



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO
CENTRO DE INFORMÁTICA
2017.2



Trabalho de Graduação

Estudo Sistemático Sobre Testes de Usabilidade
Realizados em Aplicativos Móveis

Aluno: Valdi Ferreira do Nascimento Júnior (vfnj@cin.ufpe.br)

Orientador: Alexandre Vasconcelos (amlv@cin.ufpe.br)

Co-Orientador: Ivaldir de Farias Junior (ihfj@cin.ufpe.br)

Recife

2017

Valdi Ferreira do Nascimento Júnior

Estudo Sistemático Sobre Testes de Usabilidade
Realizados em Aplicativos Móveis

Trabalho de Conclusão de Curso apresentado ao curso de Sistemas de Informação da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Dr. Alexandre Vasconcelos
Co-Orientador: Dr. Ivaldir de Farias Junior

Recife
2017

Valdi Ferreira do Nascimento Júnior

**Estudo Sistemático Sobre Testes de Usabilidade
Realizados em Aplicativos Móveis**

Trabalho de Conclusão de Curso apresentado ao curso de Sistemas de Informação da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Aprovado em _____ de _____ de _____.

BANCA EXAMINADORA:

Prof. Dr. Alexandre Vasconcelos

Prof. Dr. Leopoldo Teixeira

Dedico este trabalho aos meus pais.

AGRADECIMENTOS

Agradeço a minha família que sempre estive ao meu lado em todos os momentos da minha vida,

Gostaria de agradecer ao meu orientador Alexandre Marcos Lins de Vasconcelos e ao meu co-orientador Ivaldir de Farias Junior que, com sabedoria, paciência e dedicação, estiveram ao meu lado durante este trabalho.

Agradeço a minha família por todo apoio, carinho, e dedicação que me deram durante toda a vida, pela quais eu sou eternamente grato.

Agradeço a minha família por todo apoio, carinho, e dedicação que me deram durante toda a vida, pela quais eu sou eternamente grato.

Resumo

Os dispositivos móveis estão presentes em nossas vidas oferecendo acesso a informações, serviços e funções em diversas áreas, promovendo o aumento da produtividade e melhoria da qualidade de vida sem restrição de local ou tempo. Essas facilidades tornam os aplicativos móveis populares, atraindo um grande número de organizações e desenvolvedores, tornando a área bastante competitiva e elevando os níveis de exigência necessários para o sucesso de uma aplicação. Assim, além de entregar funcionalidades desejadas, os aplicativos devem apresentar um alto grau de usabilidade para que as expectativas dos usuários sejam atendidas de forma eficiente, eficaz e satisfatória. Um aplicativo deve ser fácil de usar, possibilitando a execução de tarefas da melhor forma possível. Para avaliar esse aspecto da qualidade dos softwares, muitos desenvolvedores utilizam testes de usabilidade, onde observam e analisam a interação dos usuários com o aplicativo com o objetivo de detectar problemas relacionados à usabilidade. Este estudo tem como objetivo investigar como são realizados os testes de usabilidade de acordo com a literatura, bem como identificar os problemas e limitações relacionados, permitindo desta forma auxiliar no planejamento para a execução deste tipo de testes de forma a evitar as limitações e problemas identificados no levantamento realizado.

Palavras-Chave: Testes, usabilidade, aplicações móveis

Abstract

Mobile devices are present in our lives providing access to information, services and functions in several areas, promoting the increase of the productivity and improvement of the quality of life, without restriction of place or time. These facilities make mobile applications popular by attracting a large number of organizations and developers making the area quite competitive and raising the levels of demand necessary for the success of an application. To do this, in addition to delivering desired functionality, applications must have a high degree of usability so that users' expectations are met efficiently, effectively, and satisfactorily. An application should be easy to use, making it possible to perform tasks in the best possible way. To evaluate this aspect of software quality, many developers use usability tests, where they observe and analyze the interaction of users with the application in order to detect problems related to the usability. This study aims to investigate how the usability tests are performed according to the literature, as well as to identify the related problems and limitations, allowing in this way to assist in the planning for the execution of this type of tests in order to avoid the limitations and problems identified.

Keywords: Tests, usability, mobile application

Lista de Figuras

Figura 1 - Resultado da busca feita nas bases selecionadas	28
Figura 2 - Interseção dos resultados da busca feita nas bases selecionadas	28
Figura 3 - Resultado da fase 2 da seleção de estudos	29
Figura 4 - Resultado da fase 3 da seleção de estudos	30
Figura 5 - Resultado da fase 4 da seleção de estudos	30
Figura 6 - Distribuição dos estudos por nível de qualidade	31
Figura 7 - Distribuição dos estudos por faixa de qualidade.....	31
Figura 8 - Distribuição das publicações selecionadas ao longo dos anos.....	32
Figura 9 - Distribuição por área	33
Figura 10 - Distribuição de autoria por países.....	33

Lista de Quadros

Quadro 1 - <i>String</i> de Busca	25
Quadro 2 - Distribuição de pontuação por classificação	26

Lista de Tabelas

Tabela 1 - Artigos X Técnica/Método/Processo/Práticas	35
---	----

Sumário

1	Introdução	11
1.1	Questões de Pesquisa	12
1.2	Objetivo	13
1.3	Estrutura do Trabalho	13
2	Referencial Teórico.....	14
2.1	Testes	14
2.2	Usabilidade	16
2.3	Aplicativos Móveis	19
2.4	Síntese do Capítulo	21
3	Metodologia de Pesquisa	22
3.1	Protocolo de Estudo Sistemático	24
3.1.1	Questões de Pesquisa.....	24
3.1.2	Estratégia de Busca.....	25
3.1.3	Seleção dos estudos	25
3.1.4	Avaliação da Qualidade.....	26
3.1.5	Extração, Análise e Síntese dos dados	26
3.2	Síntese do Capítulo	26
4	Resultados.....	27
4.1	Execução.....	27
4.1.1	Busca	27
4.1.2	Seleção por título, resumo e palavras-chave	29
4.1.3	Seleção dos estudos pela introdução e conclusão	29
4.1.4	Seleção pela leitura completa dos artigos.....	30

4.1.5	Avaliação da Qualidade.....	30
4.1.6	Visão Geral dos Estudos.....	32
4.2	Mapeamento das evidências	34
4.3	Síntese do Capítulo	39
5	Conclusão.....	41
5.1	Limitações de Pesquisa.....	42
5.2	Trabalhos Futuros.....	42
6	Bibliografia	43
	Apêndice A – Protocolo de Revisão Sistemática	45
	Apêndice B – Artigos Selecionados	57
	Apêndice C – Avaliação dos Artigos	59
	Apêndice D – Contexto dos Artigos Selecionados	60
	Apêndice E – Referências dos Artigos Selecionados.....	61

1 Introdução

Cada vez mais pessoas do mundo todo estão utilizando funcionalidades e consumindo serviços através de aplicativos em dispositivos móveis (smartphones, tablets, etc.). O aumento significativo do acesso das pessoas a esta tecnologia, bem como o aumento da oferta e disponibilidade dos serviços de comunicação, fez com que o número de usuários ativos e conectados aumentasse nos últimos anos. De maneira conveniente e eficiente, os dispositivos e seus aplicativos atendem às necessidades de seus usuários independente de local e hora (DINH et al., 2013).

Devido ao grande número de dispositivos no mercado, ao aumento da sua capacidade computacional e às novas tecnologias que se beneficiam de sua mobilidade para alcançar os usuários, como a computação em nuvem e a computação ubíqua, o ambiente móvel é uma área muito atraente para desenvolvedores e organizações que oferecem serviços nas mais diversas áreas como educação, saúde, finanças, localização e outras (CORRAL, SILLITTI e SUCCI, 2015).

