário, os cenários foram divididos em “baixo impacto” e “alto impacto”. Os cenários com baixo impacto são aqueles em que a falha do SO não impede as atividades de trabalho do usuário, com rápida

⁷ Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves.

⁸ Perception is the conscious processing of sensory information the human subject is exposed to.

⁹ Experiencing is the individual stream of perceptions (of feelings, sensory percepts and concepts) that occurs in a particular situation of reference.

recuperação da falha. Os cenários com alto impacto são aqueles em que a falha do SO prejudicou a tarefa que o usuário estava fazendo no computador. Além disso, cada cenário apresenta um valor curto ou longo de MTBF (*Mean Time Between Failures*), para que seja possível avaliar como a frequência de falhas afeta a QoE do usuário. A Tabela 1 também mostra os valores de λ utilizados para calcular a confiabilidade estimada por método clássico ($R_{(t)}$), bem como o valor de $R_{(t)}$ usado em cada cenário. Optou-se por utilizar no questionário valores hipotéticos de MTBF e descritos de forma textual para facilitar a compreensão do participante ao responder as perguntas do questionário, de modo que apresentar valores numéricos de MTBF de maneira técnica poderia prejudicar o entendimento do participante sobre os cenários apresentados.

A ($R_{(t)}$) para cada cenário foi calculada considerando a distribuição de probabilidade de falha Poisson mostrada abaixo (Lewis, 1996):

$$P(X < x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \text{ para } x = 0, 1, 2, \dots \quad (2)$$

onde λ é a taxa de falha média (constante), x é o número de eventos de interesse (ex. falha), e t é o intervalo de tempo considerado. O cálculo da confiabilidade de um sistema considerando a distribuição de probabilidade de falha Poisson é feito através da equação (LEWIS, 1996):

$$R_{(t)} = e^{-\lambda t} \quad (3)$$

na qual λ é a taxa de falha média e t é o intervalo de tempo considerado. A média e a mediana da confiabilidade percebida foram calculadas utilizando os valores coletados aplicados nas fórmulas de média e mediana do *Microsoft Excel*.

Tabela 1 – Distinção dos cenários de falha do questionário

Cenário	Impacto	Tipo de MTBF	MTBF	λ	$R_{(t)}$
Cenário 1	Baixo impacto	MTBF curto	1	1/1	36,80%
Cenário 2	Baixo impacto	MTBF longo	7	1/7	81,00%
Cenário 3	Alto impacto	MTBF curto	1	1/1	36,80%
Cenário 4	Alto impacto	MTBF longo	15	1/15	93,50%

O questionário foi respondido em duas etapas. Na primeira etapa, foram apresentadas quatro perguntas para coletar os seguintes dados descritivos dos participantes: gênero, faixa etária, escolaridade e profissão (Figura 2). Na segunda etapa, o questionário apresenta os cenários de falha, um cenário de cada vez. Os quatro cenários de falha apresentam as seguintes características: a descrição da tarefa (interação usuário-computador) realizada no cenário, como a falha acontece durante a realização da tarefa, a consequência gerada pela ocorrência da falha (que determina o impacto baixo ou alto), uma imagem mostrando a mensagem de falha gerada e uma frase relatando o MTBF. Os participantes responderam a duas perguntas para cada cenário do questionário. Para a primeira pergunta, eles responderam, por meio de uma barra deslizante com o intervalo de 0 (sem confiança alguma) até 100

(confiança total), qual o seu grau de confiança no SO descrito no cenário apresentado (Figura 3). Ao longo das próximas seções e capítulos, os valores respondidos para essa pergunta são denominados *confiabilidade percebida*, $R_{per}(t)$. A segunda pergunta foi elaborada com o intuito de medir as dimensões emocionais do participante em cada cenário, através de uma adaptação de escalas de diferencial semântico (PASQUALI, 2010). A técnica de diferencial semântico consiste em apresentar ao participante escalas bipolares definidas por adjetivos opostos para mensurar determinadas dimensões. Tais adjetivos foram escolhidos através da análise das nuvens de palavras produzidas com as respostas dos participantes do estudo (YABE *et al.*, 2018) em relação ao nível do impacto causado pela ocorrência da falha apresentada. No estudo supracitado, as respostas dos participantes foram classificadas em 10 categorias: consequência, contexto, contexto e consequência, evento de falha, expectativa de solução, familiaridade, não familiaridade, frequência da falha, mensagem de falha e recuperação. O Apêndice B contém as nuvens de palavras elaboradas por categoria. A Figura 4 mostra a segunda pergunta dos cenários e os adjetivos bipolares escolhidos. A Tabela 2 mostra as dimensões e as escalas utilizadas no questionário. A Tabela 3 mostra as principais diferenças entre o questionário *online* e o questionário presencial.

*1. Qual seu gênero?

Feminino

Masculino

*2. Indique sua idade.

até 17 anos

18 a 24 anos

25 a 44 anos

45 a 65 anos

acima de 65 anos

*3. Indique seu grau de escolaridade.

Ensino fundamental incompleto

Ensino fundamental completo

Ensino médio incompleto

Ensino médio completo

Superior incompleto

Superior completo

Pós-graduação

Outros:

*4. Informe sua profissão ou área de estudo/atividade.

Figura 3 - Perguntas da primeira etapa do questionário

*Com base no cenário descrito acima, qual o grau de confiança você atribui ao sistema operacional do computador em uso (0% a 100%):

0 100

Figura 4 - Primeira pergunta dos cenários de falha (segunda etapa)

*Para cada escala a seguir, marque a posição que melhor represente a sua opinião ao experimentar o cenário descrito acima:

Insatisfeito	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Satisfeito
Preocupado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Despreocupado
Desamparado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Amparado
Intolerante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tolerante
Improdutivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Produtivo
Nervoso	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Calmo
Atrasado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Pontual
Impaciente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Paciente
Desesperado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tranquilo
Incapaz	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Capaz

Figura 5 - Segunda pergunta dos cenários de falha (segunda etapa)

Tabela 2 - Dimensões e escalas utilizadas no questionário

Dimensão	Escala		
Satisfação	Insatisfeito	OOOOOOOOOO	Satisfeito
Preocupação	Preocupado	OOOOOOOOOO	Despreocupado
Amparo	Desamparado	OOOOOOOOOO	Amparado
Tolerância	Intolerante	OOOOOOOOOO	Tolerante
Produtividade	Improdutivo	OOOOOOOOOO	Produtivo
Estresse	Nervoso	OOOOOOOOOO	Calmo
Pontualidade	Atrasado	OOOOOOOOOO	Pontual
Paciência	Impaciente	OOOOOOOOOO	Paciente
Tranquilidade	Desesperado	OOOOOOOOOO	Tranquilo
Capacidade	Incapaz	OOOOOOOOOO	Capaz

Tabela 3 - Principais diferenças entre o questionário *online* e o questionário presencial

Característica	Questionário	
	<i>Online</i>	Presencial
Descrição dos cenários	Apresentada de forma textual para que o participante fizesse a leitura do cenário e se imaginasse inserido na situação proposta.	Em forma de atividades, que foram descritas em um caderno contendo o passo-a-passo para realizar cada atividade proposta em cada cenário.
Vivência das falhas	Cada cenário continha uma descrição de como a falha ocorria e uma imagem da falha.	As falhas eram vivenciadas conforme os participantes tentavam desempenhar as atividades propostas.
Perguntas	Apresentadas logo abaixo de cada cenário e respondidas no próprio questionário <i>online</i> .	A descrição do cenário e a imagem com a falha foram removidas do questionário presencial, deixando apenas o nome do cenário em questão e as perguntas.
Ambiente	Não controlado	Controlado
Equipamento	Não controlado	Controlado

3.2.3 Instrumental

Para avaliar a percepção dos usuários em relação ao funcionamento do SO durante a aplicação do questionário presencial, foi necessário desenvolver um instrumental que simulasse as falhas do SO. O instrumental que foi utilizado neste estudo é uma versão adaptada do instrumental desenvolvido durante o trabalho de conclusão de curso de graduação do autor (ASSUNÇÃO, 2016).

Através do instrumental, foi possível expor os participantes aos eventos de falha e suas consequências durante as sessões de coleta de dados presencial. Todas as falhas foram simuladas de forma fidedigna às respectivas falhas reais. Entretanto, devido a problemas de desempenho do computador utilizado para executar o instrumental, alguns participantes tiveram que ser descartados devido à simulação da primeira falha não ter sido executada de maneira fiel à falha real (ver Apêndice C). Nas sessões em que o computador que executou o instrumental não apresentou problemas, os participantes não relataram ter percebido diferença entre as falhas simuladas e as falhas reais.

A primeira falha simulada provoca uma interrupção na inicialização do SO e executa o serviço *StartupRepair*, procedimento padrão do Windows 7 para tentar solucionar o problema. O participante liga o computador e, durante a inicialização do SO, aparece uma mensagem com fundo preto na tela, interrompendo a execução correta do SO. Os participantes foram instruídos a escolher a opção “Iniciar o Windows Normalmente”. Após a escolha dessa opção, o SO continua funcionando. A Figura 5 mostra a tela apresentada. A simulação dessa falha foi realizada criando um programa que mostra uma janela em tela cheia com aparência idêntica à tela original quando ocorre esta falha. Para que o programa do instrumental que simula essa

falha executasse antes do programa *userinit.exe*, que o SO executa logo após a inicialização, modificou-se o valor *Userinit* presente no registro (*registry*) do Windows. O valor de *Userinit*, que originalmente é “C:\Windows\system32\userinit.exe”, foi alterado para “C:\t1.exe;C:\Windows\system32\userinit.exe”, em que t1.exe é o programa executável do instrumental responsável por simular a falha. O valor *Userinit* do registro do Windows pode ser acessado pelo caminho: HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\WindowsNT\CurrentVersion\Winlogon dentro do editor de registro do Windows (*regedit*).

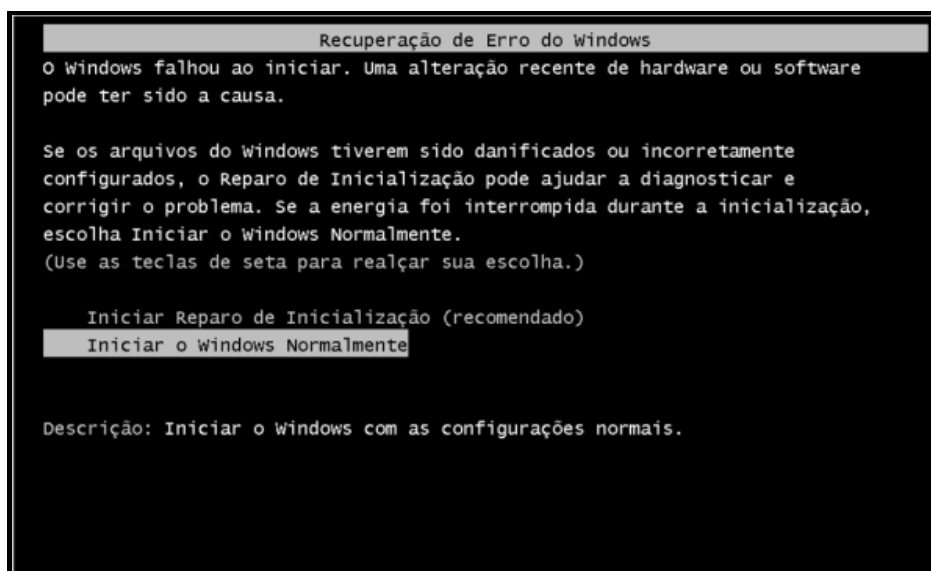


Figura 6 - Tela simulada da interrupção na inicialização do Windows

A segunda falha simulada é uma interrupção no aplicativo *explorer.exe*, que é responsável pelo gerenciamento de arquivos e pelas tarefas em GUI (*Graphic User Interface*). Essa falha geralmente ocorre quando o usuário tenta acessar alguma pasta na área de trabalho do Windows 7 e, em seguida, o *explorer.exe* “congela” a sua execução. Como consequência o usuário fica impedido de interagir com componentes gráficos da área de trabalho do sistema e até mesmo navegar na estrutura de diretórios por meio da interface gráfica. Simulou-se essa falha criando um *script* que encerra o processo *explorer.exe* e mascara o atalho do ícone do *script* na área de trabalho, alterando-o para um ícone de pasta (Figura 6). Assim, quando o participante clica no ícone tentando acessar uma pasta, a execução do *explorer.exe* é interrompida e uma mensagem de falha é mostrada na tela (Figura 7).

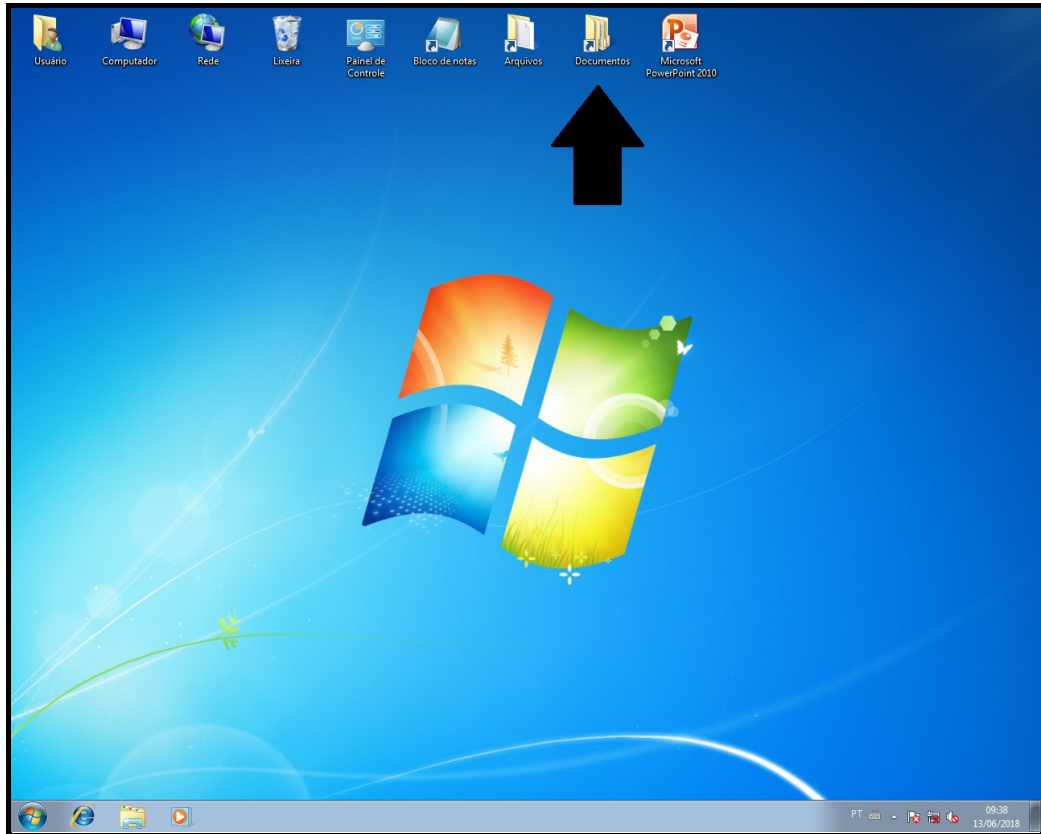


Figura 7 - Ícone de pasta no desktop que executa o *script*

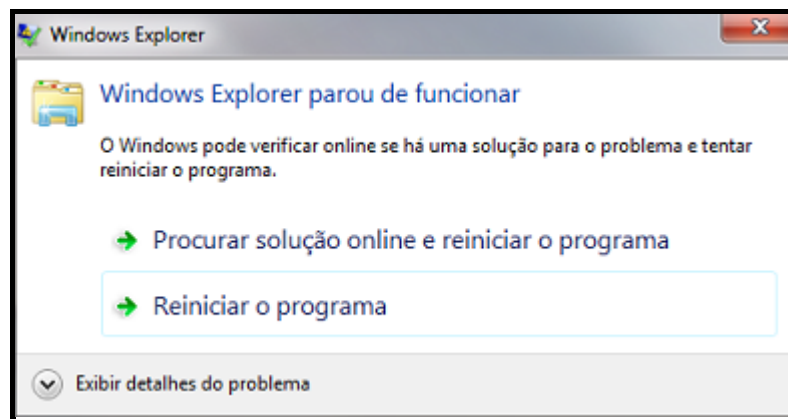


Figura 8 - Mensagem de falha da aplicação *explorer.exe*

A terceira falha simulada tem relação com o programa *mscorsvw.exe*. Esse programa pré compila o código das aplicações *.Net Framework* para melhorar o desempenho dessas aplicações durante sua execução (MICROSOFT, 2012). Essa falha foi apresentada quando um usuário, digitando algum texto em um editor de texto, subitamente tem a aplicação “congelada” e então é exibida uma janela de mensagem de falha. Após o usuário clicar em OK na janela da mensagem de falha, o editor de texto é fechado e o texto editado é perdido. Para simular essa falha, criou-se um programa com interface gráfica idêntica a do bloco de notas original (Figura 8), que após um minuto em execução tem sua tela “congelada” e então exibe a mensagem de falha (Figura 9).