Com milhares de aplicativos já lançados e publicados nas *mobile application markets* (lojas das principais plataformas de dispositivos móveis que reúnem milhares de aplicativos disponíveis para *download*), esse mercado é bastante competitivo, o que torna a qualidade um fator essencial para o sucesso. Além disso, os desenvolvedores precisam lidar com fatores críticos (limites de capacidade de processamento e armazenamento, diversidade de protocolos e meios de acesso à rede, variedade de plataformas e dispositivos, segurança e etc.) e explorar características específicas dessa tecnologia (ex. mobilidade, geolocalização, variedade de interação com o usuário) a fim de entregar aplicativos que atendam, e até mesmo superem, a expectativa dos usuários (CORRAL, SILLITTI e SUCCI, 2015).

Para atender essa expectativa, um software não precisa apenas entregar o serviço ou funcionalidade desejada, é importante também que seja fácil de aprender e usar, ou seja, apresente níveis satisfatórios de usabilidade, que é o fator que determina justamente a produtividade e aceitação de um produto (ABRAN et al., 2003). Em dispositivos móveis esse fator tem um impacto ainda maior devido a limitações de tamanho de tela e aos diferentes meios de interação (toque, gestos,

vibração, voz, teclado e etc.) fortemente presentes nestes dispositivos e que são amplamente usados pelos seus usuários (WASSERMAN, 2010).

Uma das principais formas de verificar se os níveis de qualidade estabelecidos e esperados estão sendo atendidos é através de testes. Eles são uma parte fundamental na Garantia da Qualidade de Software, e servem para avaliar o comportamento de um software a partir de observações feitas sobre seu comportamento em determinadas condições (GALIN, 2004; IEEE, 2014).

Em relação à usabilidade, os testes são fundamentais para obter informação de como as pessoas utilizam o sistema e revelar problemas relacionados à interação delas com o aplicativo (HOLZINGER, 2005). Em se tratando de aplicativos para dispositivos móveis, segundo Zhang e Adipat (2005), os testes de usabilidade oferecem avaliações de terceiros sobre a facilidade com que os usuários visualizam o conteúdo e executam determinado aplicativo, e devem levar em consideração sua natureza e atributos de uso específicos.

Realizar testes de usabilidade é importante porque alcançar um alto nível de satisfação é fundamental para um aplicativo obter sucesso. Sem eles, fica mais difícil garantir que, principalmente da perspectiva do usuário, uma aplicação é prática, efetiva e fácil de usar (ZHANG, ADIPAT, 2005).

Nesse contexto, este trabalho realizou um estudo sistemático com o objetivo de investigar como os testes de usabilidade vêm sendo realizados em aplicativos móveis. Neste estudo foram mapeadas as técnicas e métodos reportados na realização dos testes, e os problemas e limitações encontrados.

1.1 Questões de Pesquisa

Visando atingir o objetivo deste estudo foi elaborada a seguinte questão de pesquisa principal (QPP):

- QPP: Como estão sendo realizados os testes de usabilidade em aplicativos móveis?

Para orientar a condução do estudo, bem como a extração, análise e síntese dos resultados, as seguintes questões de pesquisa específicas (QPE) foram definidas:

- QPE1: Quais são as práticas, métodos, processos e/ou técnicas reportadas do processo de testes de usabilidade no desenvolvimento de aplicativos móveis?
- QPE2: Quais são os problemas e limitações dessas práticas?

1.2 Objetivo

As contribuições e os resultados esperados desse trabalho são mostrar como estão sendo conduzidos os testes de usabilidade em aplicativos móveis e, a partir dos resultados obtidos, permitir que o planejamento para a execução deste tipo de testes evite problemas e limitações identificados no levantamento da literatura.

1.3 Estrutura do Trabalho

Este trabalho está estruturado da seguinte maneira:

- Capítulo 1 – Introdução: apresenta toda parte introdutória como base para o desenvolvimento do trabalho;
- Capítulo 2 – Referencial Teórico: apresenta o referencial teórico dos principais assuntos tratados neste trabalho;
- Capítulo 3 – Metodologia da Pesquisa: apresenta os métodos utilizados nesta pesquisa;
- Capítulo 4 – Resultados: apresenta os resultados alcançados no estudo da literatura através de uma revisão sistemática.
- Capítulo 5 – Conclusão: apresenta as considerações finais deste estudo suas limitações e recomendações para trabalhos futuros.
- Apêndices: apresentam os procedimentos, resultados e formulários relevantes para a pesquisa.

2 Referencial Teórico

Esse capítulo tem como objetivo descrever os principais conceitos abordados neste trabalho dando fundamento ao estudo realizado. Para isso é feita uma descrição das principais teorias relacionadas com este trabalho.

2.1 Testes

Teste de software é o processo desenvolvido para assegurar que um programa faz aquilo para o qual foi construído e nada sem intenção. Um software não deve se comportar de maneira inconsistente e imprevisível (MYERS, 2004). É uma atividade onde um sistema ou componente do sistema, é executado sob determinadas condições a fim de avaliar seu comportamento através dos resultados observados e registrados (IEEE, 2014).

Segundo Myers (2004), os testes de software são criados com o intuito de revelar a maior quantidade de erros quanto possível. Após a correção dos erros encontrados, os testes, ao serem refeitos, estabelecem certo grau de qualidade ao software. Sendo assim, encontrar e remover erros aumenta a confiabilidade do programa e torna o processo de testes um agregador de valor para o software (MYERS, 2004). Testes de software também têm a função de auxiliar a aprovação de um determinado trecho de código implementado ou a integração de código feita a um sistema, ao final de uma fase de desenvolvimento ou antes de uma entrega. Os testes de software são parte de um processo formal de Garantia da Qualidade de Software, cujo objetivo é revisar o estado corrente de funcionamento do software, baseando-se numa lista predefinida de casos de testes que representam uma variedade de cenários esperados (GALIN, 2004).

Os testes podem ser manuais ou automatizados e devem ser planejados e projetados de acordo com o processo de desenvolvimento do software (GALIN, 2004). De maneira simples, um caso de teste pode ser classificado como um conjunto de entradas e saídas esperadas. As saídas esperadas devem ser definidas durante a fase de *design* dos testes, antes do programa ser testado, a fim de eliminar qualquer viés no processo. A ideia é que estas saídas sejam pré-computadas a partir do entendimento dos requisitos do programa (NAIK, TRIPATHY, 2008). Se as saídas não forem pré-definidas existe a chance de que resultados

errôneos, porém plausíveis, sejam interpretados como corretos (GALIN, 2004).

Mesmo com o propósito de encontrar erros estabelecidos, o processo de testes e seus casos dificilmente cobrem todas as possibilidades destes erros acontecerem (MYERS, 2004; NAIK, TRIPATHY, 2008). Para a maioria dos sistemas, é quase impossível garantir a completude dos testes devido a: o domínio de entradas possíveis é grande demais para que todas sejam consideradas no processo; o *design* pode ser muito complexo para que todas as questões de projeto sejam cobertas; e pode não ser possível reproduzir todas as possibilidades do ambiente aos quais o software será submetido (NAIK, TRIPATHY, 2008).

De acordo com Myers (2004), apesar de ser uma atividade técnica, o processo de testes, para ser bem sucedido, deve levar em consideração também outras questões, que envolvem características econômicas e psicológicas. Na maioria dos casos, criar e executar casos de testes para todas as entradas e suas saídas possíveis de um programa é impraticável, pois mesmo um programa simples pode conter milhares dessas combinações. Testar todas as permutas de dados de um programa complexo levaria muito tempo e exigiria muitos recursos humanos para ser economicamente viável (MYERS, 2004). Sendo assim, os casos de testes devem ser elaborados levando em consideração apenas um subconjunto de entradas e saídas. Entretanto isso não significa que devemos reduzir seu número de forma arbitrária com o objetivo de economizar, pois essa não seria uma boa maneira de reduzir custos (NAIK, TRIPATHY, 2008). O objetivo principal é produzir software de baixo risco utilizando um número reduzido de casos de testes, tornado o processo eficaz. Assim, é com base neste objetivo que os casos devem ser projetados (MYERS, 2004).