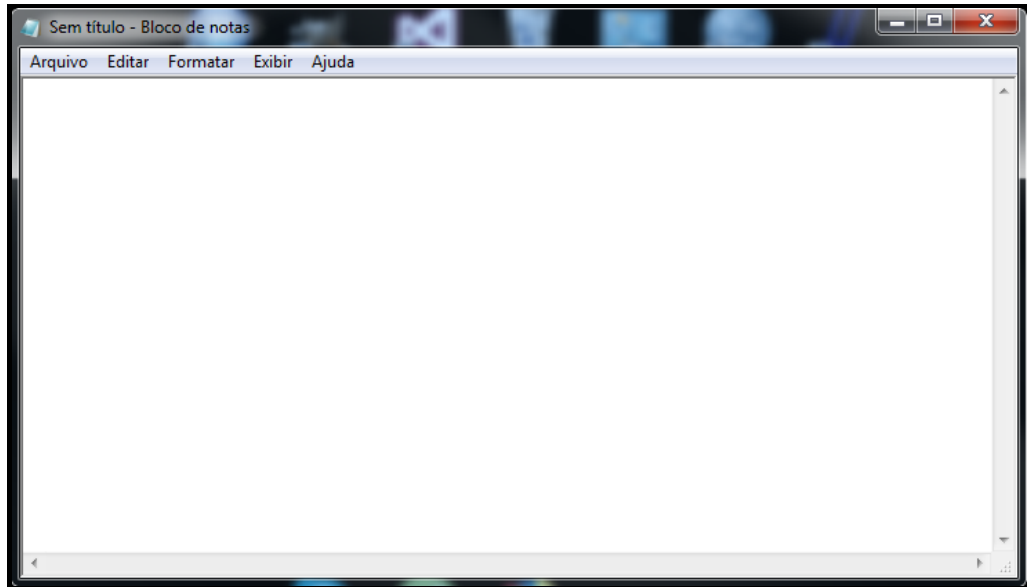


Figura 9 - Bloco de notas simulado antes da ocorrência da falha

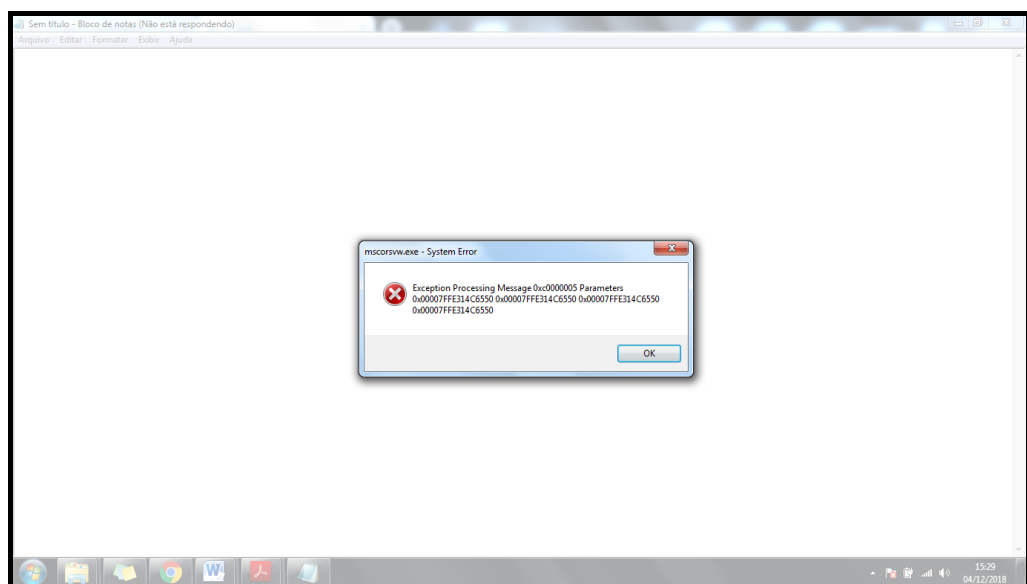


Figura 10 - Bloco de notas simulado após a ocorrência da falha

A quarta falha simulada, chamada `BAD_POOL_CALLER`, ocorre em espaço de *kernel* do SO. Essa falha acontece quando um *driver* de dispositivo em execução faz uma solicitação de um recurso indisponível, inexistente ou que está em uso por algum processo (MICROSOFT, 2017). A ocorrência da falha foi simulada enquanto o participante estava criando *slides* de uma apresentação no Microsoft PowerPoint 2010 (`POWERPNT.EXE`). Criou-se um programa que monitora a lista de processos em execução no SO e, no momento em que o programa detecta o processo `POWERPNT.EXE` na lista, um minuto após a detecção o teclado é bloqueado e uma mensagem de falha do tipo BSOD (*Blue Screen of Death*) é apresentada na tela (Figura 10), impedindo o participante de continuar a editar a apresentação de *slides*.


```

A problem has been detected and Windows has been shut down to prevent damage
to your computer.

The problem seems to be caused by the following file: usbser.sys

BAD_POOL_CALLER

If this is the first time you've seen this stop error screen,
restart your computer. If this screen appears again, follow
these steps:

Check to make sure any new hardware or software is properly installed.
If this is a new installation, ask your hardware or software manufacturer
for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware
or software. Disable BIOS memory options such as caching or shadowing.
If you need to use safe mode to remove or disable components, restart
your computer, press F8 to select Advanced Startup Options, and then
select Safe Mode.

Technical Information:

*** STOP: 0x000000c2 (0x0000000000000007, 0x000000000000109b, 0x0000000005070000,
0xfffffa80148b31bb)

*** usbser.sys - Address 0xfffff8800ce69b56 base at 0xfffff8800ce62000 DateStamp
0x4ce7a66d

```

Figura 11 - BSOD simulada para a falha BAD_POOL_CALLER

3.2.4 Participantes

O questionário *online* foi respondido por 411 sujeitos. No entanto, a amostra analisada foi composta por respostas de 400 participantes, pois 11 informaram ter idade inferior a 18 anos. O questionário presencial foi respondido por 102 participantes, dos quais 86 compuseram a amostra final. Neste caso, cada motivo de descarte dos participantes da amostra do questionário presencial está listado no Apêndice C. Neste trabalho a amostragem foi realizada por saturação (DENZIN, LINCOLN, 2005), ou seja, a coleta de dados foi interrompida quando os resultados obtidos com as análises dos dados passaram a apresentar redundância com o acréscimo de dados à amostra.

No estudo preliminar (YABE *et al.*, 2018), foram encontradas evidências estatísticas de que o nível de habilidades computacionais não influenciou a percepção do entrevistado sobre as falhas do SO nos diferentes cenários de falha considerados. Portanto, nesta dissertação não houve necessidade de estratificar os participantes com base em seu nível de conhecimento sobre o uso do computador/SO.

3.2.5 Ferramentas

O questionário foi desenvolvido utilizando a ferramenta LimeSurvey (LIMESURVEY, 2019). Esse questionário foi colocado em operação no servidor *web* do laboratório HPDCS (HPDCS, 2019) para a coleta *online*. Para coletar as respostas dos participantes nas sessões de coleta presencial, foi utilizado um computador *desktop* que executava o instrumental e um *notebook* com o questionário instalado localmente no qual os participantes respondiam as perguntas após vivenciarem as falhas (Figura 11). Para a análise estatística dos dados coletados foram utilizados o Microsoft Excel 2010 (Microsoft, 2019) e o IBM SPSS Statistics 25 (IBM, 2019). Para a elaboração das nuvens de palavras (ver Apêndice B) foi usado o *software* Iramuteq (LERASS, 2019).

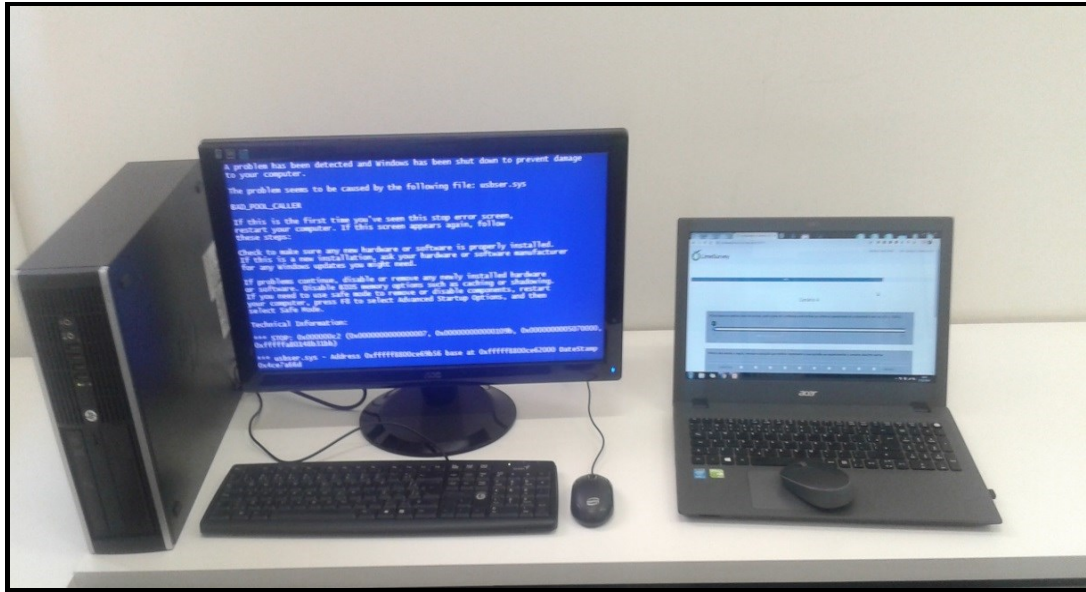


Figura 12 - Equipamentos utilizados nas sessões de coleta de dados do questionário presencial

3.3 Método

3.3.1 Teste Piloto

Para avaliar a qualidade dos materiais utilizados no estudo, foi realizado um teste piloto antes da aplicação dos questionários *online* e presencial. Participaram cinco pessoas na realização do teste piloto, nos quais todos eram alunos de pós-graduação da universidade, dentre eles quatro membros de outros laboratórios de pesquisa e um membro do laboratório no qual o autor da dissertação estava alocado, mas que não tinha relação direta com a pesquisa. Nesse teste piloto, avaliaram-se aspectos como: a duração da aplicação dos questionários (*online* e presencial), o *layout* das questões apresentadas, a clareza e objetividade dos cenários, a fidelidade da simulação das falhas, o funcionamento do equipamento utilizado, entre outros. Esses cinco participantes responderam a ambos os questionários e apresentaram as suas sugestões ao final de cada etapa. As principais sugestões dos participantes do teste piloto foram referentes ao *layout* do questionário, como troca de cores e tamanho das fontes utilizadas, posicionamento dos equipamentos no ambiente de coleta de dados, entre outras. Não foi realizado um teste piloto com mais participantes para evitar o desperdício de potenciais participantes da pesquisa oficial. Todas as sugestões apresentadas por esses participantes foram acatadas.

3.3.2 Recrutamento de Participantes

Para recrutar participantes de diferentes características (gênero, idade, grau de escolaridade, profissão), foi enviado um convite de participação de pesquisa via *e-mail* com um *link* para o questionário *online*. Esse convite foi compartilhado nas redes sociais do autor desta dissertação, na lista de e-mails da Sociedade Brasileira de Computação e enviado para os *e-mails* das coordenações de cursos de graduação e pós-graduação das seguintes universidades: Universidade Federal de Uberlândia, Universidade Federal da Bahia,

Universidade Federal de Minas Gerais, Universidade Federal de Viçosa e Universidade de São Paulo.

Após o término da coleta de dados do questionário *online*, foram iniciadas as sessões de coleta de dados presencial. Entrou-se em contato com os responsáveis de diversas empresas e universidades para que fosse possível realizar as coletas de dados nesses locais. Dentre as empresas e universidades contatadas, foram realizadas sessões de coleta de dados nas dependências da empresa Neppo Soluções em Tecnologia da Informação (11,63%) e na Universidade Federal de Viçosa – campus Rio Paranaíba (25,58%). Além desses locais, ocorreram sessões de coleta de dados no laboratório de pesquisa HPDCS (9,30%) do qual o autor faz parte e em outra sala da Universidade Federal de Uberlândia (53,49%). Nas universidades, os sujeitos eram convidados pessoalmente pelo autor desta dissertação a participarem da coleta de dados. Na empresa, uma funcionária do setor de recursos humanos direcionava os funcionários disponíveis para uma sala dedicada ao experimento.

Nesta dissertação, como as coletas de dados foram realizadas com sujeitos humanos, o projeto de pesquisa foi previamente analisado pelo Comitê de Ética em Pesquisas com Seres Humanos da Universidade Federal de Uberlândia, de acordo com as normas previstas em (RESOLUÇÃO 466, 2012). Somente após a aprovação do projeto pelo referido comitê é que os trabalhos de coleta foram iniciados (ver Anexo A). Além disso, o anonimato dos participantes foi respeitado, evitando coletar dados que pudessem associar o participante às suas respostas. Também é importante destacar que o sujeito humano que relatava ter idade inferior a 18 anos era informado que não poderia participar da sessão de coleta de dados.

3.3.3 Análise dos Dados

3.3.3.1 Análise Fatorial

Para analisar as respostas da segunda pergunta do questionário (ver Figura 4), foram utilizadas técnicas de análise fatorial (FÁVERO *et al.*, 2009). Estas são técnicas estatísticas multivariadas que tem com objetivo resumir as relações observadas entre um conjunto de variáveis inter-relacionadas para identificar fatores comuns (FÁVERO *et al.*, 2009). No âmbito deste trabalho, as técnicas de análise fatorial foram utilizadas para verificar quantos e quais fatores as dez dimensões presentes no questionário conseguem representar.

Para verificar a adequação dos dados à análise fatorial, foram calculadas duas estatísticas: o Teste de Esfericidade de Bartlett (PEREIRA, 2004) e o Teste de Kaiser-Meyer-Olkin (*KMO*) (PEREIRA, 2004). O Teste de Esfericidade de Bartlett é utilizado para avaliar a hipótese de que a matriz de correlação entre as dimensões é uma matriz identidade com determinante igual a 1. Em caso de a hipótese nula (a matriz de correlações é uma matriz identidade) não ser rejeitada, não é recomendado analisar os dados com técnicas de análise fatorial, pois isso significa que as dimensões não estão correlacionadas.

O Teste de *KMO* compara as correlações simples com as correlações parciais, o qual é calculado por meio da razão da soma dos quadrados das correlações de todas as dimensões dividida por essa mesma soma acrescentada da soma dos quadrados das correlações parciais

de todas as variáveis (PEREIRA, 2004). A expressão do Teste de *KMO* é mostrada abaixo (FÁVERO *et al.*, 2009):

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (4)$$

O valor da estatística do Teste de *KMO* varia entre 0 e 1 e avalia a adequação da amostra quanto ao grau de correlação parcial entre as variáveis. O valor de *KMO* próximo de 0 indica que a análise fatorial pode não ser adequada devido à fraca correlação entre as variáveis. Valores para a estatística *KMO* superiores a 0,60 indicam que a análise fatorial pode ser adequada (FÁVERO *et al.*, 2009).

Após a verificação da viabilidade em utilizar análise fatorial para analisar os dados, foi feita a extração dos fatores que poderiam representar a estrutura das dimensões originais. Existem duas técnicas mais usadas para a extração dos fatores: análise de componentes principais e análise de fatores comuns (HAIR *et al.*, 2009). A análise de componentes principais considera a variância total e deriva fatores que contém pequenas proporções de variância única e, em alguns casos, variância de erro. Já a análise de fator comum considera apenas variância em comum ou compartilhada, assumindo que tanto a variância de erro quanto a única não são de interesse na definição da estrutura das variáveis. A análise de componentes principais é usada quando o objetivo é resumir a maior parte da informação original (variância) a um número mínimo de fatores para fins de previsão. Em contraste, a análise de fator comum é usada prioritariamente para identificar fatores ou dimensões latentes que refletem o que as variáveis têm em comum. O objetivo da análise fatorial neste trabalho foi resumir a maior parte da informação original extraíndo a menor quantidade de fatores possível. Portanto, a análise de componentes principais mostrou-se mais adequada que a análise de fator comum.

Com o número de fatores extraído, o próximo passo foi avaliar a retenção de fatores. A retenção de fatores indica a quantidade de fatores que devem ser utilizados e quais dimensões compõem cada fator, auxiliando o pesquisador a analisar as relações entre as dimensões e os fatores investigados. Neste trabalho, foram adotados três critérios para verificar a retenção de fatores: o critério da raiz latente, o critério de porcentagem de variância e o critério do gráfico de escarpa (HAIR *et al.*, 2009). Pelo critério da raiz latente, escolhe-se o número de fatores a reter em função do número de autovalores acima de 1. Os autovalores mostram quanto cada fator consegue explicar da variância total (HAIR *et al.*, 2009). O critério de porcentagem de variância consiste em escolher o número mínimo de fatores necessário para que o percentual de variância explicada alcance o nível satisfatório desejado (HAIR *et al.*, 2009). Em (HAIR *et al.*, 2009) é recomendado adotar no mínimo 60% de variância total explicada. O critério do gráfico de escarpa é usado para identificar o número ótimo de fatores que podem ser extraídos antes que a quantidade de variância única comece a dominar a estrutura de variância comum (HAIR *et al.*, 2009). A avaliação desse critério é feita plotando o gráfico das raízes latentes em relação ao número de fatores em sua ordem de extração, analisando a forma da curva resultante para avaliar o ponto de corte (HAIR *et al.*, 2009). O

ponto no qual o gráfico começa a tomar forma horizontal (ponto de inflexão) indica a quantidade máxima de fatores a serem retidos.

Por fim, o último passo é a aplicação de rotação dos fatores. A aplicação de um método de rotação tem como objetivo principal a transformação dos coeficientes dos componentes principais retidos em uma estrutura simplificada (FÁVERO *et al.*, 2009), ou seja, auxilia o pesquisador a interpretar os fatores retidos. Os métodos de rotação podem ser ortogonais ou oblíquos (HAIR *et al.*, 2009). Os métodos ortogonais produzem fatores que não estão correlacionados entre si, sendo interpretados a partir de suas cargas fatoriais. Na rotação oblíqua, os fatores estão correlacionados e, para a interpretação da solução, torna-se necessária a consideração simultânea das correlações e das cargas fatoriais. Nas análises realizadas neste trabalho não foi necessário executar nenhum método de rotação porque apenas um fator foi retido na análise dos dados de ambos os questionários (ver Capítulo 4).

3.3.3.2 Ajuste e Análise dos Modelos

Foi utilizada regressão linear múltipla para elaborar os modelos de confiabilidade com os dados obtidos com o questionário. A regressão linear múltipla consiste em ajustar a relação entre duas ou mais variáveis explicativas (variáveis independentes), que se apresentam na forma linear, e uma variável dependente métrica (FÁVERO *et al.*, 2009).

Um modelo geral de regressão linear pode ser descrito da seguinte forma (FÁVERO *et al.*, 2009):

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + u \quad (5)$$

no qual Y é a variável dependente, α representa o intercepto (constante), β_k ($k = 1, 2, \dots, n$) são os coeficientes de cada variável, X_k são as variáveis independentes e u é a diferença entre o valor real de Y e o valor previsto de Y pelo modelo para cada observação (termo do erro).

Para fazer o ajuste dos modelos, foi utilizada a média da confiabilidade percebida nos quatro cenários por cada participante como a variável dependente e as médias dos valores de cada dimensão por participante como as variáveis independentes. As seguintes técnicas foram aplicadas para avaliar os modelos ajustados:

- R^2 e R^2 ajustado,
- Teste t,
- Teste de Durbin-Watson,
- *VIF (Variance Inflation Factor) e Tolerance.*

Em (WOOLDRIGE, 2003) é relatado que o R^2 é considerado como a proporção da variação amostral da variável dependente explicada pelas variáveis explicativas, e é utilizado como uma medida do grau de ajuste. Esta medida mostra o quanto o comportamento das variáveis independentes explica a variação da variável dependente e não, necessariamente, justifica uma relação de causalidade da variável dependente pelas variáveis independentes. O R^2 é calculado da seguinte maneira (FÁVERO *et al.*, 2009):

$$R^2 = \left(\frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}} \right)^2 \quad (6)$$

no qual X é o vetor de variáveis independentes e Y é a variável dependente. Como a estimativa amostral do R^2 tende a superestimar o parâmetro populacional, deve-se analisar também o valor do R^2 ajustado, que é uma medida do R^2 da regressão do método de mínimos quadrados ordinários ajustada pelo número de graus de liberdade. O cálculo do R^2 ajustado é feito com a expressão (FÁVERO *et al.*, 2009):

$$R_{ajus}^2 = 1 - \frac{n-1}{n-k} (1 - R^2) \quad (7)$$

em que n é o tamanho da amostra e k o número de parâmetros do modelo de regressão.

Foi utilizado o Teste t (FÁVERO *et al.*, 2009) para avaliar se as variáveis independentes dos modelos de regressão são estatisticamente significantes, bem como verificar se a relação entre cada parâmetro com a variável dependente é diferente de zero. As expressões a seguir calculam a significância estatística de cada parâmetro α e β_i (FÁVERO *et al.*, 2009):

$$t_\alpha = \frac{\alpha}{s.e.(\alpha)} \quad (8)$$

$$t_\beta = \frac{\beta_i}{s.e.(\beta_i)} \quad (9)$$

em que $s.e.$ significa o erro padrão de cada parâmetro analisado. As hipóteses do Teste t para o intercepto e para os betas são apresentadas abaixo:

$$\begin{aligned} H_0: \alpha &= 0 \\ H_1: \alpha &\neq 0 \\ H_0: \beta_i &= 0 \\ H_1: \beta_i &\neq 0 \end{aligned}$$

Pode-se utilizar o p -valor para verificar se os testes indicam a rejeição ou não rejeição da hipótese nula a um dado nível de significância. Por exemplo, considerando 95% de confiança:

Se p – valor $< 0,05$ para o intercepto, $\alpha \neq 0$;

Se p – valor $< 0,05$ para determinada variável X_i , $\beta_i \neq 0$.

Para verificar a existência de autocorrelação entre os resíduos (quando o resíduo incorpora os efeitos de uma variável que foi excluída do modelo) utilizou-se o Teste de Durbin-Watson (OTT; LONGNECKER, 2016):

$$d = \frac{\sum_{t=1}^{n-1} (e_{t+1} - e_t)^2}{\sum_t e_t^2} \quad (10)$$

em que e_t corresponde aos valores dos resíduos estimados pelo modelo em determinado instante de tempo t e n é o número de observações. Desenvolvendo a equação 10, tem-se que:

$$d = 2(1 - \hat{\rho}) \quad (11)$$

na qual:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (12)$$

As hipóteses do Teste de Durbin-Watson são:

$$H_0: \hat{\rho} = 0$$

$$H_1: \hat{\rho} \neq 0$$

Para considerar a ausência de autocorrelação entre os resíduos, a estatística do Teste de Durbin-Watson (ver Equação) deve estar inserida no intervalo $[1,5 - 2,5]$ (OTT; LONGNECKER, 2016).

Algumas das variáveis consideradas em um modelo podem apresentar elevada correlação entre si, devido ao número insuficiente de observações ou a uma amostra coletada por conveniência (FÁVERO *et al.*, 2009). A ocorrência desse fenômeno, denominado multicolinearidade, pode gerar modelos com menor desempenho de predição. Com o intuito de diagnosticar a presença de multicolinearidade nos modelos ajustados, foram utilizadas as estatísticas *VIF* e *Tolerance*. A estatística *VIF* é uma medida de quanto a variância de cada coeficiente de regressão estimado aumenta devido a colinearidade. *Tolerance* indica a proporção da variação de uma variável independente que não depende das outras variáveis independentes, ou seja, se o valor de *Tolerance* for baixo, significa que a variável explicativa analisada compartilhará um percentual elevado de sua variância com as outras variáveis independentes. Segue abaixo as expressões que calculam *VIF* e *Tolerance* (FÁVERO *et al.*, 2009):

$$Tolerance = 1 - R_k^2 \quad (13)$$

$$VIF = \frac{1}{Tolerance} \quad (14)$$

em que o R_k^2 é o coeficiente de ajuste de regressão da variável independente k com as demais variáveis independentes. Um valor de VIF superior a 10 indica a presença de multicolinearidade, assim como um valor de $Tolerance$ inferior a 0,1.

3.3.3.3 Comparação entre as Dimensões dos Cenários

Para verificar se houve ou não diferença significativa nas respostas das dimensões entre os cenários, foi utilizado o Teste de Kruskal-Wallis para amostras independentes (OTT; LONGNECKER, 2016). Abaixo estão as hipóteses do Teste de Kruskal-Wallis:

H_0 : A distribuição da dimensão D é a mesma entre todos os cenários

H_1 : A distribuição da dimensão D não é a mesma entre todos os cenários

em que D é uma das dimensões consideradas na segunda pergunta dos cenários. A estatística do Teste de Kruskal-Wallis pode ser obtida através da equação (OTT; LONGNECKER, 2016):

$$H = \frac{12}{n_T(n_T + 1)} \sum_i \frac{T_i^2}{n_i} - 3(n_t + 1) \quad (15)$$

na qual n_i é o número de observações da amostra i ($i = 1, 2, \dots, k$), n_T é o somatório combinado do tamanho da amostra (ou seja, $n_T = \sum_i n_i$) e T_i é a soma dos ranques para as medidas na amostra i após as amostras combinadas terem sido ranqueadas.

Resultados

4.1 Questionário *online*

4.1.1 Etapa 1

A amostra obtida com o questionário *online* é composta por 400 participantes, sendo 229 do gênero masculino e 171 do gênero feminino (Tabela 4).

Tabela 4 - Percentual de participantes por gênero (questionário *online*)

Gênero	Quantidade	Percentual
Feminino	171	43,00%
Masculino	229	57,00%

A Tabela 5 mostra a distribuição da faixa etária dos participantes. Observa-se maior concentração nas faixas etárias 25 – 44 anos e 18 – 24 anos, com 197 e 159 participantes respectivamente. Na faixa etária 45 – 65 anos participaram 43 participantes e apenas um na faixa etária acima de 65 anos.

Tabela 5 - Percentual de participantes por faixa etária (questionário *online*)

Faixa Etária	Quantidade	Percentual
18 a 24 anos	159	40,00%
25 a 44 anos	197	49,00%
45 a 65 anos	43	11,00%
acima de 65 anos	01	0,00%

A Figura 12 mostra o grau de escolaridade dos participantes, na qual se observa maior concentração em Pós-graduação, com 172 participantes e em Superior incompleto, com 144 participantes. Participaram 61 pessoas da categoria Superior completo e 23 da categoria Ensino médio completo. Não participaram pessoas das categorias: Ensino médio incompleto, Ensino fundamental completo e Ensino fundamental incompleto.

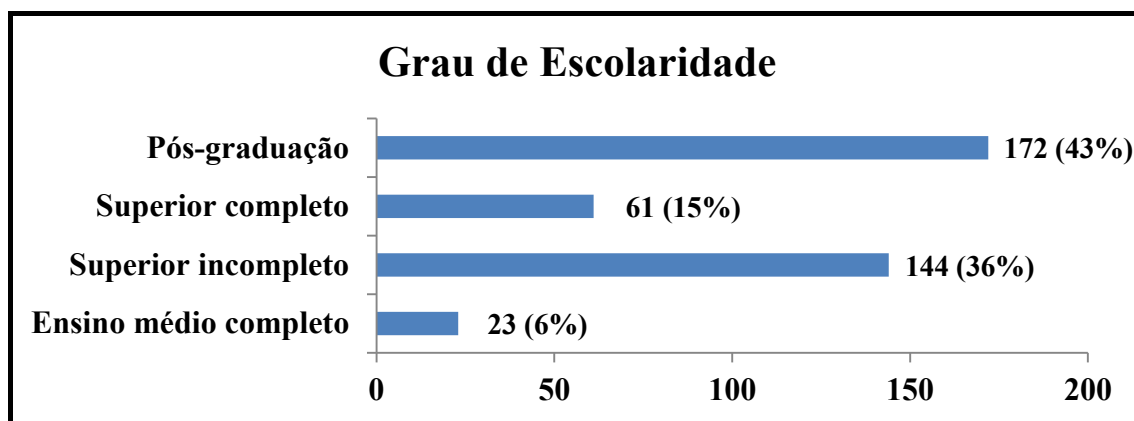


Figura 12 - Percentual de participantes por grau de escolaridade (questionário *online*)

A Tabela 6 mostra os percentuais dos participantes por área de atividade. As três áreas com maior predominância foram: Computação com 122 participantes, Ciências da Saúde com 67 e Engenharia com 34. Essas três áreas compõem cinquenta e sete por cento da amostra. A categoria Outras áreas agrupa 22 áreas com frequência inferior a 4 participantes.

Tabela 6 – Percentual de participantes por área de atividade (questionário *online*)

Categoria	Frequência	Porcentagem
Computação	122	31,00%
Ciências da Saúde	67	17,00%
Outras áreas	39	09,00%
Engenharia	34	09,00%
Educação	31	08,00%
Ciências Biológicas	27	07,00%
Biblioteconomia	24	06,00%
Estudante	19	05,00%
Administração	13	03,00%
Direito	09	02,00%
Arquitetura	05	01,00%
Comércio	05	01,00%
Economia	05	01,00%

4.1.2 Etapa 2

A Tabela 7 mostra a confiabilidade calculada com base na abordagem clássica (modelo de Poisson), por cenário, e as respectivas estatísticas descritivas da confiabilidade percebida pelos participantes. Observa-se que a média ($\bar{R}_{per}(t)$) e a mediana ($\tilde{R}_{per}(t)$) da confiabilidade percebida foi maior do que a confiabilidade calculada no cenário 1 e menor nos outros cenários. Pode-se atribuir essa diferença à recuperação rápida do sistema e à consequência relatada no cenário 1, na qual o usuário consegue continuar utilizando o computador normalmente após a ocorrência da falha, sem perda de arquivos. Nota-se que a média e a mediana da confiabilidade percebida nos cenários de baixo impacto (1 e 2) atingiu o

valor máximo de 50%, enquanto que nos cenários de alto impacto (3 e 4) esse valor não chegou a 25%. Essa é uma evidência de que a consequência possui maior influência sobre a percepção dos usuários que o MTBF. Entretanto, ao analisar a subtração entre a confiabilidade estimada pelo modelo de Poisson e a confiabilidade percebida média, bem como a subtração em relação à confiabilidade percebida mediana, observa-se que a diferença maior entre os valores ocorreu nos cenários 2 (MTBF longo, baixo impacto) e 4 (MTBF longo, alto impacto). Pode-se concluir que o MTBF longo possui maior influência na percepção dos usuários que o MTBF curto. A moda da confiabilidade percebida ($\tilde{R}_{Per}(t)$) foi 50% nos cenários 1 e 2, 10% no cenário 3 e 1% no cenário 4. Os cenários 3 e 4 apresentaram o mesmo tipo de consequência (o usuário perde o documento que estava editando), porém com MTBF e mensagem de falha diferente. O que pode ter contribuído para o cenário 4 ter sido percebido pela maioria dos participantes com apenas 1% de confiabilidade é a BSOD gerada pela falha relatada neste cenário, que pode provocar no usuário a impressão de estar vivenciando uma falha mais grave que as outras vivenciadas anteriormente. A confiabilidade percebida atingiu um valor máximo ($R_{Per}^{max}(t)$) de 100% e um valor mínimo ($R_{Per}^{min}(t)$) de 0% em todos os cenários, indicando que alguns usuários possuem confiabilidade total no funcionamento do SO, enquanto que outros usuários não confiam de forma alguma, apresentando total aversão às falhas. Por fim, pode-se observar que, embora o resultado do desvio padrão ($R_{per}^{dp}(t)$) decresceu do cenário 1 para o cenário 4, seu valor manteve-se acima de 20% em todos os cenários, indicando que as observações estão dispersas em relação à média.

Tabela 7 - Estatísticas descritivas da confiabilidade percebida pelos participantes (questionário *online*)

Impacto	Baixo		Alto	
	1	2	3	4
Cenário	1	2	3	4
MTBF	Curto	Longo	Curto	Longo
$R_{(t)}$	36,80%	81,00%	36,80%	93,50%
$\bar{R}_{Per}(t)$	46,58%	42,37%	22,15%	18,95%
$R_{(t)} - \bar{R}_{Per}(t)$	-09,78%	38,63%	14,65%	74,55%
$\tilde{R}_{Per}(t)$	50,00%	41,00%	15,00%	10,00%
$R_{(t)} - \tilde{R}_{Per}(t)$	-13,20%	40,00%	21,80%	83,50%
$\check{R}_{Per}(t)$	50,00%	50,00%	10,00%	01,00%
$R_{per}^{max}(t)$	100,00%	100,00%	100,00%	100,00%
$R_{per}^{min}(t)$	00,00%	00,00%	00,00%	00,00%
$R_{per}^{dp}(t)$	27,48%	25,07%	21,56%	20,70%

4.1.2.1 Análise Fatorial dos Dados do Questionário *Online*

O Teste de *KMO* para a amostra do questionário *online* resultou em 0,920 e o Teste de Esfericidade de Bartlett em (Graus de Liberdade (GL)(45) = 3.969,500; p-valor < 0,005), indicando ótima adequação dos dados à análise fatorial devido à forte correlação entre as dimensões (PEREIRA, 2004). A análise fatorial indicou a retenção de apenas um fator que explica 69,73% da variância total, composto pelas dez dimensões analisadas (Tabela 8). Isso mostra que as dez dimensões possuem relação entre si e estão mensurando um mesmo conceito. No âmbito deste trabalho, tal relacionamento significa que, de acordo com a análise fatorial, as dez dimensões podem ser utilizadas para mensurar a confiabilidade percebida pelos usuários.

A Figura 14 mostra o gráfico de escarpa para os fatores extraídos. Observando o gráfico, nota-se que apenas o primeiro fator extraído apresentou autovalor maior que 1 (critério da raiz latente), o que indica reter apenas esse fator. Além disso, o ponto de inflexão no gráfico é o segundo fator extraído. De acordo com o critério do gráfico de escarpa, apenas os fatores extraídos que estão acima do ponto de inflexão devem ser retidos. Portanto, o critério do gráfico de escarpa também indica a retenção de apenas um fator, assim como os outros critérios analisados (porcentagem de variância e critério da raiz latente).

Tabela 8 – Porcentagem de variância explicada por fator extraído (questionário *online*)

Fator	Porcentagem de variância	Porcentagem acumulada
1	69,73	69,73
2	09,15	78,88
3	05,60	84,49
4	04,53	89,03
5	02,70	91,73
6	02,27	94,00
7	01,90	95,90
8	01,54	97,45
9	01,40	98,85
10	01,14	100,00

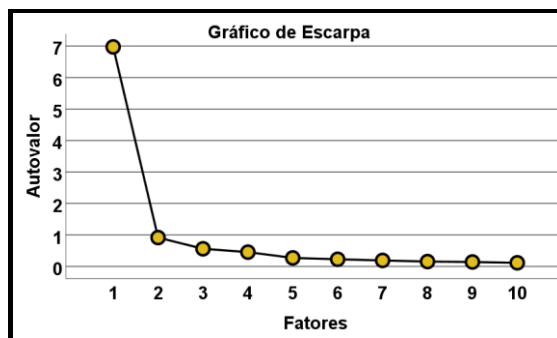


Figura 13 - Gráfico de escarpa para os fatores extraídos (questionário *online*)

Com a quantidade de fatores extraídos, é necessário verificar quais as dimensões são mais relevantes na composição do fator analisando o valor de comunalidade de cada dimensão. A Tabela 9 mostra os valores de comunalidade para as dez dimensões que compuseram o fator extraído na análise fatorial. Todas as dimensões apresentaram valor de comunalidade superior a 0,3, o que indica que tais dimensões estão relacionadas com o fator extraído. No âmbito deste trabalho, essas dimensões serão utilizadas para a elaboração de modelos de predição de confiabilidade percebida pelos usuários.

Tabela 9 – Valor de comunalidade por dimensão (questionário *online*)

Dimensão	Comunalidade
Satisfação	0,56
Preocupação	0,62
Amparo	0,71
Tolerância	0,73
Produtividade	0,74
Estresse	0,77
Pontualidade	0,77
Paciência	0,78
Tranquilidade	0,67
Capacidade	0,58

4.1.2.2 Ajuste dos modelos de regressão linear múltipla com os dados do questionário *online*

No ajuste do primeiro modelo de predição de confiabilidade percebida, serão consideradas três dimensões com maior valor de comunalidade e, conseqüentemente, as dimensões mais relevantes na composição do fator. Analisando os valores da Tabela 9, observa-se que as três dimensões com maior comunalidade são: Paciência, Estresse e Pontualidade. Para ajustar o modelo, foi calculado o escore médio dessas três dimensões por cenário para cada usuário. Por exemplo: um usuário hipotético atribuiu para a dimensão Paciência escore 8 no cenário 1, 6 no cenário 2, 4 no cenário 3 e 2 no cenário 4, obtendo escore médio 5 ($((8+6+4+2)/4 = 5)$) na dimensão Paciência para esse usuário. Esse escore médio também foi calculado para a confiabilidade percebida pelos usuários nos cenários. Sendo assim, a confiabilidade percebida média por cada participante foi utilizada como **variável dependente** e os escores médios das dimensões Estresse, Pontualidade e Paciência

foram utilizados como **variáveis independentes** para ajustar o modelo 1 de confiabilidade predita mostrado em (16):

$$R_{Pred_1}(t) = 14,869 + 2,350 * a + 3,572 * b - 0,210 * c \quad (16)$$

no qual a é o escore de Estresse, b é o escore de Pontualidade, c é o escore de Paciência e t é o momento de ocorrência da falha. A Tabela 10 mostra o ajuste do modelo descrito em (16) considerando uma, duas e três dimensões. O campo Constante representa o ponto em que a reta de regressão passa pelo eixo y no plano cartesiano. No âmbito deste trabalho, a constante representa o valor mínimo de confiabilidade predita sem considerar os escores das dimensões presentes no modelo. Pode-se observar que houve melhora do modelo ao considerar duas dimensões (Estresse e Pontualidade) ao invés de apenas uma (Estresse) analisando o valor de R^2 e R^2 ajustado (R^2 ajus). Entretanto, o mesmo não ocorre ao considerar as três dimensões, pois o valor de R^2 não é alterado e o R^2 ajustado é reduzido em 0,002. De acordo com a literatura, o resultado do R^2 e R^2 ajustado deve ser utilizado para avaliar o quanto o poder de predição do modelo ajustado melhora com a inserção de novas variáveis, não sendo recomendado analisar esse valor como indicativo definitivo da precisão de um modelo (FÁVERO *et al.*, 2009). Sendo assim, a precisão dos modelos ajustados neste trabalho será analisada pelo o quão próximo o resultado produzido por um modelo se aproxima da confiabilidade percebida pelos usuários.