É importante estabelecer o propósito certo do processo de testes para aqueles que irão testar as aplicações, pois entendê-lo pode ser decisivo para o seu sucesso. Segundo Myers (2004), nós somos orientados a objetivos e esses têm um efeito psicológico importante nas nossas atividades. O autor defende que se for estabelecido que o processo de testes deve demonstrar que um programa não tem erros, os testadores, subconscientemente, podem selecionar casos de testes com baixa chance de encontrar erros. Por outro lado, se os desenvolvedores veem o processo de testes como algo destrutivo, isto pode acabar por prejudicar o processo. Um teste bem sucedido é aquele que, ao ser executado, encontra erros que podem

ser concertados e que após ser novamente executado, e feitas as devidas correções, estabelece que não existem mais erros como os que foram encontrados.

O processo de testes foi uma das primeiras ferramentas da garantia da qualidade de software e inicialmente sua aplicação era restrita apenas aos estágios finais de desenvolvimento, pouco antes da entrega do programa. Com o tempo, a importância de detecção precoce de defeitos nos softwares aumentou o que motivou os profissionais a aplicar o processo de teste a partes do programa ainda em construção (GALIN, 2004). Considerando a importância que os testes têm nesse processo de garantia, pode-se dizer que o sucesso de um programa está intimamente ligado à qualidade dos testes ao qual foi submetido.

O processo de testes é constantemente relacionado com dois conceitos usados em desenvolvimento de software, que possuem seus próprios objetivos e conjuntos de atividades (NAIK, TRIPATHY, 2008):

- **Verificação:** São atividades que avaliam se um software satisfaz os requisitos estabelecidos nas especificações do seu projeto. São responsáveis por avaliar se o software foi construído corretamente.
- **Validação:** São testes executados a partir do ponto de vista dos usuários e têm como objetivo estabelecer o quanto um produto, em uso, atende as suas expectativas.

O processo de testes é um agregador de valor ao software, pois através dele é possível avaliar o comportamento de um programa encontrando o maior número possível de erros e verificando se as correções necessárias foram feitas, aumentando a sua confiabilidade. Além disso, é possível avaliar por meio de atividades de validação, como testes de usabilidade, o quanto um determinado software atende as expectativas do usuário, elevando suas chances de sucesso.

2.2 Usabilidade

A norma 9241-11 define usabilidade como sendo a “medida na qual um produto é usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002). Tais produtos são projetados para incorporar características e atributos de maneira que, em determinado contexto, beneficiem seus usuários (ABRAN et al., 2003). Em Radatz, Geraci e Katri (1990),

usabilidade é descrita como a facilidade com que o usuário aprende a operar um sistema, preparar suas entradas e interpretar suas saídas.

Um provedor de software precisa ser capaz de verificar o grau de confiança com que o produto atende ou não o nível de qualidade esperada (ABRAN et al., 2003). Devemos considerar usabilidade como um atributo fundamental durante a construção do software, para o alcance da qualidade (JURISTO, WINDL e CONSTANTINE, 2001). Usabilidade não é opcional, mas sim uma característica básica que determina a aceitação e a produtividade, que é pode ser descrita sendo a relação entre as tarefas realizadas com sucesso pelo usuário e os recursos investidos, como tempo e esforço. O termo se relaciona com um conjunto de múltiplos conceitos (tempo de execução, facilidade de aprendizagem, satisfação do usuário, etc.) que devem ser desenvolvidos juntos (ABRAN et al., 2003).

Para maior compreensão do que de fato se trata a usabilidade, é necessário compreender melhor os termos com as quais ela se relaciona em sua definição. Sendo assim a norma 9241-11 fornece as seguintes definições:

- **Eficácia:** Acurácia e completude com as quais usuários alcançam objetivos específicos.
- **Eficiência:** Recursos gastos em relação à acurácia e abrangência com as quais usuários atingem objetivos.
- **Satisfação:** Ausência do desconforto e atitudes positivas para com o uso de um produto
- **Contexto de uso:** Usuários, tarefas, equipamento e o ambiente físico e social no qual o produto é usado.

Planejar a usabilidade de um produto envolve identificar sistematicamente seus requisitos. É preciso determinar as medidas da usabilidade, extraídas a partir da decomposição da eficácia, eficiência e satisfação desejadas em atributos mensuráveis, para aferir o grau em que a esta atende esses requisitos e descrever o contexto de uso, de modo que esse também seja verificável. Essas medidas são particularmente importantes para visualizar as interações dentro do contexto (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002).

Para verificar se o nível de usabilidade requerido foi atingido é preciso mais do que apenas incorporar atributos. É necessário medir a satisfação e o desempenho dos usuários ao utilizarem o software. Esta medida está associada a

uma interação complexa entre usuário e contexto, o que pode gerar diferentes níveis de desempenho para um mesmo produto em diferentes contextos. A especificação e avaliação da usabilidade devem abordar os vários ambientes onde o software será utilizado (ABRAN et al., 2003). É importante notar que a usabilidade de um software varia de acordo com os *stakeholders*, e que características de contexto (usuário, tarefa, ambiente) podem ser determinantes, dado que qualquer alteração relevante pode alterar a usabilidade do produto (ABRAN et al., 2003).. Um produto feito para usuários experientes, por exemplo, pode não ser utilizável para iniciantes. Mudanças drásticas no ambiente podem trazer desconforto, ou inviabilizar funcionalidades, como por exemplo, barulho excessivo em softwares com respostas sonoras.

A especificação de contextos de uso adequados é uma parte fundamental na avaliação do grau de usabilidade de um produto, cujos níveis podem apresentar mudanças a partir de alterações feitas nas características do contexto no qual é usado. Além disso, a descrição dessas especificações é importante para comparar medidas de satisfação e desempenho entre soluções diferentes para um mesmo problema em contextos iguais (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002). Ainda segundo a norma, não existe regra geral para estabelecer medidas de eficácia, eficiência e satisfação e suas relações. A importância que cada um desses componentes representa e o modo como se relacionam dependem do contexto e do objetivo e é fundamental para a interpretação e análise dessas medidas que elas e todos os componentes estejam bem definidos.

A usabilidade é um atributo difícil de incorporar a qualquer sistema e requer conhecimento profundo das características e comportamento do usuário, algo que se torna ainda mais difícil pelo fato de que muitos desenvolvedores estão focados apenas nas questões técnicas do sistema (JURISTO, WINDL e CONSTANTINE, 2001). É importante assegurar que as interações planejadas do usuário com o sistema estão sendo observadas e são levadas em conta em todo o projeto de desenvolvimento, desde o início, pois a introdução de um alto grau de usabilidade em um software pode reduzir tempo de desenvolvimento, evitando que entregas sejam invalidadas causando retrabalho, aumentar vendas, melhorar produtividade dos usuários e reduzir custos de manutenção e suporte (JURISTO, WINDL e CONSTANTINE, 2001).

2.3 Aplicativos Móveis

Dispositivos móveis se tornaram uma parte essencial do cotidiano da vida de grande parte das pessoas como o meio mais efetivo e rápido de comunicação, acesso à informação e serviços, em qualquer lugar e a qualquer hora (DINH et al., 2013). Usuários são capazes de interagir com inúmeras funcionalidades a fim de executar tarefas que tornem suas vidas mais fáceis e seus dias mais produtivos. Sistemas móveis realizam operações numa ampla gama de áreas, desde as mais tradicionais como voz e mensagens de texto até aplicações com funcionalidades mais complexas como serviços baseados em localização ou realidade aumentada por exemplo (CORRAL, SILLITTI e SUCCI, 2015).

Devido à popularização dos dispositivos e suas aplicações, o ambiente móvel se tornou uma importante plataforma para o desenvolvimento de produtos de software e serviços (CORRAL, SILLITTI e SUCCI, 2015). O número de softwares voltados para esses dispositivos cresce à medida que se tornam mais populares. A sua complexidade também aumenta, conforme a capacidade computacional e a dos meios de comunicação. Sendo assim os ambientes móveis se tornaram uma área de desenvolvimento importante para organizações dos mais diversos setores que têm como objetivo levar seus produtos e serviços para mais perto de seus usuários (DINH et al., 2013).

Sendo uma das áreas tecnológicas de maior crescimento, e ainda tendo muito a ser explorado, os aplicativos móveis atraem muitos desenvolvedores fazendo o setor ser bastante competitivo. Isso acabou elevando o interesse em conhecer mais detalhadamente as etapas de desenvolvimento do ponto de vista do produto e do processo (CORRAL, SILLITTI e SUCCI, 2015).

Entretanto, frente a esse crescimento rápido, os aplicativos móveis enfrentam muitos desafios relacionados à limitação dos dispositivos e dos ambientes móveis. São questões relacionadas à segurança, mobilidade, e restrições na oferta de recursos dos dispositivos como processamento, bateria, memória e capacidade de armazenamento. Tais limitações podem afetar consideravelmente a qualidade dos serviços nas áreas de comércio eletrônico, educação, saúde, finanças, entretenimento, entre outras.

Com objetivo de minimizar essas limitações, é utilizada uma arquitetura onde a maior parte da carga de processamento e armazenamento é transferida para

serviços de computação em nuvem, consumindo menos os recursos do dispositivo, melhorando a experiência do usuário e tornando possível a entrega de serviços e funcionalidades mais robustas e complexas. Esse paradigma é amplamente usado pelas organizações e consolidou os dispositivos móveis como sendo o principal meio de acesso dos usuários a serviços disponíveis nos aplicativos (DINH et al, 2013; LOMOTEY e DETERS, 2014).

Ainda segundo Lomotey e Deters (2014), o mais comum é que os usuários queiram acessar suas informações e serviços independentes do local, da hora e do dispositivo. Esse paradigma onde um usuário está conectado, de forma sincronizada e sem perda de qualidade, a um mesmo serviço através de múltiplos dispositivos é chamado de computação de nuvem ubíqua e tem impulsionado ainda mais a exigência de qualidade dos aplicativos móveis, pois estes são o meio de acesso principal dos usuários finais dessa arquitetura (LOMOTEY e DETERS, 2014).

Embora essas abordagens tenham auxiliado a expandir a capacidade de entrega dos dispositivos móveis e tornado possível o aumento da oferta de serviços complexos por meio dos aplicativos, os ambientes móveis ainda são limitados e seus desenvolvedores devem levar esse fato em consideração. Segundo Corral, Sillitti e Succi (2015), para produzir um aplicativo de sucesso eles precisam, sob essas condições, criar produtos com disponibilidade, eficiência e capacidade de resposta e que atendam os regulamentos dos canais de distribuição e satisfaçam as expectativas dos usuários.

Dentre os canais de distribuição, os que mais se destacam são os mercados de aplicações móveis, onde é possível tornar o aplicativo disponível para download diretamente para os milhões de usuários do mundo todo. Esses canais hospedam milhares de aplicativos que devem seguir políticas de qualidade estabelecidas referentes a conteúdo, funcionalidade, usabilidade e muitas outras, a fim de terem suas submissões de aplicativos aprovadas, e os mesmos publicados (CORRAL, SILLITTI e SUCCI, 2015). Além desses padrões de qualidade estabelecidos, o aplicativo também é avaliado no mercado por *feedbacks* e pontuações dadas pelos usuários.

Sendo assim, se faz necessário garantir que os aplicativos apresentem, não só as funcionalidades desejadas e solicitadas pelos usuários, mas também níveis adequados de usabilidade que atendam ou superem suas expectativas. Por ser um

dos principais processos para garantir qualidade, os testes desempenham papel importante na análise da usabilidade de aplicativos móveis, pois é fundamental para que as organizações obtenham informações junto aos potenciais usuários sobre eficiência, eficácia e satisfação, o que possibilita a correção de erros e a melhoria da experiência deles ao usar o aplicativo.

2.4 Síntese do Capítulo

Neste capítulo foram apresentados os conceitos e fundamentos teóricos relacionados a testes, usabilidade e aplicativos móveis a serem usados como base para este trabalho. Ao longo do capítulo, foram abordados a importância de testes e usabilidade, além da contextualização dos ambientes e aplicativos móveis e o modo com estes se relacionam.

3 Metodologia de Pesquisa

Este capítulo descreve as técnicas de revisão sistemática, método utilizado nesta pesquisa, e suas etapas conforme especificado na literatura. Esta técnica surgiu na medicina, devido à necessidade de ter evidências que dessem suporte aos estudos conduzidos. Em engenharia de software esse método é utilizado com o mesmo objetivo. Com a finalidade de responder uma questão de pesquisa, explorar um tópico de uma área ou investigar um fenômeno, a revisão sistemática procura identificar, avaliar e interpretar pesquisas relacionadas. Dentre as muitas motivações possíveis ela pode servir para sumarizar evidências, identificar lacunas nas demais pesquisas a fim de identificar necessidades de pesquisas futuras relacionadas, embasar teorias ou auxiliar na geração de novas (KITCHENHAM e CHARTERS, 2007).

Ainda segundo Kitchenham e Charters (2007), uma revisão sistemática deve sintetizar de maneira imparcial, os trabalhos já realizados, e isso deve ser notório no processo e nos resultados, pois além de ser um trabalho de exploração e síntese importante nas áreas onde é aplicada, muitas das revisões são base para outras pesquisas. As revisões permitem que, ao se basear em trabalhos anteriores, estudos possam produzir resultados e análises que impulsionam uma área do conhecimento a evoluir.

Uma revisão sistemática possui algumas características específicas que visam tornar o processo justo e confiável a fim de que ele tenha valor científico. Essas especificidades diferenciam esse tipo de revisão literária de outros tipos mais *ad-hoc*. Algumas delas são (KITCHENHAM e CHARTERS, 2007):

- Deve ser definido um protocolo que especifique as questões de pesquisa e os métodos a serem utilizados;
- Possui uma estratégia de pesquisa que visa encontrar trabalhos relevantes relacionados;
- O processo é documentado para que possa ser possível sua reprodução e para ter seu rigor e completude avaliados.
- Possui critérios de inclusão e exclusão explícitos a fim de encontrar estudos mais relevantes;
- Estabelece critérios de qualidade para avaliar os estudos durante o processo.
- Especifica as informações a serem obtidas de cada estudo

- Funciona como pré-requisito para meta-análises quantitativas.

A revisão sistemática pode ser dividida nas seguintes fases:

- Planejamento: Nessa fase é identificada a necessidade do trabalho, é feita a especificação da(s) questão(ões), é definida a estratégia e um protocolo de pesquisa é desenvolvido e avaliado.
- Condução: A estratégia é aplicada e todas as informações referentes aos artigos encontrados (ano, autor, título, fonte de busca e etc.) são documentadas. São aplicados os critérios de inclusão e exclusão nos estudos, da maneira estabelecida no protocolo, e em seguida os de qualidade. Após isso os dados relevantes para pesquisa são extraídos e sintetizados.
- Apresentação dos resultados: É construído um relatório que descreve todas as especificações do processo e as informações geradas em todas as atividades, assim como o resultado alcançado a partir da revisão.

Durante a fase de planejamento é preciso se certificar das razões pelas quais a revisão está sendo conduzida. A necessidade da revisão surge com o desejo de obter conclusões mais generalistas sobre uma área ou fenômeno ou apoiar pesquisas futuras (KITCHENHAM e CHARTERS, 2007). Essa fase possui como parte mais importante a definição das questões de pesquisa que a revisão busca responder e que guiam a elaboração do protocolo de pesquisa (incluindo a estratégia de busca), bem como a extração de dados e sua análise.

O protocolo de pesquisa especifica todas as etapas e métodos da revisão e deve ser definido com antecedência de modo a evitar vieses durante o trabalho, pois um pesquisador poderia conduzir a revisão de modo a atender suas expectativas, direcionando os resultados e diminuindo assim seu valor científico. Neste protocolo os itens fundamentais encontrados são:

- A razão da pesquisa;
- A(s) questão(ões) de pesquisa;
- A estratégia a ser usada, que indica os termos de pesquisa, os métodos e as fontes de busca usadas durante o trabalho;
- Os critérios de seleção e o modo como eles são aplicados;

- Os critérios de avaliação da qualidade e como são aplicados;
- A estratégia para extração e síntese dos dados.