O resultado do Teste de Durbin-Watson (DW) indicou a independência dos resíduos, o que mostra que o modelo é apropriado. O resultado de VIF e $Tolerance$ (Tol) mostrou a ausência de multicolinearidade entre as dimensões em todos os casos. Apesar disso, o nível de significância (p-valor < 0,005) indicou que a inserção da dimensão Paciência (Teste t = 0,830) não melhorou o poder de predição do modelo. Pode-se concluir que apesar da dimensão Paciência ter sido indicada como a de maior relevância para a composição do fator na análise fatorial, essa dimensão não foi adequada para melhorar o poder de predição do Modelo 1.

Tabela 10 – Resumo estatístico do Modelo 1

Ajuste	Coef.	R^2	R^2 ajus	DW	Teste t	Sig. Teste t	VIF	Tol
(Constante)	17,900				9,710	0,000	-	-
Estresse	4,596	0,206	0,204	-	10,154	0,000	1,000	1,000
(Constante)	14,858				7,829	0,000	-	-
Estresse	2,208	0,251	0,247	-	3,356	0,001	2,234	0,448
Pontualidade	3,518				4,885	0,000	2,234	0,448
(Constante)	14,869				7,822	0,000	-	-
Estresse	2,350	0,251	0,245	2,069	2,517	0,012	4,487	0,223
Pontualidade	3,572				4,681	0,000	2,501	0,400
Paciência	-0,210				-0,215	0,830	4,675	0,214

O segundo modelo ajustado foi elaborado em duas etapas. Na primeira etapa foi ajustado um modelo considerando as dez dimensões. Então foram selecionadas as dimensões mais relevantes apontadas pelo Teste t (p-valor < 0,005) para que o segundo modelo fosse produzido considerando tais dimensões. A Tabela 11 mostra o nível de significância no Teste t das dez dimensões ajustadas. Os resultados do Teste t apontaram as dimensões Satisfação, Tolerância e Paciência como as mais relevantes.

Tabela 11 - Nível de significância no Teste t das dez dimensões utilizadas na primeira etapa de ajuste do Modelo 2 (questionário *online*)

Dimensão	Significância
Satisfação	0,000
Preocupação	0,737
Amparo	0,874
Tolerância	0,000
Produtividade	0,230
Estresse	0,213
Pontualidade	0,687
Paciência	0,001
Tranquilidade	0,630
Capacidade	0,343

A Tabela 12 mostra o ajuste do segundo modelo considerando uma, duas e três dimensões. O resultado do Teste de Durbin-Watson para o segundo modelo também indicou a independência dos resíduos. Pode-se observar que houve melhora do modelo ao considerar duas dimensões (Satisfação e Tolerância) e uma melhora menor ao considerar as três dimensões (Satisfação, Tolerância e Paciência) analisando o resultado do R^2 e do R^2 ajustado. Entretanto, ao analisar o nível de significância de cada dimensão individualmente através do resultado do Teste t, observa-se que o nível de significância da dimensão Paciência (Teste t = 0,008 > p = 0,005) indicou a remoção de tal dimensão do modelo. Portanto, o segundo modelo foi ajustado apenas com as dimensões Satisfação e Tolerância, sendo mostrado em (17):

$$R_{Pred_2}(t) = 9,789 + 5,539 * c + 2,839 * d \quad (17)$$

no qual c é o escore de Satisfação, d é o escore de Tolerância e t é o momento de ocorrência da falha.

Tabela 12 – Resumo estatístico do Modelo 2

Modelo 2								
Ajuste	Coef.	R^2	R^2 ajust	DW	Teste t	Sig. Teste t	VIF	Tol
(Constante)	14,398	0,353	0,351	-	9,110	0,000	-	-
Satisfação	7,743				14,731	0,000	1,000	1,000
(Constante)	9,789	0,390	0,387	-	5,446	0,000	-	-
Satisfação	5,539				8,165	0,000	1,763	0,567
Tolerância	2,839				4,938	0,000	1,763	0,567
(Constante)	14,869	0,401	0,396	1,949	5,535	0,000	-	-
Satisfação	2,350				8,641	0,000	1,925	0,519
Tolerância	3,572				5,508	0,000	3,026	0,330
Paciência	-0,210				-2,647	0,008	3,018	0,331

A Tabela 13 mostra uma comparação entre a confiabilidade estimada por método clássico ($R_{(t)}$), a média ($\bar{R}_{Per}(t)$) e mediana ($\tilde{R}_{Per}(t)$) da confiabilidade percebida pelos participantes e a confiabilidade predita pelo Modelo 1 ($R_{Pred_1}(t)$) e pelo Modelo 2 ($R_{Pred_2}(t)$) nos quatro cenários do questionário *online*. Analisando as diferenças entre ($\bar{R}_{Per}(t)$), ($R_{(t)}$), ($R_{Pred_1}(t)$) e ($R_{Pred_2}(t)$), nota-se que o resultado calculado com ambos os modelos aproxima-se mais da confiabilidade percebida pelos usuários que o resultado obtido pelo método clássico, com o Modelo 2 apresentando maior aproximação que o Modelo 1. O mesmo ocorre ao fazer essa análise com ($\tilde{R}_{Per}(t)$), porém observa-se que nesse caso a precisão dos modelos é reduzida em relação à comparação feita utilizando a ($\bar{R}_{Per}(t)$). Observa-se que o valor de confiabilidade obtido com os dois modelos é menor que a confiabilidade percebida para os cenários 1 e 2, definidos com baixo impacto, porém para os cenários 3 e 4, definidos com alto impacto, o valor de confiabilidade obtido com ambos os modelos é maior que o valor de confiabilidade percebida. Com essas análises, conclui-se que ao calcular a confiabilidade de um sistema considerando fatores de QoS e QoE (análise quanti-qualitativa), consegue-se obter um resultado que se aproxima melhor da confiabilidade percebida pelos usuários (QoE - análise qualitativa) que apenas considerando fatores de QoS (análise qualitativa). Além disso, observa-se que considerar as dimensões Satisfação e Tolerância na elaboração de modelos de predição de confiabilidade produz modelos mais precisos que considerar as dimensões Estresse, Pontualidade e Paciência. Portanto, nesse caso as dimensões apontadas pela análise fatorial como as mais relacionadas com o fator não foram as dimensões que compuseram o modelo de predição mais preciso.

Tabela 13 - Comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelos Modelos 1 e 2 (questionário *online*)

Impacto	Baixo		Alto	
	1	2	3	4
MTBF	Curto	Longo	Curto	Longo
$R_{(t)}$	36,80%	81,00%	36,80%	93,50%
$\bar{R}_{Per}(t)$	47,63%	43,47%	23,83%	20,71%
$\tilde{R}_{Per}(t)$	50,00%	43,00%	16,00%	11,00%
$R_{Pred_1}(t)$	40,69%	37,64%	28,90%	28,41%
$R_{Pred_2}(t)$	44,75%	37,95%	26,67%	26,27%
$\bar{R}_{Per}(t) - R_{(t)}$	10,83%	-37,53%	-12,97%	-72,79%
$\bar{R}_{Per}(t) - R_{Pred_1}(t)$	06,94%	05,83%	-05,07%	-07,70%
$\bar{R}_{Per}(t) - R_{Pred_2}(t)$	02,88%	05,52%	-02,84%	-05,56%
$\tilde{R}_{Per}(t) - R_{(t)}$	13,20%	-38,00%	-20,80%	-82,50%
$\tilde{R}_{Per}(t) - R_{Pred_1}(t)$	09,31%	05,36%	-12,90%	-17,41%
$\tilde{R}_{Per}(t) - R_{Pred_2}(t)$	05,25%	05,05%	-10,67%	-15,27%

Analisando os intervalos de confiança de 95% mostrados na Figura 15 pode-se observar que a confiabilidade estimada por método clássico não se aproxima da confiabilidade percebida pelos participantes em nenhum cenário, com maior diferença nos cenários definidos com MTBF longo (cenários 2 e 4). Pode-se concluir que um valor de MTBF longo impacta positivamente na confiabilidade estimada por método clássico (QoS), mas que o mesmo não ocorre com a confiabilidade percebida pelos usuários (QoE), que é impactada negativamente quando os usuários vivenciam falhas com MTBF longo. Observa-se que os valores da média e mediana se aproximam dentro do intervalo de confiança nos cenários 1 e 2, definidos com baixo impacto. Entretanto, o mesmo não ocorre nos cenários 3 e 4, definidos com alto impacto. Isso indica que embora a maioria dos usuários percebesse o sistema com confiabilidade baixa nos cenários de alto impacto, alguns usuários perceberam o sistema com alta confiabilidade, o que elevou a média nesse caso. Esse fenômeno pode ter ocorrido devido à familiaridade que esses usuários possuem com as falhas apresentadas nos cenários, que impacta positivamente na QoE dos usuários, conforme relatado em (YABE *et al.*, 2018).

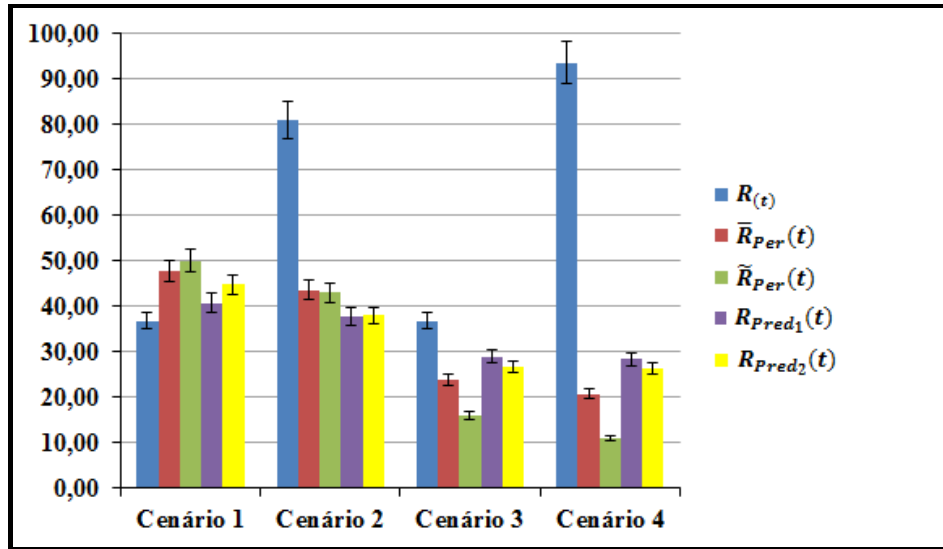


Figura 14 - Intervalo de confiança de 95% para os resultados dos cálculos de confiabilidade (questionário *online*)

4.1.2.3 Comparação entre as dimensões dos cenários do questionário *online*

Para examinar as diferenças entre os cenários por dimensão foi utilizado o Teste de Kruskal-Wallis. A Tabela 14 mostra os resultados do teste para cada par de cenários sem diferença estatística de acordo com o teste. Portanto, caso uma combinação não estiver na tabela significa que houve diferença estatística na percepção de um cenário em comparação com outro. Por exemplo, na dimensão Tranquilidade a Tabela 14 não mostra o par de cenários 2 e 3 (C2 – C3), portanto, houve diferença significativa da dimensão Tranquilidade entre os cenários 2 e 3.

Observa-se que os cenários 3 e 4, definidos com alto impacto na QoE dos usuários, não apresentaram diferenças significativas em todas as dimensões. Portanto, pode-se concluir que os cenários 3 e 4 foram percebidos de forma semelhante pelos usuários. O cenário 1 e o cenário 2 não apresentaram diferença significativa nas dimensões Preocupação, Amparo, Estresse, Pontualidade, Paciência e Tranquilidade. Esse fenômeno pode ter acontecido devido a esses cenários serem definidos com baixo impacto. O fato da dimensão Amparo não ter apresentado diferença significativa pode indicar que as falhas vivenciadas pelos usuários nos cenários 1 e 2 são solucionadas de forma que os usuários não se sentem desamparados pelo sistema.

Nota-se que em todas as dimensões avaliadas as combinações entre cenários de baixo e alto impacto apresentaram diferença estatística, ou seja, as falhas de alto impacto foram percebidas de forma diferente das falhas de baixo impacto pelos participantes deste estudo.

Tabela 14 – Cenários sem diferença estatística de acordo com o teste de Kruskal-Wallis (questionário online)

Dimensão	Cenário (p-valor < 0,005)	Graus de Liberdade
Satisfação	C3 = C4 / Sig = 0,613	GL(3) = 277,43
Preocupação	C1 = C2 / Sig = 0,142	
	C3 = C4 / Sig = 0,605	GL(3) = 255,72
Amparo	C1 = C2 / Sig = 0,579	
	C3 = C4 / Sig = 0,260	GL(3) = 263,22
Tolerância	C3 = C4 / Sig = 0,396	GL(3) = 321,37
Produtividade	C3 = C4 / Sig = 0,813	GL(3) = 312,93
Estresse	C1 = C2 / Sig = 0,010	
	C3 = C4 / Sig = 0,355	GL(3) = 264,30
Pontualidade	C1 = C2 / Sig = 0,097	
	C3 = C4 / Sig = 0,645	GL(3) = 229,20
Paciência	C1 = C2 / Sig = 0,030	
	C3 = C4 / Sig = 0,344	GL(3) = 125,43
Tranquilidade	C3 = C4 / Sig = 0,017	GL(3) = 342,60
Capacidade	C1 = C2 / Sig = 0,019	
	C3 = C4 / Sig = 0,193	GL(3) = 203,31

A Figura 16 mostra as médias dos escores das dimensões por cenário com um intervalo de confiança de 95%. As dimensões Preocupação e Amparo não apresentaram diferença significativa nos cenários 1 e 2, diferentemente das outras dimensões. A dimensão Preocupação foi a única que apresentou escore médio maior no cenário 2 que no cenário 1. Nos cenários 3 e 4 houve diferença significativa apenas na dimensão Tranquilidade. De modo geral, pode-se concluir que o nível de impacto dos cenários influenciou negativamente na percepção dos usuários.

A Figura 17 mostra as medianas dos escores das dimensões por cenário, também com um intervalo de confiança de 95%. Pode-se observar que a diminuição nos escores entre os cenários 1 e 2 é maior que entre os cenários 3 e 4. As dimensões Preocupação, Amparo e Pontualidade apresentaram a mesma variação nos cenários 1 e 2. Já no caso dos cenários 3 e 4, somente as dimensões Amparo, Tranquilidade e Capacidade não apresentaram mesma variação. Observa-se que a mediana do escore de paciência reduziu de 3,00 para 1,00 entre os cenários 1 e 2, e estabilizou em 1,00 até o último cenário. Essa redução dessa maneira não ocorreu nos escores das outras dimensões. Analisando o gráfico, conclui-se que a ocorrência de uma falha em cenários com alto impacto afeta a QoE dos usuários ao ponto de metade dos participantes terem avaliado todas as dimensões com escore mínimo, exceto as dimensões Amparo (cenário 3), Tranquilidade (cenários 3 e 4) e Capacidade (cenários 3 e 4). Portanto,

para melhorar a QoE dos usuários em relação à confiabilidade do *software*, deve-se elaborar medidas para que as falhas do sistema não sejam percebidas de maneira tão negativa.

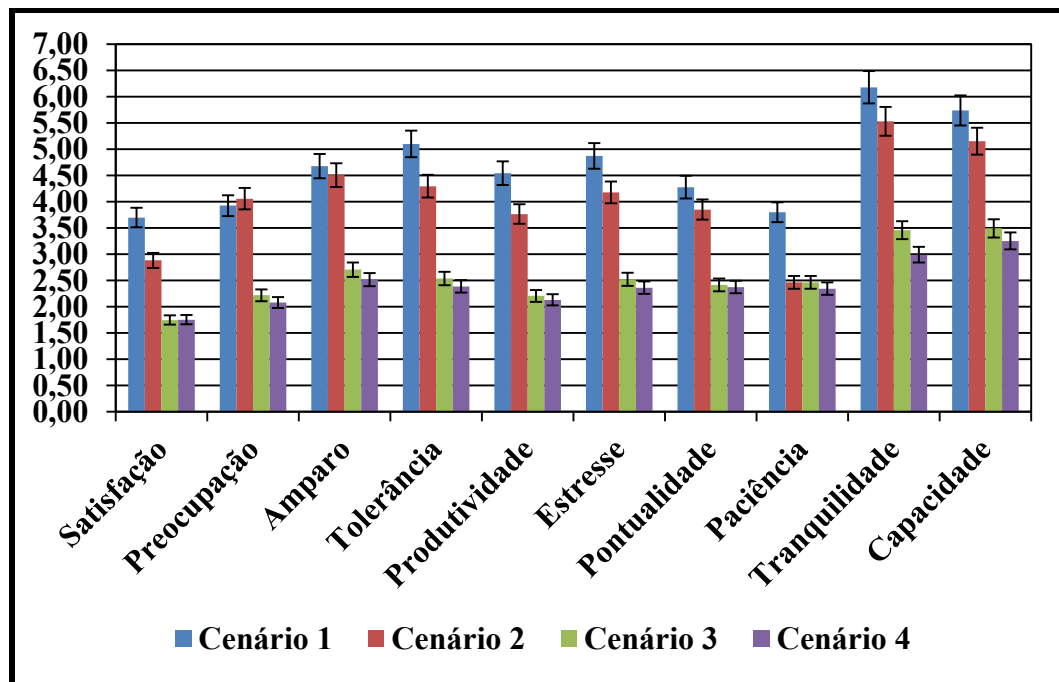


Figura 15 - Médias dos escores das dimensões por cenário (questionário *online*)

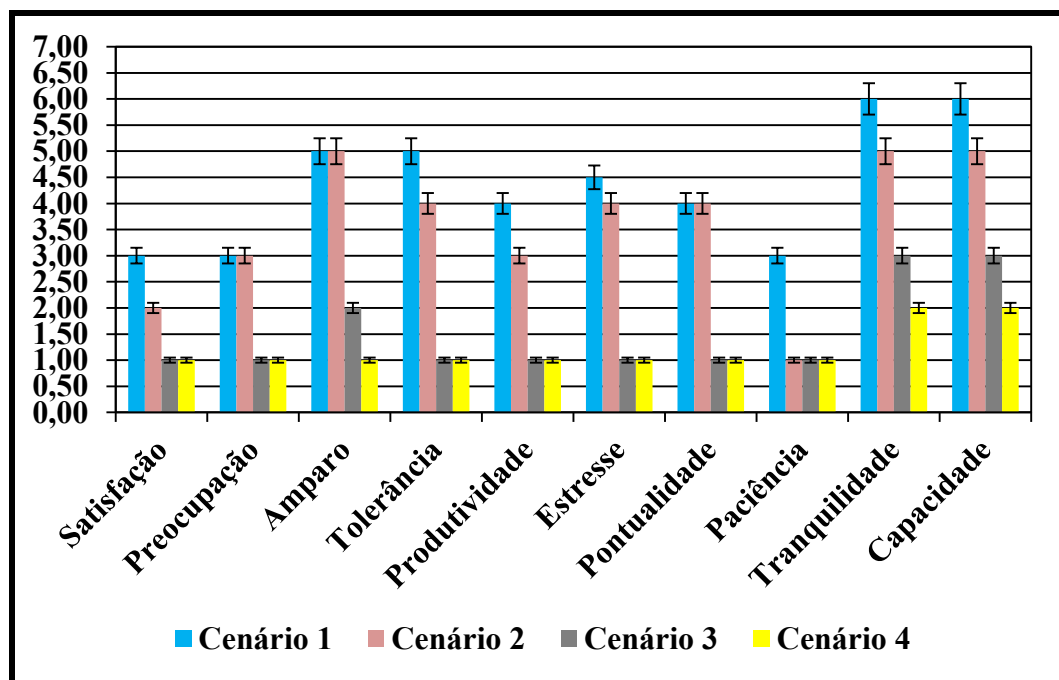


Figura 16 - Medianas dos escores das dimensões por cenário (questionário *online*)

4.2 Questionário presencial

4.2.1 Etapa 1

A amostra do questionário presencial foi composta por 86 participantes, dos quais 61 são do gênero masculino e 25 do gênero feminino (Tabela 15).

Tabela 15 - Percentual de participantes por gênero (questionário presencial)

Gênero	Quantidade	Percentual
Feminino	25	29,00%
Masculino	61	71,00%

A Tabela 16 mostra a distribuição de faixa etária dos participantes do questionário presencial. A faixa etária 18 a 24 anos concentrou a maioria dos participantes da amostra, com 60 pessoas. A segunda maior faixa etária foi 25 a 44 anos, com 24 participantes. A faixa etária 45 a 65 anos concentrou 2 participantes e por fim, não houve participantes com faixa etária acima de 65 anos.

Tabela 16 - Percentual de participantes por faixa etária (questionário presencial)

Faixa Etária	Quantidade	Percentual
18 a 24 anos	60	70,00%
25 a 44 anos	24	28,00%
45 a 65 anos	02	02,00%

A Figura 18 mostra a distribuição dos participantes por grau de escolaridade. Observa-se maior concentração em Superior incompleto, com 61 participantes. Participaram 12 pessoas com Pós-graduação e Superior completo. Uma pessoa participou sendo da categoria Ensino médio completo. Por fim, não houve participantes das categorias Ensino médio incompleto, Ensino fundamental completo e Ensino fundamental incompleto.

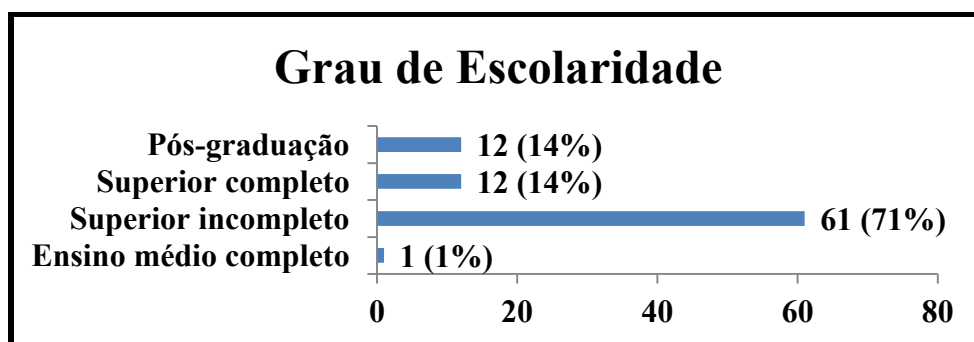


Figura 17 - Percentual de participantes por grau de escolaridade (questionário presencial)

A Tabela 17 mostra os percentuais dos participantes por área de atividade. A categoria com maior predominância foi Computação com 71 participantes, compondo 83% da amostra total. A segunda categoria mais predominante foi Engenharia com 12%. As categorias Ciências Contábeis, Educação e Ciências da Saúde compuseram juntas 5% da amostra total.

Tabela 17 – Percentual de participantes por atividade (questionário presencial)

Categoria	Frequência	Porcentagem
Computação	71	83,00%
Engenharia	10	12,00%
Ciências Contábeis	02	02,00%
Educação	02	02,00%
Ciências da Saúde	01	01,00%

4.2.2 Etapa 2

A Tabela 18 mostra a confiabilidade calculada com base na abordagem clássica por cenário e as respectivas estatísticas descritivas da confiabilidade percebida pelos participantes. Observa-se que a média e a mediana da confiabilidade percebida foi maior do que a confiabilidade calculada no cenário 1 e menor nos outros cenários, assim como no questionário *online*. Nota-se que a média e a mediana da confiabilidade percebida nos cenários de baixo impacto (1 e 2) atingiu o valor máximo de 56,83%, enquanto que nos cenários de alto impacto (3 e 4) esse valor não chegou a 15%. Esse resultado é semelhante ao observado no questionário *online*, porém a diferença entre os cenários de baixo e alto impacto no questionário presencial foi ainda maior, reforçando a evidência de que a consequência possui maior influência sobre a percepção dos usuários que o MTBF. Analisando a subtração entre a confiabilidade estimada pelo modelo de Poisson e a confiabilidade percebida média, bem como a subtração em relação à confiabilidade percebida mediana, observa-se que a diferença maior entre os valores ocorreu nos cenários 2 (MTBF longo, baixo impacto) e 4 (MTBF longo, alto impacto). Pode-se concluir que o MTBF longo possui maior influência na percepção dos usuários que o MTBF curto, assim como no questionário *online*. A moda da confiabilidade percebida foi 50% nos cenários 1 e 2 e 0% nos cenários 3 e 4. Os cenários 3 e 4 apresentaram o mesmo tipo de consequência (o usuário perde o documento que estava editando), porém com MTBF e mensagem de falha diferente. O que pode ter contribuído para os cenários 3 e 4 ter sido percebidos pela maioria dos participantes como sem confiabilidade alguma é a BSOD gerada pela falha relatada neste cenário, que pode provocar no usuário a impressão de estar vivenciando uma falha mais grave que as outras vivenciadas anteriormente. A confiabilidade percebida atingiu um valor máximo de 100% no cenário 1, 85% no cenário 2, 75% no cenário 3 e 70% no cenário 4. O valor mínimo de confiabilidade percebida foi de 4% no cenário 1 e 0% nos outros cenários. Esse resultado aparenta ser mais realista em relação àquele observado no questionário *online*, em que a confiabilidade percebida máxima foi 100% em todos os cenários. Isso é uma evidência de que os usuários percebem menos confiabilidade em um sistema à medida que eles vivenciam falhas em sequência. Por fim, pode-se observar que, embora o resultado do desvio padrão decresceu do

cenário 1 para o cenário 4, seu valor manteve-se abaixo de 22% em todos os cenários, indicando que as observações do questionário presencial estão menos dispersas em relação à média que as observações do questionário *online*. Por fim, pode-se observar maior diferença entre o resultado do cálculo de confiabilidade clássica e a confiabilidade percebida pelos participantes no questionário presencial em comparação com o questionário *online*.

Tabela 17 - Estatísticas descritivas da confiabilidade percebida pelos participantes (questionário presencial)

Cenário	1	2	3	4
Impacto	Baixo		Alto	
MTBF	Curto	Longo	Curto	Longo
$R_{(t)}$	36,80%	81,00%	36,80%	93,50%
$\bar{R}_{per}(t)$	56,83%	40,53%	14,13%	09,81%
$R_{(t)} - \bar{R}_{per}(t)$	-20,03%	40,47%	22,67%	83,69%
$\tilde{R}_{per}(t)$	53,50%	40,00%	10,00%	01,50%
$R_{(t)} - \tilde{R}_{per}(t)$	-16,70%	41,00%	26,80%	92,00%
$\check{R}_{per}(t)$	50,00%	50,00%	00,00%	00,00%
$R_{per}^{max}(t)$	100,00%	85,00%	75,00%	70,00%
$R_{per}^{min}(t)$	04,00%	00,00%	00,00%	00,00%
$R_{per}^{dp}(t)$	21,02%	21,48%	17,00%	15,65%

4.2.2.1 Análise Fatorial dos Dados do Questionário *Presencial*

O Teste de *KMO* para a amostra do questionário presencial resultou em 0,859 e o Teste de Esfericidade de Bartlett em $(GL(45) = 729,76; p\text{-valor} < 0,005)$, indicando boa adequação dos dados à análise fatorial devido à forte correlação entre as dimensões (PEREIRA, 2004). A análise fatorial indicou a retenção de apenas um fator que explica 62,39% da variância total, composto pelas dez dimensões analisadas (Tabela 19), indicando que as dez dimensões possuem relação entre si e estão mensurando um mesmo conceito. Portanto, as dez dimensões podem ser utilizadas para mensurar a confiabilidade percebida pelos usuários.

A Figura 19 mostra o gráfico de escarpa para os fatores extraídos. Observando o gráfico, nota-se que apenas o primeiro fator extraído apresentou autovalor maior que 1 (critério da raiz latente), o que indica reter apenas esse fator. Além disso, o ponto de inflexão no gráfico é o segundo fator extraído. De acordo com o critério do gráfico de escarpa, apenas os fatores extraídos que estão acima do ponto de inflexão devem ser retidos. Portanto, o

critério do gráfico de escurpa também indica a retenção de apenas um fator, assim como os outros critérios analisados (porcentagem de variância e critério da raiz latente).

Tabela 18 - Porcentagem de variância explicada por fator (questionário presencial)

Fator	Porcentagem de variância	Porcentagem acumulada
1	62,39	62,39
2	09,12	71,51
3	08,08	79,59
4	06,67	86,26
5	05,15	91,41
6	03,41	94,82
7	01,75	96,57
8	01,67	98,24
9	00,98	99,22
10	00,78	100,00

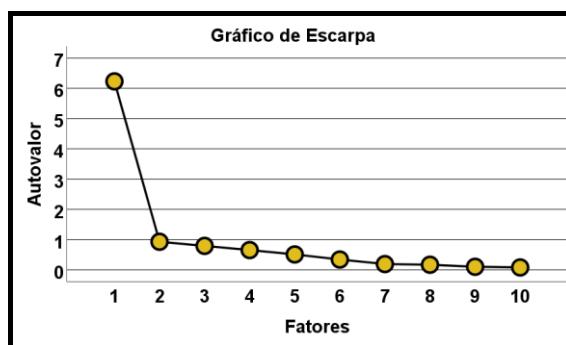


Figura 18 - Gráfico de escurpa para os fatores extraídos (questionário presencial)

Com a quantidade de fatores extraídos, é necessário verificar quais as dimensões são mais relevantes na composição do fator analisando o valor de comunalidade de cada dimensão. A Tabela 20 mostra os valores de comunalidade para as dez dimensões que compuseram o fator extraído na análise fatorial. Todas as dimensões apresentaram valor de comunalidade superior a 0,3, o que indica que tais dimensões estão relacionadas com o fator extraído. Para ajustar os modelos de predição de confiabilidade percebida, serão consideradas as três dimensões com maior valor de comunalidade e, conseqüentemente, as dimensões mais relevantes na composição do fator. Analisando os valores da Tabela 19, observa-se que as três dimensões com maior comunalidade são: Estresse, Pontualidade e Paciência. Essas dimensões também foram indicadas pela análise fatorial dos dados do questionário *online*, porém em ordem de relevância diferente. A análise fatorial realizada com os dados do questionário *online* também indicou a retenção de um fator composto pelas dez dimensões avaliadas.

Tabela 19 - Valor de comunalidade por dimensão (questionário presencial)

Dimensão	Comunalidade
Satisfação	0,52
Preocupação	0,70
Amparo	0,78
Tolerância	0,83
Produtividade	0,82
Estresse	0,86
Pontualidade	0,85
Paciência	0,85
Tranquilidade	0,83
Capacidade	0,73

4.2.2.2 Ajuste dos modelos de regressão linear múltipla com os dados do questionário presencial

No ajuste do terceiro modelo de predição de confiabilidade percebida, serão consideradas três dimensões com maior valor de comunalidade e, conseqüentemente, as dimensões mais relevantes na composição do fator. Analisando os valores da Tabela 20, observa-se que as três dimensões com maior comunalidade são: Estresse, Pontualidade e Paciência. O mesmo procedimento adotado para ajustar o Modelo 1 foi adotado para ajustar o Modelo 3. Sendo assim, a confiabilidade percebida média por cada participante foi utilizada como **variável dependente** e os escores médios das dimensões Estresse, Pontualidade e Paciência foram utilizados como **variáveis independentes** para ajustar o modelo 3 de confiabilidade predita mostrado em (18):

$$R_{Pred_3}(t) = 23,846 - 0,674 * a1 + 1,152 * b1 + 1,524 * c1 \quad (18)$$

no qual $a1$ é o escore de Estresse, $b1$ é o escore de Pontualidade, $c1$ é o escore de Paciência e t é o momento de ocorrência da falha. A Tabela 21 mostra o ajuste do modelo descrito em (18) considerando uma, duas e três dimensões. Pode-se observar que houve melhora significativa do modelo ao considerar duas dimensões (Estresse e Pontualidade) ao invés de apenas uma (Estresse) analisando o valor de R^2 e R^2 ajustado (R^2 ajust). Entretanto, o modelo não é melhorado ao adicionar a dimensão Paciência, pois o valor de R^2 ajustado é reduzido em 0,006. Ao analisar o nível de significância das três dimensões no Teste t (Estresse 0,701; p-valor < 0,005), (Pontualidade 0,493; p-valor < 0,005), (Paciência 0,477; p-valor < 0,005) observa-se que nenhuma dessas três dimensões é adequada para montar o modelo. Entretanto, o Modelo 3 será mantido com essas dimensões para avaliar se o seu resultado se aproxima mais da percepção dos usuários que o resultado da confiabilidade estimada pelo método clássico. Por fim, o resultado do Teste de DW indicou a independência dos resíduos e o valor de VIF e Tol mostrou a ausência de multicolinearidade entre as dimensões em todos os casos, indicando que o modelo é apropriado.

Tabela 20 – Resumo estatístico do Modelo 3

Modelo 3								
Ajuste	Coef.	R^2	R^2 ajust	DW	Teste t	Sig. Teste t	VIF	Tol
(Constante)	25,092	0,026	0,015	-	6,604	0,000	1,000	1,000
Estresse	1,382				1,501	0,137		
(Constante)	25,543	0,049	0,026	-	5,987	0,000	-	-
Estresse	0,205				0,166	0,869	1,827	0,547
Pontualidade	1,876				1,414	0,161	1,827	0,547
(Constante)	23,846	0,055	0,020	2,098	6,012	0,000	-	-
Estresse	-0,674				-0,386	0,701	3,627	0,276
Pontualidade	1,152				0,689	0,493	2,887	0,346
Paciência	1,524				0,714	0,477	5,720	0,175

Assim como o Modelo 2, o Modelo 4 também foi elaborado em duas etapas. Primeiramente, foi ajustado um modelo considerando as dez dimensões. Então, foram selecionadas as dimensões mais relevantes apontadas pelo Teste t ($p < 0,005$). A Tabela 22 mostra o nível de significância no Teste t das dez dimensões. Os resultados apontaram apenas a dimensão Satisfação como relevante.

Tabela 21 - Nível de significância no Teste t das dez dimensões utilizadas na primeira etapa de ajuste do Modelo 4 (questionário presencial)

Dimensão	Significância
Satisfação	0,000
Preocupação	0,345
Amparo	0,797
Tolerância	0,409
Produtividade	0,450
Estresse	0,040
Pontualidade	0,511
Paciência	0,285
Tranquilidade	0,338
Capacidade	0,174

Portanto, o quarto modelo ajustado é um modelo de regressão linear simples, por considerar apenas uma dimensão em seu ajuste. Tal modelo pode ser visto em (19):

$$R_{Pred_4}(t) = 11,008 + 7,184 * s \quad (19)$$

no qual s é o escore de Satisfação e t é o momento de ocorrência da falha.

A Tabela 23 mostra o resumo estatístico do Modelo 4. O resultado do Teste de Durbin-Watson indicou a independência dos resíduos. Os valores de *VIF* e *Tol* são iguais a 1 porque apenas uma dimensão foi considerada no ajuste e, portanto, não existe a ocorrência de multicolinearidade.

Tabela 22 - Resumo estatístico do Modelo 4

Modelo 4								
Ajuste	Coef.	R^2	R^2 ajust	DW	Teste t	Sig. Teste t	<i>VIF</i>	<i>Tol</i>
(Constante)	11,008	0,293	0,285	2,232	3,130	0,002	1,000	1,000
Satisfação	7,184				5,902	0,000		

A Tabela 24 mostra uma comparação entre a confiabilidade estimada por método clássico ($R_{(t)}$), a média ($\bar{R}_{Per}(t)$) e mediana ($\tilde{R}_{Per}(t)$) da confiabilidade percebida pelos participantes e a confiabilidade predita pelo Modelo 3 ($R_{Pred_3}(t)$) e pelo Modelo 4 ($R_{Pred_4}(t)$) nos quatro cenários do questionário *online*. Analisando as diferenças entre os cálculos, nota-se que o resultado calculado com ambos os modelos aproxima-se mais da confiabilidade percebida pelos usuários que o resultado obtido pelo método clássico, exceto no cenário 1, no qual o método clássico se aproximou mais da percepção do usuário que o Modelo 3. O mesmo ocorre ao analisar ($\tilde{R}_{Per}(t)$), porém observa-se que nesse caso a precisão dos modelos é reduzida em relação à comparação feita utilizando a ($\bar{R}_{Per}(t)$). Nota-se que o resultado obtido com os dois modelos é menor que a confiabilidade percebida para os cenários 1 e 2, definidos com baixo impacto, porém para os cenários 3 e 4, definidos com alto impacto, o resultado obtido com ambos os modelos é maior que o valor de confiabilidade percebida. Com todas essas análises, reforça-se a conclusão de que ao calcular a confiabilidade de um sistema considerando fatores de QoS e QoE (análise quanti-qualitativa), o resultado obtido se aproxima melhor da confiabilidade percebida pelos usuários (QoE - análise qualitativa) que apenas considerando fatores de QoS (análise qualitativa). Além disso, observa-se que considerar a dimensão Satisfação na elaboração de modelos de predição de confiabilidade produz modelos mais precisos que considerar as dimensões Estresse, Pontualidade e Paciência. Portanto, as dimensões apontadas pela análise fatorial como as mais relacionadas com o fator não foram as dimensões que compuseram o modelo de predição mais preciso, assim como foi observado no questionário *online*.