O protocolo é uma parte crítica de qualquer revisão e é importante que seja avaliado por especialistas a fim de encontrar falhas a serem corrigidas, e receber sugestões que melhorem a qualidade do trabalho. A consistência do trabalho pode ser verificada, dentre outros fatores, pela derivação adequada dos termos de busca a partir das questões de pesquisa, examinando a relação dos dados extraídos com a(s) questão(ões) de pesquisa e se a análise dos dados é apropriada para respondê-las (KITCHENHAM e CHARTERS, 2007).

O processo de revisão é iterativo e boa parte das atividades da fase de desenvolvimento podem ser revistas e ter parte da estratégia refinada à medida que ele avança para torná-lo mais eficiente na captura e filtragem dos estudos. Além disso, os critérios de coleta e síntese dos dados, projetados no início, podem ser melhorados conforme os critérios de qualidade forem sendo aplicados e os dados coletados.

3.1 Protocolo de Estudo Sistemático

Nesta seção é apresentado, de forma resumida, o protocolo utilizado no desenvolvimento deste trabalho. Ele segue as especificações contidas em (KITCHENHAM e CHARTERS, 2007), além de adotar estratégia para o processamento dos dados contidos em (DYBA e DINGOYR, 2007). A versão completa do protocolo se encontra no APÊNDICE A.

3.1.1 Questões de Pesquisa

Visando atingir o objetivo deste estudo foi elaborada a seguinte questão de pesquisa principal (QPP):

- QPP: Como estão sendo realizados os testes de usabilidade em aplicativos móveis?

Para orientar a condução do estudo, bem como a extração, análise e síntese dos resultados, as seguintes questões de pesquisa específicas (QPE) foram definidas:

- QPE1: Quais são as práticas, métodos, processos e/ou técnicas reportadas

do processo de testes de usabilidade no desenvolvimento de aplicativos móveis?

- QPE2: Quais são os problemas e limitações dessas práticas?

3.1.2 Estratégia de Busca

Com o objetivo de encontrar o maior número de artigos relacionados ao tema foram extraídos os termos: Teste, usabilidade e aplicativos móveis. Esses termos foram traduzidos para o inglês, por ser o principal idioma usado para buscas nas bases escolhidas (ACM Library, IEEE Xplore e Scopus), e identificados termos relacionados e sinônimos que foram concatenados utilizando o operador AND para que no final fossem unidos pelo operador booleano OR. O resultado foi a seguinte *string* de pesquisa:

Quadro 1: *String* de Busca

<i>String</i> de Busca
(test OR testing) AND (usability) AND ("mobile software" OR "mobile application" OR "mobile app")

Fonte: Próprio Autor

3.1.3 Seleção dos estudos

Na fase 1, foi feita uma busca nas bases escolhidas utilizando a *string* de busca e os metadados dos estudos achados foram exportados e organizados em planilhas do Google *Spreadsheets* armazenadas no Google Drive, sendo atribuído um identificador único para cada um deles.

Após isso foi realizada a fase 2, onde foram lidos o título, resumo e palavras-chave e aplicados os critérios de inclusão e exclusão. Os artigos aprovados, depois de submetidos aos critérios de seleção, foram registrados utilizando a mesma ferramenta da fase anterior.

Na fase 3, os artigos foram baixados e em seguida, armazenados usando o Google Drive. A introdução, e conclusão dos artigos foram lidas e os critérios de inclusão e exclusão são novamente aplicados. Os resultados foram registrados nas planilhas do Google e os aprovados passam para a próxima fase.

A fase 4 foi a última fase da seleção e nela os artigos aprovados da fase anterior foram lidos na íntegra. Os critérios de seleção foram aplicados uma última vez e os resultados foram registrados de forma semelhante nas fases anteriores. Os

artigos resultantes dessa fase foram usados no prosseguimento do trabalho.

3.1.4 Avaliação da Qualidade

Em seguida uma avaliação foi feita usando a escala de 3 valores proposta em Likert (1932) onde são atribuídas respostas gradativas para avaliar o quanto o estudo está de acordo com cada um dos critérios de qualidade.

- Não atende (0): deve ser atribuído caso não exista nada no trabalho que atenda ao critério avaliado;
- Dúbio (0.5): deve ser atribuído na falta de clareza do critério avaliado;
- Atende (1): deve ser atribuído se o critério avaliado é atendido;

Os artigos foram classificados em 4 faixas de qualidade, observando os critérios do quadro 2, de acordo com o somatório alcançado pela a avaliação dos itens de avaliação de qualidade:

Quadro 2: Distribuição de pontuação por classificação

Baixa	Média	Alta	Muito Alta
$0 \leq N \leq 3$	$3 < N \leq 7$	$7 < N \leq 8,5$	$8,5 < N \leq 10$

Fonte: Próprio Autor

3.1.5 Extração, Análise e Síntese dos dados

Depois de avaliados, os estudos aprovados na última fase tiveram os dados relevantes, como referências, contexto e evidências, extraídos com o objetivo de responder as questões de pesquisa. Após isso, foram extraídos os dados e evidências foram extraídas e codificadas. Os códigos foram identificados e traduzidos em temas, que foram interpretados e explorados, buscando estabelecer temas superiores e suas relações (DYBA e DINGOYR, 2007).

3.2 Síntese do Capítulo

Este capítulo apresentou o ciclo da pesquisa e os passos conduzidos para a construção deste trabalho com o objetivo de tornar os resultados mais confiáveis, auditáveis e reproduzíveis. O capítulo a seguir apresenta os resultados obtidos a partir deste processo.

4 Resultados

Este capítulo apresenta o resultado do estudo sistemático feito na literatura com base em uma revisão sistemática, seguindo processo definido no capítulo anterior. Todas as etapas do processo são descritas e os principais pontos de cada uma delas são destacados. Os dados resultantes de cada uma das etapas, e do processo como um todo, são analisados com o objetivo de responder às questões de pesquisa.

A busca dos artigos foi realizada apenas pelo autor do trabalho, o que segundo, pode acarretar a possibilidade de vieses na pesquisa. Com o intuito de diminuir essa possibilidade, foram conduzidos reavaliações de amostras dos estudos selecionados e excluídos em cada fase, para verificar a consistência da pesquisa (KITCHENHAM e CHARTERS, 2007).

4.1 Execução

A revisão foi realizada conforme o protocolo resumido apresentado no capítulo 3, disponível por completo no APÊNDICE A. As próximas seções descrevem o desenvolvimento de cada etapa e os seus resultados, bem como uma análise de cada um delas.

4.1.1 Busca

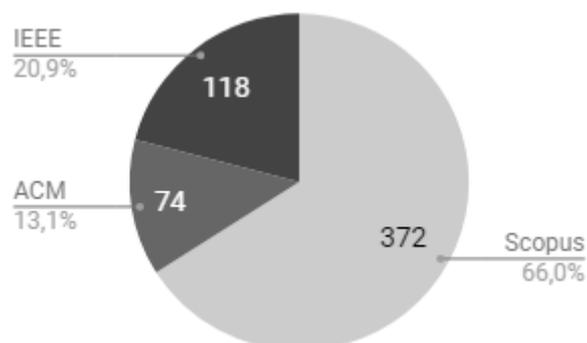
Após serem definidas as questões de pesquisa e a partir delas extraídos os termos, com identificação de sinônimos e termos relacionados, a *string* de busca foi desenvolvida e inseridas de forma manual nas bases digitais selecionadas para obtenção dos estudos que seriam submetidos à fase 2.

Os dados dos estudos (Ex. título, ano, autores, palavras-chave, resumo, etc.) encontrados a partir da *string* de busca foram exportados em arquivos com formato Comma Separated Values (.csv) usando funcionalidades oferecidas pelas próprias bases. Os arquivos foram separados em planilhas do Google *Spreadsheets* presentes na plataforma Google Drive para melhor organização e acesso. A cada artigo foi atribuído um identificador único, usado para que referências mais precisas fossem criadas.

A partir das buscas primárias realizadas nas bases digitais um total de 566 estudos foram obtidos, somados aos valores das três bases escolhidas. A

porcentagem e a quantidade de artigos de cada umas das bases estão na Figura 1.

Figura 1: Resultado da busca feita nas bases selecionadas

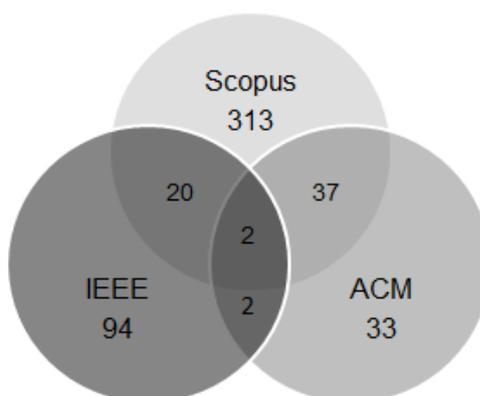


Fonte: Próprio Autor

Conforme pode ser observado, a base digital Scopus retornou mais que o triplo de artigos da base IEEE Xplore, e mais que o quádruplo de artigos da base ACM Digital Library. Isso se deve ao fato de que a base Scopus indexa resultados de outras bases e de diversas revistas e conferências do mundo inteiro o que a torna bastante diversificada.

Além disso, a Scopus inclui alguns resultados das outras duas bases escolhidas, que também apresentam resultados semelhantes entre si, conforme observado na Figura 2. As duplicatas foram identificadas usando informações exportadas das bases e monitoradas em cada fase, sendo eliminadas sempre que essa condição era confirmada.

Figura 2: Interseção dos resultados da busca feita nas bases selecionadas

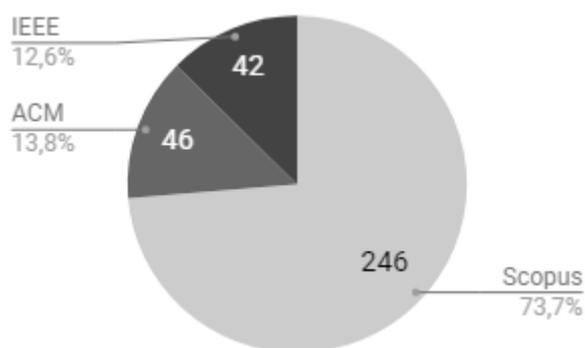


Fonte: Próprio Autor

4.1.2 Seleção por título, resumo e palavras-chave

Após a organização dos dados exportados e identificação de cada estudo, foi efetuada a leitura de cada título, resumo e palavras-chave a fim de determinar se o artigo em questão passaria para a próxima fase. Para isso essas informações foram submetidas aos critérios de inclusão e exclusão, sendo prioritariamente eliminados os artigos que não se encaixavam na área tratada por este trabalho. O resultado dessa etapa pode ser visualizado na Figura 3.

Figura 3: Resultado da fase 2 da seleção de estudos



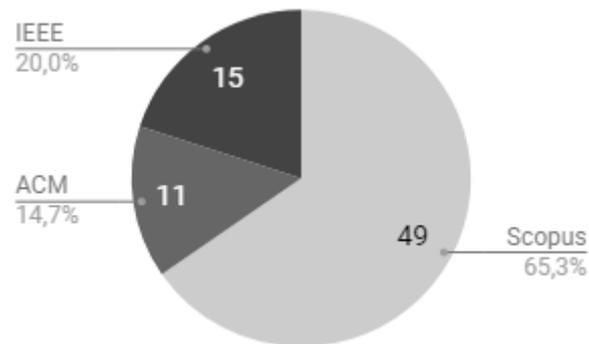
Fonte: Próprio Autor

Os artigos cuja aprovação ou rejeição não ficou clara foi aprovado para ser melhor avaliado na próxima fase. Sendo assim o total de aprovados foi de 332 artigos.

4.1.3 Seleção dos estudos pela introdução e conclusão

Os 332 artigos aprovados na fase anterior tiveram seus dados registrados e o download foi feito. Os artigos cujo download não foi possível ser realizado gratuitamente nos ambientes do CIn/UFPE, foram eliminados conforme critério estabelecido no APÊNDICE A. Foram eliminados neste critério 122 artigos, todos indexados pela biblioteca Scopus, o que representa 36% dos aprovados para essa fase e 21% do total das buscas. Para os artigos que tiveram o download feito, foi realizada a leitura da introdução e da conclusão para que eles fossem submetidos aos critérios de inclusão e exclusão, novamente aplicando a regra de aprovar os artigos cuja leitura levantou dúvidas sobre sua exclusão ou não. A quantidade de artigos aprovados nesta etapa se encontra, dividida por base, na Figura 4.

Figura 4: Resultado da fase 3 da seleção de estudos

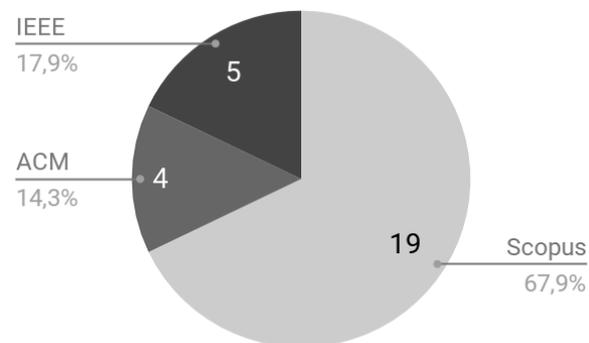


Fonte: Próprio Autor

4.1.4 Seleção pela leitura completa dos artigos

Nesta fase foi realizada a leitura completa dos artigos aprovados na fase anterior e aplicados os critérios de seleção a fim de obter o conjunto final de artigos a serem utilizados para responder as questões de pesquisa, cuja lista completa pode ser visualizada no APÊNDICE B. Nessa fase foi aprovado um total de 28 artigos, distribuídos entre as bases conforme a Figura 5.

Figura 5: Resultado da fase 4 da seleção de estudos



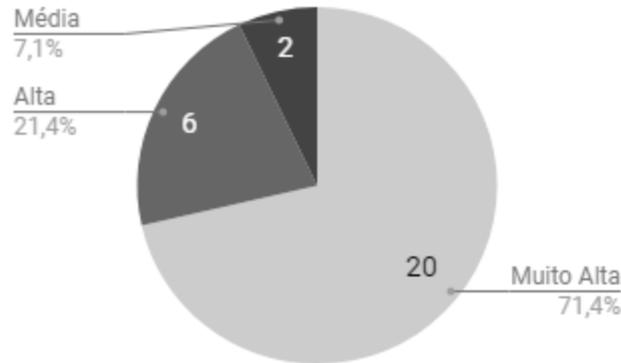
Fonte: Próprio Autor

4.1.5 Avaliação da Qualidade

Com a finalização das etapas de seleção dos estudos, os artigos restantes foram avaliados segundo os critérios de qualidade estabelecidos no APÊNDICE A, aos quais atribuiu-se valores para representar em que grau determinado quesito era satisfeito pelo estudo. A pontuação final foi dada com base no somatório de todos os quesitos e, assim como a pontuação individual por critério de qualidade, encontra-se no APÊNDICE C. A distribuição da avaliação final dos estudos por nível de qualidade pode ser observada na Figura 6, e nela é possível notar que os trabalhos

resultantes foram bem avaliados na maioria dos quesitos.