Em comparação com os resultados do questionário *online*, pode-se observar que a dimensão Tolerância não foi apontada pelo Teste t como relevante para montar o Modelo 4. Contudo, a dimensão Satisfação apareceu como relevante nas duas análises, o que indica que considerar essa dimensão é importante para produzir modelos de predição de confiabilidade que se aproximem da percepção dos usuários. Além disso, os Modelos 2 e 4, que foram montados considerando a dimensão Satisfação, se aproximaram mais da confiabilidade percebida pelos usuários que os Modelos 1 e 3, que não foram montados considerando essa dimensão. Esse resultado reforça a importância de considerar a Satisfação para avaliar a QoE dos usuários.

Tabela 23 - Comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelos Modelos 3 e 4 (questionário presencial)

Cenário	1	2	3	4
Impacto	Baixo		Alto	
MTBF	Curto	Longo	Curto	Longo
$R(t)$	36,80%	81,00%	36,80%	93,50%
$\bar{R}_{Per}(t)$	56,83%	40,53%	14,13%	09,81%
$\tilde{R}_{Per}(t)$	53,50%	40,00%	10,00%	01,50%
$R_{Pred_3}(t)$	34,63%	31,95%	27,47%	27,26%
$R_{Pred_4}(t)$	47,01%	33,48%	21,12%	19,70%
$\bar{R}_{Per}(t) - R(t)$	20,03%	-40,47%	-22,67%	-83,69%
$\bar{R}_{Per}(t) - R_{Pred_3}(t)$	22,19%	08,59%	-13,34%	-17,45%
$\bar{R}_{Per}(t) - R_{Pred_4}(t)$	09,81%	07,06%	-06,99%	-09,88%
$\tilde{R}_{Per}(t) - R(t)$	16,70%	-41,00%	-26,80%	-92,00%
$\tilde{R}_{Per}(t) - R_{Pred_3}(t)$	18,87%	08,05%	-17,47%	-25,76%
$\tilde{R}_{Per}(t) - R_{Pred_4}(t)$	06,49%	06,52%	-11,12%	-18,20%

A Figura 20 mostra uma comparação gráfica entre os cálculos de confiabilidade com um intervalo de confiança de 95%. Observa-se o mesmo padrão dos resultados do questionário *online*, porém os participantes do questionário presencial perceberam um valor ainda menor de confiabilidade nos cenários definidos como de alto impacto. Pode-se concluir que vivenciar as falhas descritas nos cenários na prática afetou mais a experiência dos usuários que apenas ler esses cenários, como foi no questionário *online*.

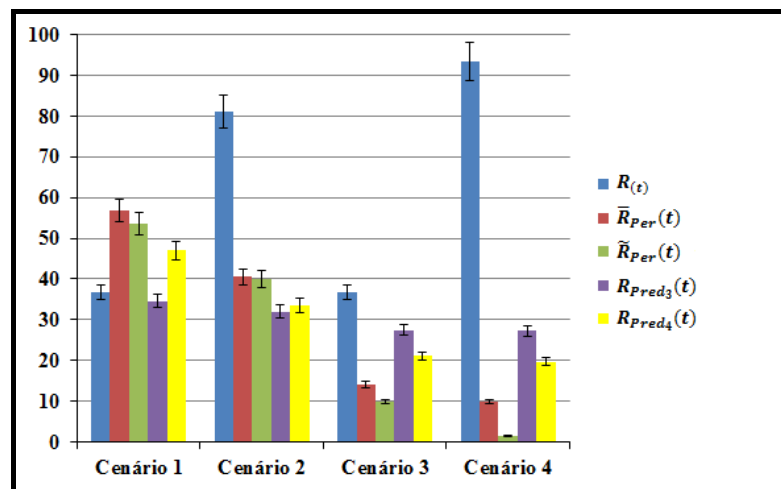


Figura 19 - Intervalo de confiança de 95% para os resultados dos cálculos de confiabilidade (questionário presencial)

4.2.2.3 Comparação entre as dimensões dos cenários do questionário presencial

Para examinar as diferenças entre os cenários por dimensão foi utilizado o Teste de Kruskal-Wallis. A Tabela 25 mostra os resultados do teste para cada par de cenários sem diferença estatística de acordo com o teste. Portanto, caso uma combinação não estiver na tabela significa que houve diferença estatística na percepção de um cenário em comparação com outro. Por exemplo, na dimensão Estresse a Tabela 25 não mostra o par de cenários 2 e 3 (C2 – C3), portanto, houve diferença significativa da dimensão Estresse entre os cenários 2 e 3.

Não foram encontradas diferenças significativas entre o cenário 3 e o cenário 4 para todas as dimensões, assim como no questionário *online*. Isso reforça a conclusão de que os cenários 3 e 4 foram vivenciados pelos usuários com intensidade semelhante. As combinações entre cenários de baixo e alto apresentaram diferença significativa em todas as dez dimensões consideradas neste estudo. Esse resultado também foi observado no questionário *online* e através dele reforça-se a conclusão de que os cenários de impacto diferente foram vivenciados com intensidade discrepante, o que ressalta a relevância que a consequência de uma falha têm na QoE dos usuários. Os cenários 1 e 2 não apresentaram diferença significativa apenas nas dimensões Amparo e Pontualidade, o que indica que falhas sem grandes consequências e de rápida recuperação não fazem com que os usuários se sintam desamparados e atrasados para realizar as suas tarefas.

Tabela 24 – Comparação entre os cenários por dimensão (questionário presencial)

Dimensão	Cenário (p-valor < 0,005)	Graus de Liberdade
Satisfação	C3 = C4 / Sig = 0,416	GL(3) = 180,49
Preocupação	C3 = C4 / Sig = 0,043	GL(3) = 157,40
Amparo	C1 = C2 / Sig = 0,009 C3 = C4 / Sig = 0,531	GL(3) = 150,72
Tolerância	C3 = C4 / Sig = 0,897	GL(3) = 172,14
Produtividade	C3 = C4 / Sig = 0,954	GL(3) = 187,47
Estresse	C3 = C4 / Sig = 0,179	GL(3) = 153,34
Pontualidade	C1 = C2 / Sig = 0,016 C3 = C4 / Sig = 0,834	GL(3) = 156,10
Paciência	C3 = C4 / Sig = 0,694	GL(3) = 129,98
Tranquilidade	C3 = C4 / Sig = 0,162	GL(3) = 147,33
Capacidade	C3 = C4 / Sig = 0,246	GL(3) = 139,50

A Figura 21 mostra a média dos escores das dimensões por cenário do questionário presencial com um intervalo de confiança de 95%. Todas as dimensões apresentaram diferença significativa nos cenários 1 e 2, enquanto que no questionário *online* as dimensões

Preocupação e Amparo não apresentaram diferença significativa nesses cenários. Observa-se diferença significativa entre os cenários 2 e 3, assim como nos resultados do questionário *online*. Entre os cenários 3 e 4 houve diferença significativa nas dimensões: Satisfação, Preocupação, Estresse, Tranquilidade e Capacidade. No questionário *online*, essa diferença ocorreu apenas na dimensão Tranquilidade. Os resultados obtidos com a análise dos dados na Figura 21 evidenciam ainda mais o quanto o impacto dos cenários influenciou na percepção dos usuários, pois nesse caso a diferença de percepção entre os cenários de baixo e alto impacto foi mais evidente que a diferença observada no questionário *online* (ver Figura 16).

A Figura 22 mostra a mediana dos escores das dimensões por cenário do questionário presencial com um intervalo de confiança de 95%. Assim como na análise das médias (Figura 21), todas as dimensões apresentaram diferença significativa entre os cenários 1,2 e 3. No entanto, entre os cenários 3 e 4, houve diferença significativa apenas nas dimensões Tranquilidade e Capacidade, pois as medianas das outras dimensões desses cenários foram iguais a 1. Comparando os resultados mostrados na Figura 22 (questionário presencial) com os resultados mostrados na Figura 17 (questionário *online*), pode-se observar que os valores de todas as dimensões no questionário presencial foram maiores que no questionário *online* para o cenário 1. A recuperação rápida do sistema nesse cenário, bem como a ausência de perda de arquivos podem ter contribuído para o menor impacto na QoE dos usuários nesse cenário.

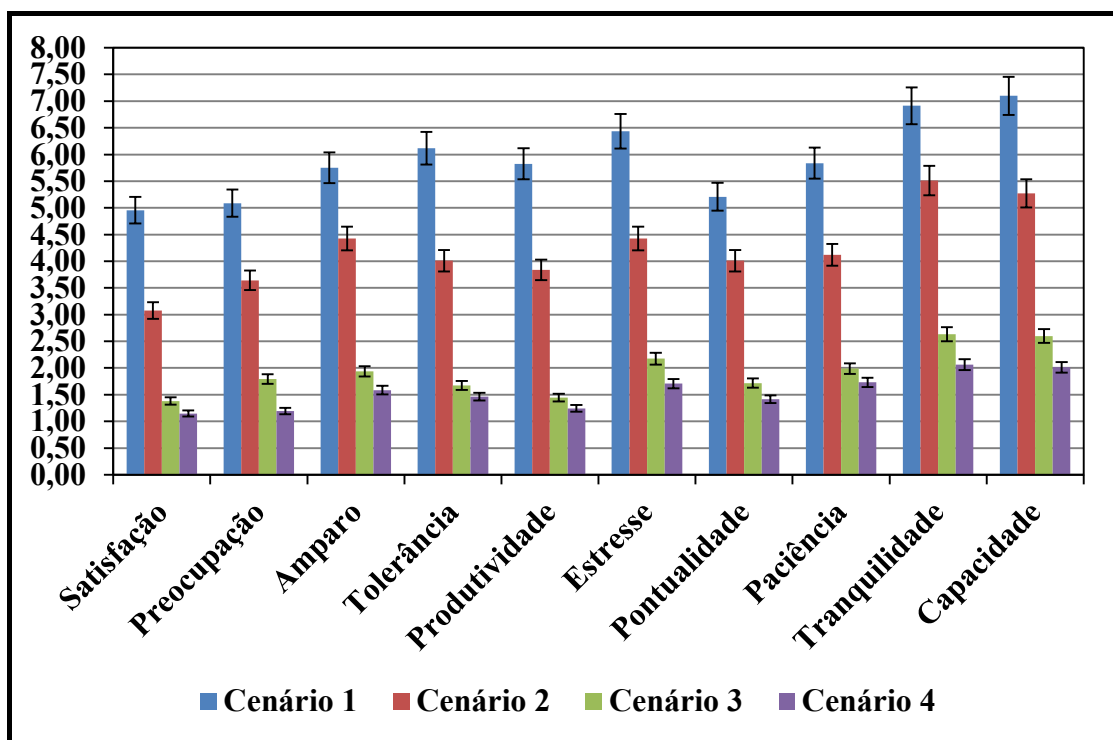


Figura 20 - Médias dos escores das dimensões por cenário (questionário presencial)

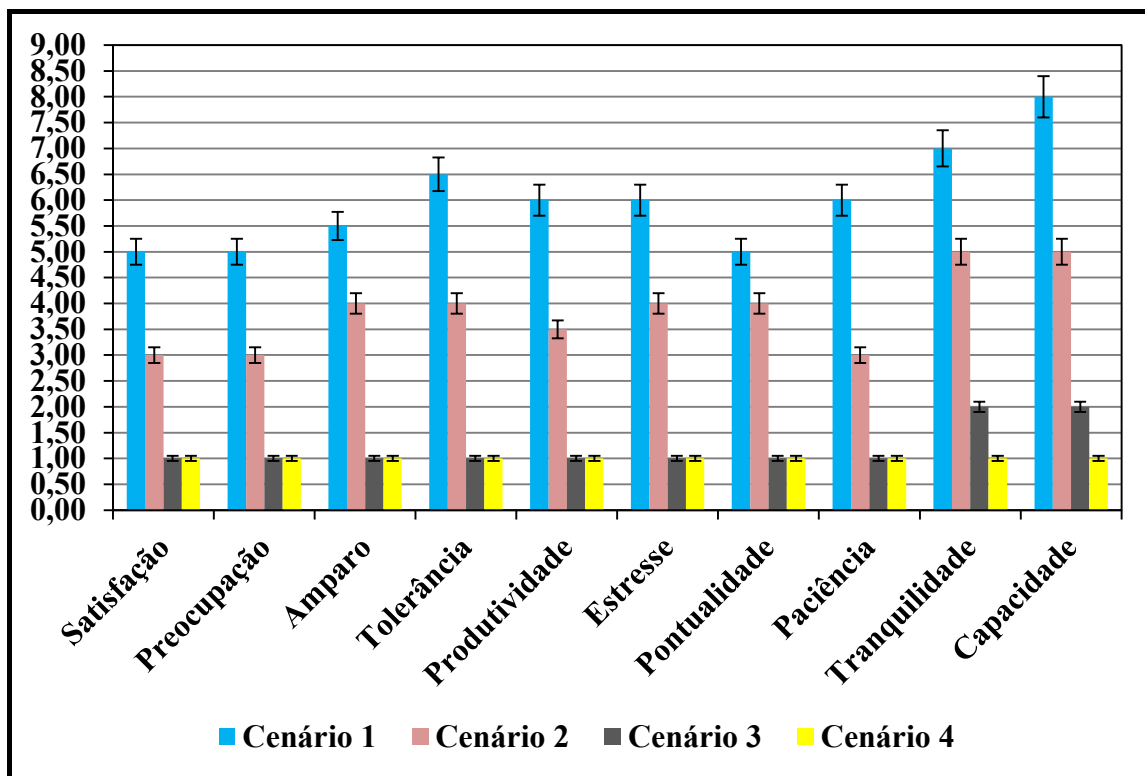


Figura 21 - Medianas dos escores das dimensões por cenário (questionário presencial)

4.3 Validação com amostra de modelagem e amostra de teste

Com o objetivo de validar a precisão da técnica de modelagem do cálculo de confiabilidade de *software* considerando a percepção dos usuários, dividiu-se a amostra coletada com o questionário *online* em duas partes: amostra de modelagem e amostra de teste. A amostra de modelagem é composta pelos dados dos participantes de 1 a 200, enquanto que a amostra de teste é composta pelos participantes de 201 a 400. Os dados da amostra de modelagem foram utilizados para ajustar o modelo de cálculo de confiabilidade enquanto que os dados da amostra de teste foram usados como entrada de dados no modelo ajustado. Para o ajuste do modelo foi utilizada apenas a variável Satisfação, por já ter sido observado anteriormente que ela é variável mais precisa dentre as observadas neste estudo. Além disso, a validação foi realizada apenas com os dados coletados no questionário *online*, pois a amostra de dados coletada com o questionário presencial ficaria muito pequena para essa análise. Segue abaixo em (20) o modelo ajustado (Modelo 5):

$$R_{Pred_5}(t) = 13,749 + 7,771 * s \quad (20)$$

A Tabela 26 mostra o resumo estatístico do Modelo 5. O resultado do Teste de Durbin-Watson indicou a independência dos resíduos. Os valores de *VIF* e *Tol* são iguais a 1 porque apenas uma dimensão foi considerada no ajuste, mostrando que não existe a ocorrência de multicolinearidade. Portanto, o Modelo 5 também é adequado para uso.

Tabela 26 - Resumo estatístico do Modelo 5 feito com dados do questionário *online* usando a variável Satisfação

Modelo 5								
Ajuste	Coef.	R^2	R^2 ajust	DW	Teste t	Sig. Teste t	VIF	Tol
(Constante)	13,749	0,386	0,383	1,958	6,535	0,000	1,000	1,000
Satisfação	7,771				11,161	0,000		

A Tabela 27 mostra a comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelo Modelo 5. Nota-se que com exceção do cenário 1, em todos os outros casos a confiabilidade predita pelo Modelo 5 aproximou-se mais da confiabilidade percebida pelos usuários que o cálculo de confiabilidade pelo modelo clássico. A redução da amostra nessa análise, bem como o tipo de impacto provocado pela falha descrita no cenário 1 podem ter colaborado para isso, já que no cenário 3 o valor de confiabilidade calculado pelo modelo clássico é o mesmo do cenário 1. Apesar disso, o valor de confiabilidade predito pelo Modelo 5 no cenário 3 aproximou-se mais da confiabilidade percebida pelos usuários que o modelo clássico. Esses resultados reforçam ainda mais a imprecisão do modelo de confiabilidade clássico em relação à percepção de falhas de *software* pelos usuários, bem como a importância de considerar o impacto que tais falhas provocam nos usuários durante o uso dos sistemas.

Tabela 27 - Comparação por cenário entre os resultados do cálculo de confiabilidade clássico, a média e mediana da confiabilidade percebida e a confiabilidade predita pelo Modelo 5 (questionário *online*)

Impacto	Baixo		Alto	
	1	2	3	4
Cenário				
MTBF	Curto	Longo	Curto	Longo
$R_{(t)}$	36,80%	81,00%	36,80%	93,50%
$\bar{R}_{Per}(t)$	51,06%	45,03%	24,55%	21,09%
$\tilde{R}_{Per}(t)$	50,00%	45,00%	20,00%	15,00%
$R_{Pred_5}(t)$	40,29%	34,69%	27,27%	27,31%
$\bar{R}_{Per}(t) - R_{(t)}$	14,26%	-35,97%	-12,25%	-72,41%
$\bar{R}_{Per}(t) - R_{Pred_5}(t)$	10,77%	10,34%	-02,72%	-06,22%
$\tilde{R}_{Per}(t) - R_{(t)}$	13,20%	-36,00%	-16,80%	-78,50%
$\tilde{R}_{Per}(t) - R_{Pred_5}(t)$	09,71%	10,31%	-07,27%	-12,31%

Conclusão

5.1 Síntese dos Resultados

Este trabalho teve como objetivo principal realizar uma análise da confiabilidade de *software* estimada com fatores quantitativos e qualitativos relacionados à QoE do usuário. Essa análise foi realizada utilizando falhas do SO Microsoft Windows 7 como um estudo de caso. Para realizar este estudo, foi elaborado um questionário contendo quatro cenários que descreviam situações hipotéticas, mas realísticas, em que ocorriam falhas do SO durante o uso do computador. Os cenários 1 e 2 do questionário foram definidos com baixo impacto, ou seja, são aqueles em que a falha do SO não impede as atividades de trabalho do usuário, com rápida recuperação da falha. Já os cenários 3 e 4 foram definidos com alto impacto, sendo os cenários em que a falha do SO prejudicou a tarefa que o usuário estava fazendo no computador (consequência da falha). Além disso, cada cenário apresentou um valor curto ou longo de MTBF, para que fosse possível avaliar como a frequência de falhas afetou a QoE do usuário. Por fim, a confiabilidade relatada em cada cenário foi calculada pelo modelo clássico de Poisson.