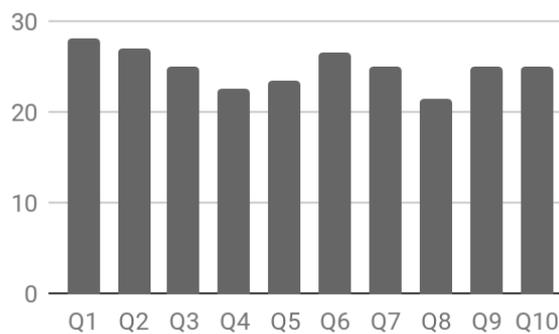
Figura 6: Distribuição dos estudos por nível de qualidade



Fonte: Próprio Autor

A Figura 7 apresenta o total de pontos obtidos pelos 28 artigos por quesito. Os quesitos avaliam a clareza dos objetivos da pesquisa, da estratégia adotada, da descrição da tecnologia avaliada e dos métodos de coleta de dados assim como do relato dos resultados. Além disso, avaliam a descrição dos testes e suas variáveis, se o trabalho é bem referenciado, se os objetivos foram alcançados e o valor da contribuição para academia ou indústria. Enquanto o quesito 1 (Q1- Existe uma clara declaração dos objetivos da pesquisa?) obteve nota máxima em todos eles, que indica que todos apresentaram os contextos nos quais estão inseridos de e os objetivos da pesquisa de maneira satisfatória, através do gráfico é possível observar que os quesitos Q4 e Q8 obtiveram as pontuações mais baixas, o que indica que em boa parte das vezes ficou dúvida a maneira como aquele quesito foi atendido, ou que ele não foi atendido por completo.

Figura 7: Distribuição dos estudos por faixa de qualidade



Fonte: Próprio Autor

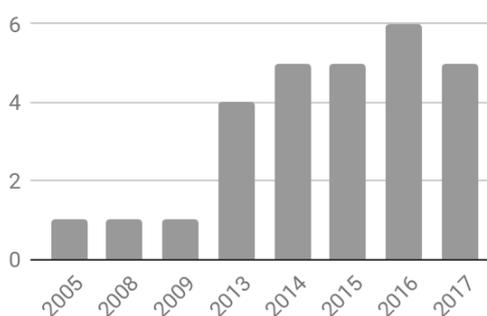
A pontuação atingida no quesito 4 (Q4 – Os métodos de coleta de dados foram adequadamente descritos?), que tem como objetivo avaliar a forma de coleta de dados, indica que muitos estudos não reportaram de maneira clara os dados resultantes dos testes, se limitando apenas a discutir os resultados, sem apresentar evidências que os comprovem.

A menor pontuação do item correspondente à descrição clara do contexto em que os testes foram realizados (Q8 – Os participantes ou unidades de observação estão adequadamente descritos?) indica que os estudos não descreveram bem o ambiente, os usuários, ou as plataformas usadas para realização dos testes. Isso é particularmente importante em testes de usabilidade, cuja medida varia conforme se altera qualquer uma dessas características. Além disso, fica mais difícil replicar os testes ou realizar testes de soluções em contextos semelhantes para comparação.

4.1.6 Visão Geral dos Estudos

Depois de concluídas as fases de seleção, qualificação e extração dos dados dos artigos aprovados, verificou-se que nos últimos anos o número de artigos que reportaram testes de usabilidade feitos em aplicativos móveis permaneceu constante, o que pode ser visualizado na Figura 8. Isso demonstra que junto com o aumento do desenvolvimento de aplicativos móveis, houve a consolidação do processo e da sua importância no desenvolvimento destes.

Figura 8: Distribuição das publicações selecionadas ao longo dos anos

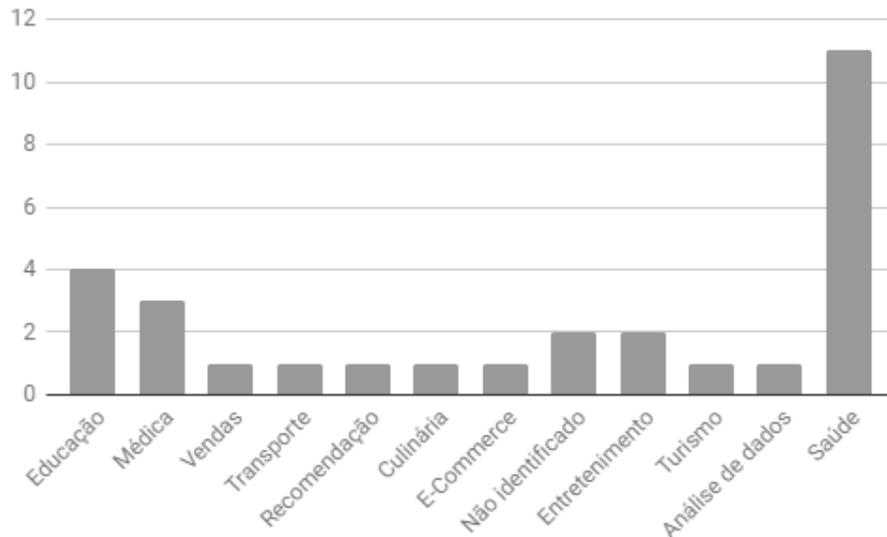


Fonte: Próprio Autor

A Figura 9 apresenta as áreas de aplicação em que os testes de usabilidade têm sido realizados. A área de cuidado com a saúde, representada no gráfico apenas como Saúde, teve mais aplicativos testados por ser uma área crítica, relacionada muitas vezes a mudanças de hábito. Para estes, a usabilidade é fundamental para que o aplicativo seja incorporado à rotina e tenha impactos

positivos na qualidade de vida das pessoas. É importante destacar que a variação de áreas mostra que a preocupação com a usabilidade está presente em diversos domínios.

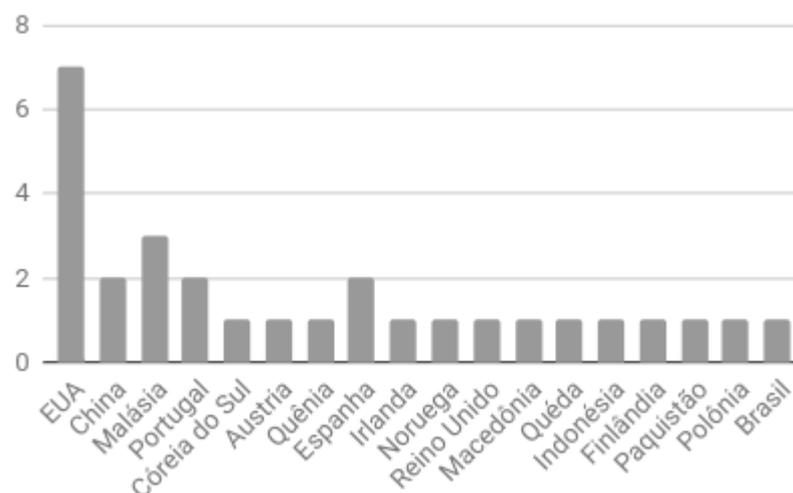
Figura 9: Distribuição por área dos aplicativos testados nos estudos selecionados



Fonte: Próprio Autor

Os relatos de testes de usabilidade e estudos relacionados têm autorias distribuídas pelos países de acordo com a Figura 10. Apesar dos Estados Unidos liderarem no ranking, as pesquisas se encontram muito bem distribuídas pelos continentes, em diversos países, incluindo o Brasil, o que reafirma a relevância desses testes em ambientes móveis.

Figura 10. Distribuição de autoria por países



Fonte: Próprio Autor

4.2 Mapeamento das evidências

Nesta seção são apresentados os resultados do estudo sistemático feito utilizando técnicas de mapeamento sistemático de acordo com as questões de pesquisa.

QPE1: Quais são as práticas, métodos, processos e/ou técnicas reportadas do processo de testes de usabilidade no desenvolvimento de aplicativos móveis?

A tabela 1 apresenta as práticas, métodos, processos e técnicas levantadas a partir dos estudos resultantes. De acordo com o levantamento, os pesquisadores utilizam 11 técnicas para realizar os testes e obter resultados.