Para cada cenário, os participantes respondiam a duas perguntas. Na primeira pergunta os participantes responderam qual o seu grau de confiança no SO descrito no cenário apresentado em um intervalo de 0 (sem confiança alguma) até 100 (confiança total). Na segunda pergunta foram medidas dez dimensões emocionais (Satisfação, Preocupação, Amparo, Tolerância, Produtividade, Estresse, Pontualidade, Paciência, Tranquilidade e Capacidade) através de uma adaptação de escalas de diferencial semântico. O questionário foi aplicado de forma *online* e presencial, sendo que os participantes presentes na amostra coletada com o questionário *online* (400 participantes) não compuseram a amostra coletada com o questionário presencial (86 participantes).

Ao analisar o intervalo de confiança das médias e medianas das respostas da primeira pergunta do questionário, observa-se que a confiabilidade percebida pelos usuários é significativamente diferente da confiabilidade estimada pelo método clássico, pois o sistema foi percebido com maior confiabilidade que a confiabilidade calculada pelo método clássico no cenário 1 e com menor confiabilidade nos outros cenários. Portanto, conclui-se que o resultado do cálculo da confiabilidade de um *software* considerando apenas aspectos de QoS não é compatível com a QoE dos usuários em relação à confiabilidade do *software* usado.

Ao explorar as dimensões presentes na segunda pergunta do questionário, nota-se que a média e a mediana dos escores das dimensões nos cenários com MTBF longo foram significativamente mais baixos em comparação aos escores dos cenários de mesmo impacto, mas com MTBF curto. Esse fenômeno também foi observado ao analisar a confiabilidade

percebida nos cenários. O cenário 1 (MTBF curto) foi percebido com maior confiabilidade que o cenário 2 (MTBF longo), apesar de ambos os cenários serem definidos com baixo impacto. O mesmo ocorre entre os cenários 3 e 4, em que o cenário 3 (MTBF curto) foi percebido com maior confiabilidade que o cenário 4 (MTBF longo). Portanto, conclui-se que o MTBF é um fator de influência de QoE, afetando negativamente a percepção dos usuários em relação à confiabilidade de *software*, sendo que um valor de MTBF longo afeta mais que um valor de MTBF curto. Note que ao calcular a confiabilidade de um sistema pelo método clássico, observa-se que o sistema apresenta maior confiabilidade com um MTBF longo em comparação à confiabilidade calculada com um MTBF curto.

Ao comparar a confiabilidade percebida nos cenários de baixo e alto impacto, nota-se que os cenários com alto impacto foram percebidos significativamente com menor confiabilidade que os cenários de baixo impacto, evidenciando que a consequência provocada pela ocorrência de uma falha é um fator que afeta negativamente a QoE dos usuários em relação à confiabilidade do *software*, assim como foi observado em (YABE *et al.*, 2018).

Para verificar a relação entre as dimensões da segunda pergunta do questionário, foi aplicada uma análise fatorial, que indicou que as dez dimensões estão relacionadas com um fator e, portanto, todas elas poderiam ser utilizadas como variáveis na elaboração de modelos de predição. A análise fatorial dos dados do questionário *online* mostrou que as dimensões Paciência, Estresse e Pontualidade são as três dimensões dentre as dez que possuem maior relevância na composição do fator. Sendo assim, essas dimensões foram utilizadas para montar o primeiro modelo de predição de confiabilidade dentre os quatro modelos ajustados neste trabalho.

O ajuste dos modelos de predição foi realizado através de regressão linear múltipla, na qual as médias dos escores atribuídos pelos participantes às dimensões em cada cenário foram utilizadas como variáveis independentes e a média da confiabilidade percebida pelos participantes em cada cenário foram utilizadas como variável independente em todos os modelos. Os Modelos 1 e 2 foram ajustados com os dados do questionário *online*. No Modelo 1, as três dimensões mais relevantes apontadas pela análise fatorial dos dados do questionário *online* foram utilizadas no ajuste do modelo. O Modelo 2 foi ajustado utilizando as dimensões apontadas pelo Teste t como significantes dentre as dez dimensões. Esse teste é utilizado para verificar quais são as variáveis relevantes para ajustar um modelo de regressão linear múltipla, que no caso do Modelo 2 foram as dimensões Satisfação e Tolerância. Ao comparar o resultado da confiabilidade predita pelos Modelos 1 e 2 com a confiabilidade percebida pelos usuários e com a confiabilidade estimada pelo método clássico, nota-se que ambos os modelos tiveram um resultado que se aproximou significativamente mais da percepção dos usuários que o resultado da confiabilidade estimada pelo método clássico, com o Modelo 2 apresentando melhor precisão que o Modelo 1.

Os Modelos 3 e 4 foram ajustados com os dados do questionário presencial. A análise fatorial dos dados do questionário presencial também mostrou que as dez dimensões possuem relação com um fator e além disso, indicou as dimensões Estresse, Pontualidade e Paciência como as de maior relevância na composição do fator. Entretanto, na análise fatorial dos dados

do questionário *online* a ordem de relevância foi: Paciência, Estresse e Pontualidade. Sendo assim, o Modelo 3 foi ajustado considerando essas três dimensões, assim como foi feito no ajuste do Modelo 1. Para ajustar o Modelo 4, foram consideradas as dimensões mais relevantes apontadas pelo Teste t, que neste caso indicou apenas a dimensão Satisfação como relevante. Ao comparar o resultado da confiabilidade predita pelos Modelos 3 e 4 com a confiabilidade percebida pelos usuários e com a confiabilidade estimada pelo método clássico, pode-se observar que o método clássico obteve melhor desempenho que o Modelo 3 apenas no cenário 1, por pouca diferença. Nos outros cenários, os Modelos 3 e 4 apresentaram melhor desempenho que a confiabilidade estimada pelo método clássico. Portanto, o Modelo 4 apresentou melhor desempenho que o Modelo 3. Note que tanto para os dados do questionário *online* quanto para os dados do questionário presencial as dimensões apontadas pela análise fatorial como relevantes na composição do fator não compuseram os modelos de predição de confiabilidade mais precisos. Portanto, pode-se concluir que apesar de a análise fatorial mostrar a relação entre as dimensões, essa técnica não auxilia na indicação das melhores dimensões para o ajuste de modelos de predição.

Note que os modelos ajustados que apresentaram maior precisão na predição de confiabilidade (Modelos 2 e 4) tiveram em comum a dimensão Satisfação. Conclui-se que para elaborar modelos de predição de confiabilidade de *software* que considerem a percepção dos usuários deve-se levar em consideração a satisfação desses usuários ao utilizar o sistema.

Por fim, ao comparar como cada dimensão foi percebida em cada cenário com o Teste de Kruskal-Wallis, observa-se que não houve diferença significativa nas dimensões entre os cenários com alto impacto (cenários 3 e 4), indicando que as falhas apresentadas nesses cenários foram vivenciadas com mesma intensidade. Esse fenômeno foi observado nos dados de ambos os questionários. Entre os cenários 2 e 3, foi observada diferença significativa em todas as dimensões, exceto na dimensão Paciência nos dados do questionário *online*. Esse resultado reforça a conclusão de que a consequência provocada por uma falha influencia a experiência do usuário, pois o cenário 2 é de baixo impacto e o cenário 3 de alto impacto. Entre os cenários 1 e 2, foi observada diferença significativa nos dados do questionário *online* nas dimensões Satisfação, Tolerância, Produtividade, Paciência, Tranquilidade e Capacidade. Já nos dados do questionário presencial foi observada diferença significativa nas dimensões Satisfação, Preocupação, Tolerância, Produtividade, Estresse, Tranquilidade e Capacidade. Essa diferença entre os cenários pode ser explicada pela recuperação mais rápida da falha apresentada no cenário 1 em relação à falha descrita no cenário 2.

5.2 Resposta para as Perguntas de Pesquisa

Fatores que influenciam a QoE do usuário afetam a percepção do usuário em relação à confiabilidade de *software*?

Resposta: Os cenários apresentados no questionário foram vivenciados com diferentes intensidades devido a dois fatores: a consequência provocada pela falha e o MTBF. Os cenários em que a falha prejudicava a tarefa que o usuário estava fazendo no computador foram percebidos com confiabilidade muito menor que a confiabilidade estimada pelo método clássico. Outro achado importante é que o MTBF longo impacta significativamente de forma

negativa na percepção dos usuários em relação à confiabilidade do *software*, diferentemente da confiabilidade estimada pelo método clássico que é aumentada com um valor maior de MTBF. Portanto, pode-se concluir que os fatores que influenciam a QoE afetam sim a percepção do usuário em relação à confiabilidade do *software*. Neste estudo, os fatores sobressaíram nas análises estatísticas realizadas foram: Estresse, Pontualidade, Paciência, Tolerância e Satisfação. Embora a Satisfação tenha se destacado como o principal fator para o cálculo da confiabilidade de *software* considerando a percepção dos usuários perante falhas de sistema, os outros fatores levantados podem ser levados em consideração em pesquisas futuras para a análise de outros aspectos de qualidade de *software*.

O resultado do cálculo da confiabilidade de um *software* (abordagem atual, QoS) é compatível com a real percepção dos usuários (QoE) sobre a confiabilidade do *software* usado?

Resposta: Neste trabalho foram produzidos modelos preditores de confiabilidade percebida pelos usuários considerando aspectos de QoE. Ao analisar a diferença entre o resultado da confiabilidade estimada pelo método clássico (QoS) e a confiabilidade percebida pelos usuários (QoE), bem como a diferença entre a confiabilidade percebida pelos usuários e os modelos preditores de confiabilidade (QoE), observa-se que os resultados apresentados pelos modelos se aproximaram significativamente mais da confiabilidade percebida pelos usuários do que o resultado da confiabilidade estimada pelo método clássico. Portanto, pode-se concluir que a confiabilidade do *software* calculada com base apenas em parâmetros de QoS não representa a real percepção de confiabilidade que os usuários têm em relação ao *software*. Com o uso de análise fatorial e regressão linear, observou-se que a Satisfação dos usuários com relação ao uso do *software* é um dos principais fatores de influência que pode auxiliar a calcular a percepção de confiabilidade de sistemas dos usuários de forma mais precisa que o modelo de cálculo de confiabilidade com Poisson, que foi o modelo analisado neste estudo.

5.3 Ameaças à Validade e Limitações da Pesquisa

Neste trabalho foram coletados dados através de um questionário *online*. Embora essa abordagem possua suas vantagens, devem-se levar em consideração algumas características ao avaliar os resultados apresentados nesta dissertação. A primeira delas é que os participantes podem não ter entendido as perguntas ao lerem o questionário *online*. Outro ponto importante é que as respostas dos participantes nas escalas concentraram-se no polo negativo, o que pode ter ocorrido devido aos participantes clicarem na primeira opção para responderem ao questionário mais rápido. Portanto, também foi aplicado um questionário presencial supervisionado pelo autor desta dissertação para amenizar essa ameaça à validade.

Para mitigar a influência que a opinião de um participante pode ter nas respostas de outro participante durante as sessões de coleta de dados, entrevistou-se um participante por vez. Além disso, ao final da coleta de dados, o participante entrevistado era orientado a não relatar a outras pessoas informações relacionadas ao estudo. Devido à necessidade de utilização de equipamento específico para a realização das sessões de coleta de dados (questionário presencial), o estudo foi realizado em apenas duas cidades diferentes. Portanto, fatores culturais e regionais podem ter influenciado as respostas dos participantes.

Outra ameaça é a ordem em que os cenários do questionário foram apresentados para os participantes. Para atenuar essa ameaça, os cenários elaborados como de alto impacto foram apresentados após os cenários de baixo impacto, para evitar que os participantes respondessem ao restante do questionário generalizando sua experiência com base em uma péssima primeira experiência. Entretanto, todos os participantes foram expostos à mesma ordem de falhas e a mais de uma falha, o que pode ter influenciado a experiência dos usuários no decorrer do estudo. Para mitigar essa ameaça à validade seria necessário avaliar as falhas individualmente e em diferentes ordens, para verificar se houve alteração significativa na experiência dos usuários alterando a ordem, quantidade e sequência de falhas vivenciadas pelos usuários. Algumas características, como o cansaço físico e mental dos participantes e do pesquisador que supervisionou o experimento, os diferentes horários em que a coleta de dados foi realizada (das 08:00 às 21:00 horas, em média) as condições de iluminação e temperatura do local de coleta de dados, bem como o *hardware* utilizado pelos participantes também podem ter influenciado os resultados obtidos.

Por fim, este estudo apresenta algumas limitações, como a avaliação da confiabilidade percebida pelos usuários em uma única versão do Microsoft Windows 7, bem como não foi feita a avaliação em diferentes sistemas operacionais.

5.4 Dificuldades Encontradas

A principal dificuldade encontrada neste trabalho foi encontrar pessoas com disponibilidade para participar da coleta de dados do questionário presencial. Para mitigar essa dificuldade, procurou-se realizar as coletas de dados em diversos horários.

Outra dificuldade encontrada foi elaborar os cenários utilizados no questionário. Procurou-se elaborar os cenários da forma mais realista possível para que os participantes se sentissem em um ambiente real durante a aplicação do questionário presencial e conseguissem se sentir vivenciando as falhas ao lerem os cenários no questionário *online*.

5.5 Trabalhos Futuros

Em curto prazo, pretende-se realizar sessões de coleta de dados com usuários de outros países para avaliar se o aspecto cultural realmente influencia na QoE dos usuários, comparando os resultados deste trabalho com as análises da amostra que será coletada futuramente. Além disso, podem-se realizar coletas alterando a ordem em que as falhas foram expostas aos participantes para verificar se haverá diferença em relação aos resultados deste estudo. Por fim, espera-se investigar a percepção de confiabilidade dos usuários em relação a outros sistemas operacionais, como Android e GNU/Linux e outras versões de SO da família Windows.

Em médio prazo, espera-se conseguir aprimorar o método de avaliação de confiabilidade utilizado nesta dissertação, acrescentando novas falhas ao instrumental e melhorando a simulação das falhas já existentes. Também, planeja-se capturar informações biométricas dos participantes de forma não invasiva durante as sessões de coleta de dados, como movimento ocular, batimentos cardíacos, entre outros, para comparar as respostas dos participantes com suas informações biométricas.

Em longo prazo, conjectura-se desenvolver um método de predição e melhoria da QoE do usuário em relação à confiabilidade do SO durante o uso do computador. Espera-se prever a confiabilidade percebida pelos usuários utilizando conceitos de predição de falhas e padrões de ocorrência de falhas. Pode-se melhorar a QoE dos usuários baseando-se em perfis de uso e padrões de percepção de falha. Tais informações podem ser coletadas através de avaliações respondidas pelos usuários, por exemplo, utilizando escalas bipolares como as que foram utilizadas neste trabalho.

Referências

- ABOWD, D. et al. **Towards a Better Understanding of Context and Context-Awareness**. Handheld and Ubiquitous Computing. Berlin. 1999. p. 304 - 307. doi: 10.1007/3-540-48157-5_29 https://doi.org/10.1007/3-540-48157-5_29
- AL-BADAREEN, B. et al. **Users' Perspective of Software Quality**. Proceedings of the 10th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems. 2011. p. 84 - 89. ISBN: 978-960-474-277-6
- AMERICAN NATIONAL STANDARDS INSTITUTE; INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. **Standard Glossary of Software Engineering Terminology**. 1991. <https://doi.org/10.1109/IEEESTD.1990.101064>
- ASSUNÇÃO, L. M. **Instrumental para Coleta de Dados sobre a Experiência do Usuário em Confiabilidade de Sistemas Operacionais**. Trabalho de Conclusão de Curso (Sistemas de Informação) - UFU. Uberlândia. 2016.
- AVIZIENIS, A. et al. Basic Concepts and Taxonomy of Dependable and Secure Computing. **IEEE Transactions on Dependable and Secure Computing**, v. 1, n. 1, p. 11-33, 2004. <https://doi.org/10.1109/TDSC.2004.2>
- CHEN, K.-T.; TU, C.-C.; XIAO, W.-C. **OneClick: A Framework for Measuring Network Quality of Experience**. IEEE INFOCOM. Rio de Janeiro. 2009. p. 702 - 710. doi: 10.1109/infcom.2009.5061978 <https://doi.org/10.1109/INFCOM.2009.5061978>
- DOS SANTOS, C. A. R. **Um Estudo Exploratório Sobre Padrões de Falhas de Software de Sistemas Operacionais**. Dissertação (Ciência da Computação) - UFU. Uberlândia. 2016. <https://doi.org/10.5753/ctd.2017.3456>
- FÁVERO, L. P. et al. **Análise de dados: modelagem multivariada para tomada de decisões**. 1ª. ed. Rio de Janeiro: Elsevier, 2009. ISBN: 8535230467
- FIEDLER, M.; HOSSFELD, T.; TRAN-GIA, P. A Generic Quantitative Relationship between Quality of Experience and Quality of Service. **IEEE Network**, v. 24, n. 2, p. 36 - 41, Março 2010. <https://doi.org/10.1109/MNET.2010.5430142>
- GANAPATHI, A.; GANAPATHI, V.; PATTERSON, D. **Windows XP Kernel Crash Analysis**. Proceedings of the 20th conference on Large Installation System Administration. Berkeley. 2006. p. 149 - 159.
- GOLDENBERG, M. **A arte de pesquisar - Como fazer pesquisa qualitativa em Ciências Sociais**. 8ª. ed. Rio de Janeiro: Record, 2004. ISBN: 978-85-0109680-7

HAIR, J. F. et al. **Análise Multivariada de Dados**. Tradução de Adonai Schlup Sant'Anna. 6ª. ed. Porto Alegre: Bookman, 2009.

HASSENZAHN, M. **The hedonic/pragmatic model of user experience**. British HCI Group Annual Conference on People and Computers: HCI.but not as we know it. Lancaster. 2007. p. 10 - 14.

HASSENZAHN, M.; DIEFENBACH, S.; GORITZ, A. Needs, affect, and interactive products – Facets of user experience. **Interacting with Computers**, v. 22, p. 353 - 362, 2010. <https://doi.org/10.1016/j.intcom.2010.04.002>

HASSENZAHN, M.; TRACTINSKY, N. User Experience - A Research Agenda. **Behavior & Information Technology**, v. 25, p. 91 - 97, 2006. <https://doi.org/10.1080/01449290500330331>

ICKIN, S. et al. Factors Influencing Quality of Experience of Commonly-Used Mobile Applications. **IEEE Communications Magazine**, v. 50, n. 4, p. 48 - 56, Abril 2012. <https://doi.org/10.1109/MCOM.2012.6178833>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Ergonomics of human-system interaction**. 2010. <https://doi.org/10.3403/BSISO9241>

INTERNATIONAL TELECOMMUNICATION UNION. **Methods for subjective determination of transmission quality**. Geneva. 1996.

JALOTE, et al. **Measuring Reliability of Software Products**. Proceedings of 15th International Symposium on Software Reliability Engineering. 2004. p. 47 - 53.

LAW, E. L.-C. et al. **Understanding, Scoping and Defining User eXperience**. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston. 2009. p. 719 - 728. <https://doi.org/10.1145/1518701.1518813>

LEWIS, E. E. **Introduction to Reliability Engineering**. 2ª. ed. Evanston: John Wiley & Sons, Inc., v. 1, 1996. ISBN: 9780471018339

LIOTOU, E. et al. Quality of experience management in mobile cellular networks: key issues and design challenges. **IEEE Communications Magazine**, v. 53, n. 7, p. 145 - 153, Julho 2015. <https://doi.org/10.1109/MCOM.2015.7158278>

LYU, M. R. **Software Reliability Engineering: A Roadmap**. Future of Software Engineering. 2007. p. 153 - 170. <https://doi.org/10.1109/FOSE.2007.24>

MARSDEN, V.; WRIGHT, J. (Eds.). **Handbook of Survey Research**. 2ª. ed. Emerald, 2010. ISBN: 9781848552241

MATIAS, R. et al. **An Empirical Exploratory Study on Operating System Reliability**. Proceedings of the 29th Annual ACM Symposium on Applied Computing. Gyeongju. 2014. p. 1523 - 1528. <https://doi.org/10.1145/2554850.2555021>

MATIAS, R.; ANTUNES, M. P. **Confiabilidade de Sistemas Operacionais de Propósito Geral: Um Estudo de Caso.** Proceedings of the 4th Brazilian Symposium on Computing Systems Engineering. 2014. p. 1 - 6.

MATIAS, R.; OLIVEIRA, ; ARAÚJO, L. **Operating System Reliability from the Quality of Experience Viewpoint: an Exploratory Study.** Proceedings of 28th ACM Symposium on Applied Computing. 2013. p. 1644 - 1649. <https://doi.org/10.1145/2480362.2480669>

MICROSOFT. Assemblies in the Common Language Runtime. **Docs**, 2012. Disponível em: <[https://docs.microsoft.com/en-us/previous-versions/visualstudio/visual-studio-2008/hk5f40ct\(v=vs.90\)>](https://docs.microsoft.com/en-us/previous-versions/visualstudio/visual-studio-2008/hk5f40ct(v=vs.90)>). Acesso em: 17 dez. 2018.

MICROSOFT. (Developer Content) Bug Check 0xC2: BAD_POOL_CALLER. **Docs**, 2017. Disponível em: <<https://docs.microsoft.com/en-us/windows-hardware/drivers/debugger/bug-check-0xc2--bad-pool-caller>>. Acesso em: 17 dez. 2018.

MOLLER, S.; RAAKE, A. **Quality of Experience Advanced Concepts, Applications and Methods.** 1ª. ed. Berlin: Springer, v. 1, 2014. ISBN: 9783319026800
<https://doi.org/10.1007/978-3-319-02681-7>

NETMARKETSHARE. **Operating System Share by Version**, 2018. Disponível em: <<https://netmarketshare.com/operating-system-market-share.aspx?id=platformsDesktopVersions>>. Acesso em: 22 nov. 2018.

OTT, R. L.; LONGNECKER, M. **An Introduction to Statistical Methods & Data Analysis.** 7ª. ed. Boston: Cengage Learning, 2016. ISBN: 1305269470

PASQUALI, L. **Instrumentação Psicológica: fundamentos e práticas.** 1ª. ed. Porto Alegre: Artmed, 2010. ISBN: 9788536321066

PEREIRA, J. C. R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais.** 3ª. ed. São Paulo: Edusp, 2004. ISBN: 9788531405235

RAINER, B.; TIMMERER, C. A Generic Utility Model Representing the Quality of Sensory Experience. **ACM Transactions on Multimedia Computing, Communications, and Applications**, v. 11, n. 1, p. 1 - 14, 2014. <https://doi.org/10.1145/2648429>

REICHL, P.; TUFFIN, B.; SCHATZ, R. Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. **Telecommunication Systems**, v. 52, n. 2, p. 587 - 600, Fevereiro 2013. <https://doi.org/10.1007/s11235-011-9503-7>

SWIFT, M. M.; BERSHAD, B. N.; LEVY, H. M. **Improving the Reliability of Commodity Operating Systems.** Proceedings of the nineteenth ACM Symposium on Operating Systems Principles. New York. 2003. p. 207 - 222. <https://doi.org/10.1145/945465.945466>

TÉCNICAS, A. B. D. N. **Engenharia de software - Qualidade de produto.** 2003.

TSOLKAS, D. et al. A survey on parametric QoE estimation for popular services. **Journal of Network and Computer Applications**, v. 77, p. 1 - 17, Janeiro 2017. <https://doi.org/10.1016/j.jnca.2016.10.016>

UNION, I. T. **Quality of telecommunication services: concepts, models, objectives and dependability planning - Terms and definitions related to the quality of telecommunication services**. p. 3. 2008.

VEERARAGAVAN, R. et al. Modeling QoE in Dependable Tele-Immersive Applications: A Case Study of World Opera. **IEEE Transactions on Parallel and Distributed Systems**, v. 27, n. 9, p. 2667 - 2681, Setembro 2016. <https://doi.org/10.1109/TPDS.2015.2503291>

VIEIRA, S. **Como elaborar questionários**. 1^a. ed. São Paulo: Atlas, 2009. ISBN: 8522455732

WOOLDRIGE, J. M. **Introductory econometrics: a modern approach**. 5^a. ed. South Western: Thompson, 2012. ISBN: 1111531048

XU, J.; KALBARCZYK, Z.; IYER, R. K. **Networked Windows NT System Field Failure Data Analysis**. Proceeding of the 1999 Pacific Rim International Symposium on Dependable Computing. Washington. 1999. p. 178 - 185. <https://doi.org/10.1109/PRDC.1999.816227>

YABE, et al. **Influence Factors on the Quality of User Experience in OS Reliability: A Qualitative Experimental Study**. Proceedings of the 13th International Conference on Availability, Reliability, and Security. Hamburg. 2018. <https://doi.org/10.1145/3230833.3230862>

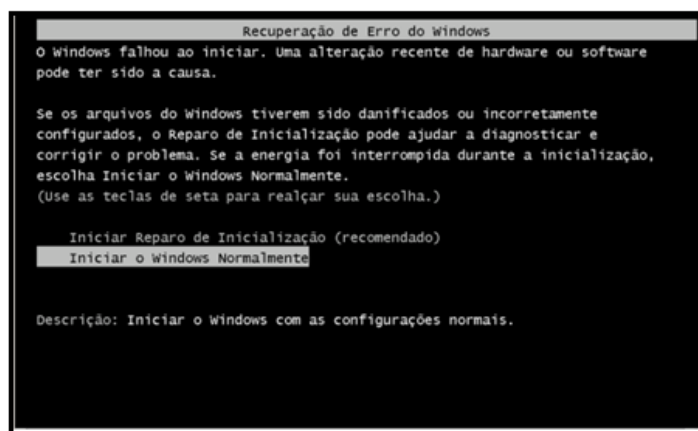
YOUNG, R. M.; BARNARD, P. **The use of scenarios in human-computer interaction research: turbocharging the tortoise of cumulative science**. Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface. Toronto. 1987. p. 291 - 296. <https://doi.org/10.1145/29933.275645>

ZANELLI, J. C. Pesquisa qualitativa em estudos da gestão de pessoas. **Estudos de Psicologia**, v. 7, p. 79 - 88, 2002. <https://doi.org/10.1590/S1413-294X2002000300009>

Apêndices

Cenários do Questionário

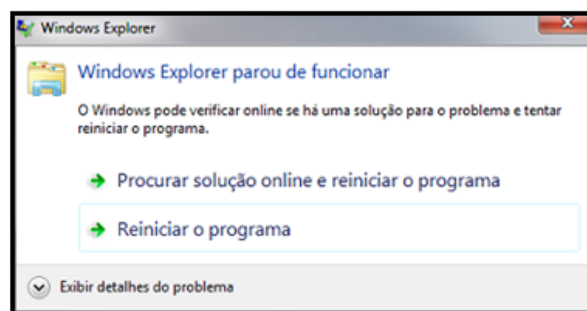
Você está no trabalho e liga o computador para começar suas atividades. Durante a inicialização do sistema operacional ocorre uma falha e aparece na tela a mensagem da figura abaixo. Você escolhe a opção *Iniciar o Windows Normalmente* e continua a utilizar o computador.



Nos últimos seis meses, esta falha tem ocorrido a cada inicialização do computador, em média 1 vez por dia, normalmente no início do dia de trabalho.

Figura A.1 - Cenário 1

Você está procurando um arquivo armazenado no computador da empresa. Ao acessar diferentes pastas (diretórios) em busca do arquivo, aleatoriamente aparece a mensagem abaixo:



Esta falha vem ocorrendo há seis meses, em média uma vez por semana. Sempre que esta falha ocorre você clica na opção *Reiniciar o programa* e consegue dar continuidade ao que estava fazendo.

Figura A.2 - Cenário 2

Você está participando de uma reunião e é solicitado(a) a realizar anotações para escrever um relatório. Você abre o editor de texto e digita as anotações na medida em que a reunião se desenvolve. Durante a reunião o computador parou de funcionar (“travou”) uma vez e apresentou a mensagem abaixo. Este problema começou há aproximadamente seis meses e ocorre aleatoriamente, em média uma vez por dia. A cada ocorrência desta falha o computador deve ser reinicializado, o que pode causar a perda de trabalho, como ocorreu neste caso. Parte das anotações foi perdida e precisou ser redigitada.

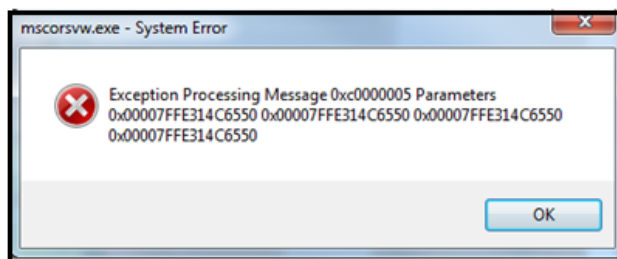


Figura A.3 - Cenário 3

Você está no trabalho preparando os slides de uma apresentação. Durante esta atividade ocorre uma falha no sistema operacional, gerando a mensagem da figura abaixo:

```
A problem has been detected and Windows has been shut down to prevent damage
to your computer.

The problem seems to be caused by the following file: usbser.sys
BAD_POOL_CALLER

If this is the first time you've seen this stop error screen,
restart your computer. If this screen appears again, follow
these steps:

Check to make sure any new hardware or software is properly installed.
If this is a new installation, ask your hardware or software manufacturer
for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware
or software. Disable BIOS memory options such as caching or shadowing.
If you need to use safe mode to remove or disable components, restart
your computer, press F8 to select Advanced Startup Options, and then
select Safe Mode.

Technical Information:

*** STOP: 0x000000c2 (0x0000000000000007, 0x000000000000109b, 0x0000000005070000,
0xfffffa80148b31bb)

*** usbser.sys - Address 0xfffff8800ce69b56 base at 0xfffff8800ce62000 DateStamp
0x4ce7a66d
```

Após reiniciar o computador, você percebe que perdeu o arquivo com os slides da apresentação em que estava trabalhando, lhe exigindo refazê-los. Esta falha ocorre ao menos uma vez a cada 15 dias e sempre causa algum retrabalho.

Figura A.4 - Cenário 4

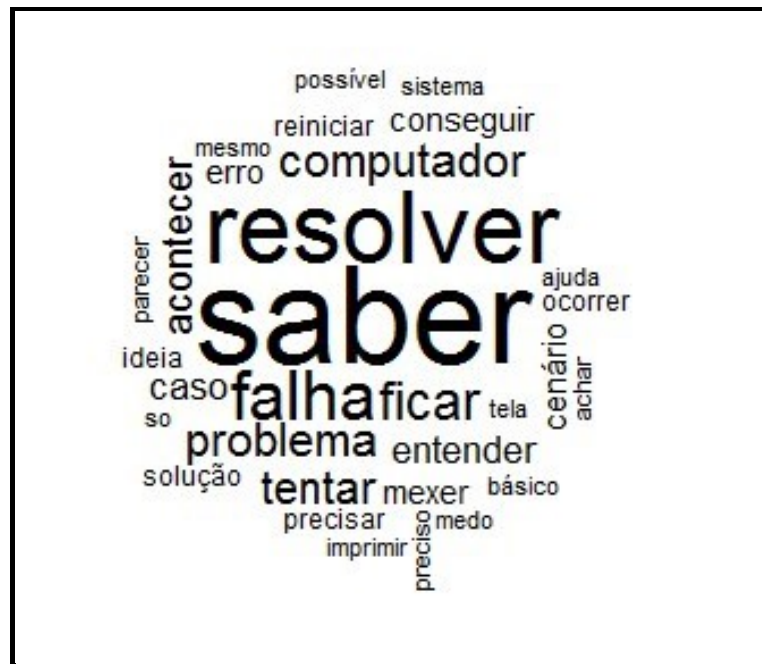


Figura B.7 - Nuvem de palavras produzida com as respostas da categoria Não Familiaridade

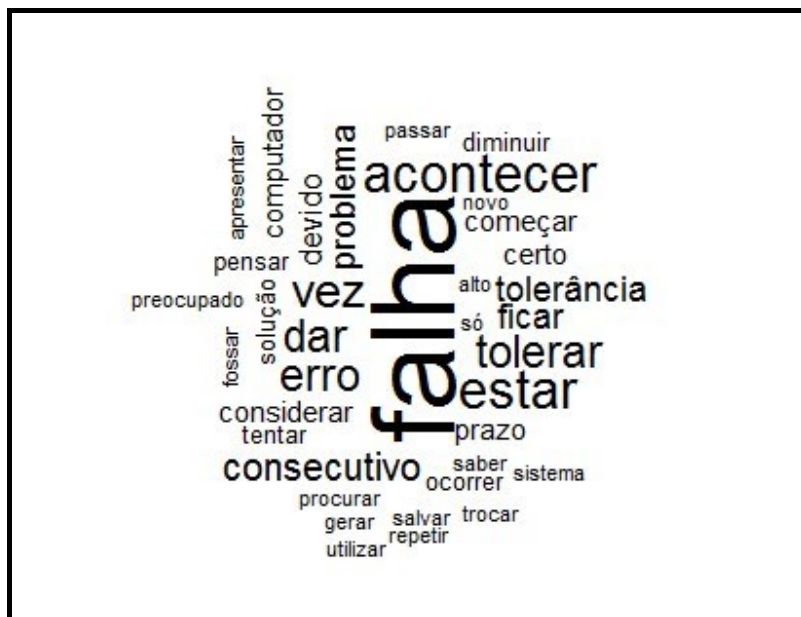


Figura B.8 - Nuvem de palavras produzida com as respostas da categoria Frequência de Falha

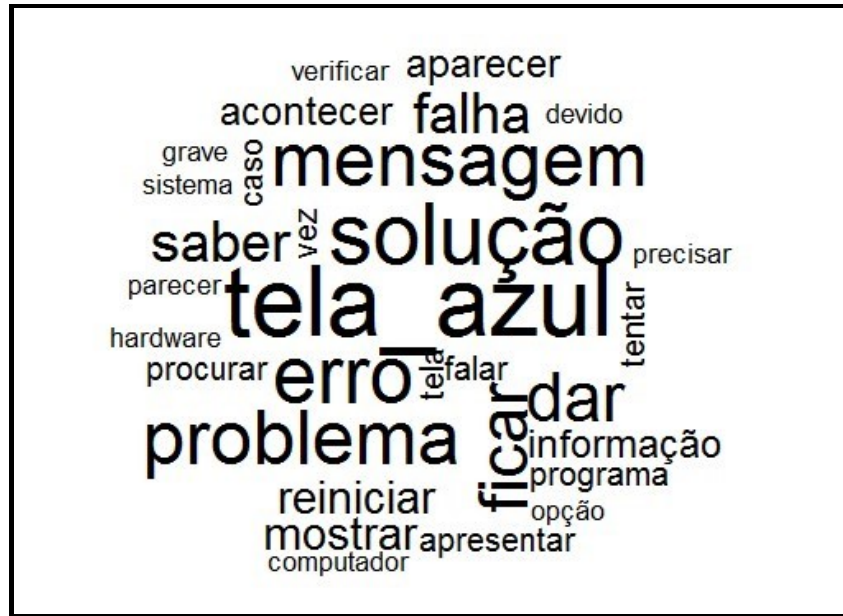


Figura B.9 - Nuvem de palavras produzida com as respostas da categoria Mensagem de Falha



Figura B.10 - Nuvem de palavras produzida com as respostas da categoria Recuperação

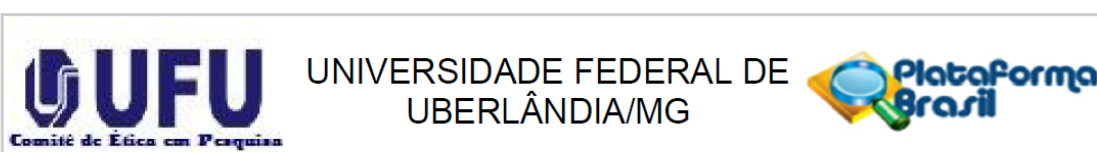
Participantes Descartados

Tabela C.1 - Participantes descartados da amostra do questionário presencial

Número dos participantes	Motivo de descarte
10, 87	Ocorreu um erro no <i>notebook</i> em que o participante respondeu as perguntas, fazendo com que as respostas inseridas fossem perdidas.
14	O participante respondeu a dimensão Pontualidade de forma aleatória, pois ficou com dúvida durante a coleta de dados, mas ficou com receio de sanar sua dúvida com o pesquisador.
49, 52, 54, 61, 62,70	Participou do questionário <i>online</i> .
56	O participante se distraiu e fechou a mensagem de falha sem querer, o que afetou sua percepção de falha.
64	O participante respondeu as escalas de forma invertida, considerando o polo negativo como positivo e vice versa.
67, 95	A primeira falha não foi simulada da forma correta devido a uma queda de desempenho no computador <i>desktop</i> que executou o instrumental.
69	O participante se sentiu mal durante a coleta de dados e pediu para não prosseguir.
93	O participante respondeu a dimensão Capacidade de forma aleatória, pois ficou com dúvida durante a coleta de dados e não informou isso ao pesquisador.
94	O participante não seguiu o roteiro de tarefas da forma correta, prejudicando a simulação do cenário 2.

Anexos

Parecer do Comitê de Ética em Pesquisas com seres humanos da Universidade Federal de Uberlândia



UNIVERSIDADE FEDERAL DE
UBERLÂNDIA/MG



PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Análise qualitativa da experiência do usuário em confiabilidade de sistemas operacionais

Pesquisador: RIVALINO MATIAS JUNIOR

Área Temática:

Versão: 4

CAAE: 64081616.4.0000.5152

Instituição Proponente: Faculdade de Computação

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.708.755

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não