Dentre as técnicas relacionadas aos testes de usabilidade realizados, é possível verificar que a maioria dos estudos inclui questionários para avaliar a interação dos aplicativos com os usuários. Alguns desses questionários se baseiam em modelos que utilizam algumas escalas numéricas definidas para avaliar o grau de satisfação dos usuários. Os artigos [A5, A6, A18 e A19] utilizaram o System Usability Scale (SUS) para aplicar questionários, os artigos [A17 e A4] utilizaram o Software Usability Measurement Inventory (SUMI), o artigo [A20] usou o QUIS Questionnaire for User Interaction Satisfaction (QUIS) e o [A28] utilizou o Usefulness, Satisfaction and Ease of use (USE). A técnica de ThinkAloud¹ foi usada em mais de um terço dos estudos, enquanto que quase um quarto se utiliza de gravação de áudio e vídeo para observações e registro dos testes.

¹ Essa técnica consiste em solicitar aos usuários que verbalizem continuamente seus pensamentos durante o uso do software, permitindo que aos testadores compreendam como eles visualizam o sistema, facilitando a identificação dos problemas de usabilidade e suas razões (HOLZINGER, 2005).

Tabela 1: Artigos X Técnica/Método/Processo/Práticas

	Think Aloud	Questionário	Não Identificado	Gravação de Vídeo	Gravação de Áudio	Captura de logs	Eye Tracker	Captura de expressões	Deteção de Contexto	Entrevista	Gravação de telas	Total
A1		X										1
A2		X		X	X		X					4
A3			X									1
A4		X										1
A5	X	X		X	X					X		5
A6	X	X										2
A7		X										1
A8		X									X	2
A9		X										1
A10			X							X		2
A11	X	X										2
A12				X	X							2
A13	X	X		X							X	4
A14		X			X				X	X		4
A15	X				X				X			3
A16									X			1
A17		X							X			2
A18		X		X								2
A19									X			1
A20		X		X								2
A21	X	X										2
A22	X											1
A23	X					X						2
A24				X	X							2
A25	X			X	X							3
A26						X						1
A27						X		X	X			3
A28	X	X										2
Total	10	16	2	8	7	3	1	1	1	6	3	

Fonte: Próprio Autor

QPE2: Quais são os problemas e limitações dessas práticas?

Após os estudos terem sido selecionados e processados, foram submetidos à extração de dados, que tinha como objetivo, além de revelar como eram feitos os testes de usabilidade, encontrar as principais questões reportadas ligadas a esse

processo. Estas questões são relatadas a seguir, fundamentadas pelas transcrições de afirmativas extraídas do estudo:

Dificuldade de generalização dos resultados dos testes:

O artigo [A2] relata problemas em relação a grupos de usuários com número reduzido de participantes, que impediria uma generalização dos resultados, e a dificuldade de estabelecer um grupo apropriado para os testes.

Os artigos [A3 e A19] apontaram que a falta de diversidade entre os usuários, como a falta de mulheres e outras características que não correspondem à variedade do público alvo, interferem nas conclusões sobre os testes e podem prejudicar a validade externa dos resultados. Além disso, [A17] apontou que por ser realizado com apenas um grupo de um determinado local, o teste não seria suficiente para generalização.

“Moreover, a series of problems were detected in the experiment. The main problem was to obtain the proper participant group” - [A2]

“There are few limitations in this research. First of all there were no female participants in this research. Therefore these results cannot be generalized for all illiterate population of KPK.” - [A3]

“Given that testing was undertaken with a group of adolescents attending a single urban hospital for weight management, the results cannot be generalized.” - [A17]

“Participants were mostly of high education status, which limits the external validity of the study. It would be useful to include a more diverse and representative sample in future research, and also a more homogenous clinical sample.” - [A19]

Grande esforço necessário para execução dos testes

O estudo [A2] aponta que testes de usabilidade podem ser bastante trabalhosos, já que exige a captação e registro de uma quantidade grande de informações, muitas delas subjetivas, que devem ser analisadas em conjunto para identificar falhas na usabilidade desses aplicativos.

“Usability testing is quite laborious. In the present case, the usability study for 4 scenarios cost 14 man-days.”- [A2]

Execução de testes sem considerar a experiência prévia dos participantes:

Outra limitação apontada pelo estudo [A3] é que não foi considerada a experiência prévia dos participantes, o que pode influenciar nos resultados.

“This research also did not considered other experience factors which may become handy in such education related scenarios. For example, experience with mobile and computers could affect the interaction of participants with mobile application which is not considered the provided scenario.”- [A3]

Testes conduzidos usando apenas uma plataforma:

Os artigos [A4] e [A12] apontam que é desejável que a solução apresentada seja testada em mais de uma plataforma dada à heterogeneidade das plataformas e dos dispositivos móveis. Embora o estudo [A12] assuma que os testes foram feitos em uma aplicação baseada em no sistema Android, é importante destacar que nesta plataforma os sistemas também apresentam variações.

“The use of different mobile devices during the test also should be considered.”- [A4]

“Regarding the limitations of this study, we only developed and tested an Android-based application; thus, we did not verify the differences in usability across different device operation systems.” - [A12]

Técnicas de testes que podem interferir no comportamento dos usuários durante os testes:

Também foi mencionado um possível problema referente à técnica de *ThinkAloud*, uma das mais usadas segundo os dados levantados, segundo o estudo [A13]. De acordo com este estudo, esta técnica, assim como outras que tem como base interação com avaliadores durante os testes, poderia interferir nas reações dos usuários durante os testes. O estudo também aponta que utilizar protótipos, ao invés do programa real, limitaria a possibilidade de testes de funcionalidades.

“One major limitation of this study is the possibility that participants were influenced by the study situation itself, since the task of thinking aloud and simultaneously being observed might have provoked unintended reactions or statements” - A[13]

“Another limitation is that participants tested a prototype rather than the actual program, so they could not use or test many of the program functionalities” - [A13]

Limitações da técnica de Captura de Log:

Os artigos [A23, A26 e A27] propõem técnicas de testes que utilizam a captura de log para análise de alguns atributos da usabilidade com o objetivo de automatizá-los. Entretanto o artigo [A23] destaca que, embora tenha êxito em grande parte das questões, essa abordagem possui limitações que devem ser observadas e tratadas de outras formas. Em [A27] a abordagem é combinada com outras formas de análise (emoção e contexto) para suplantiar essas limitações. A expressão facial dos usuários é monitorada, usando a câmera frontal do dispositivo, para identificar emoções positivas e negativas. Além disso, é feita a captura de informações do contexto em que o usuário de encontra (clima, localização e *status* de mobilidade e de conexão).

Essas informações são sistematicamente ligadas aos eventos de interação do usuário com o aplicativo testado para geração automatizada de gráficos e relatórios para análise. Entretanto, segundo o artigo, sua técnica ainda precisa ser comparada a outros métodos a fim de validar seus benefícios.

“However, the shortcoming of the event logging method is, for most of the issues it identified, that although it can point out the location of a problem, it cannot tell the cause of that problem. This is an issue common to all event logging methods because they lack the information that can be gleaned from listening to participants’ verbalized thoughts and observing participants’ facial expressions and other nonverbal signals.” - [A23]

“Additionally, this technique will be compared to other usability methodologies to validate the benefit of our approach.” - [A27]

Número limitado de sessões:

O número de sessões foi apontado como limitação pelo estudo [A16]. Assim como boa parte dos estudos, é relatada apenas uma sessão de testes de usabilidade, o que segundo ele seria insuficiente para encontrar todos os problemas.

“A potential limitation of the study is that the usability of Painometer was tested in a single session. Some problems might arise when the app is used for a period of time; to address this potential issue, an e-mail address is provided within the app description.” - [A16]

4.3 Síntese do Capítulo

Este capítulo descreveu os resultados do estudo sistemático realizado de acordo com a aplicação dos métodos de busca e coleta já mencionados. Esse estudo foi realizado com o objetivo de verificar como os testes de usabilidade vêm sendo executados e relatados em aplicativos móveis, bem como os desafios e limitações observadas relacionadas a estes testes.

Foram descritas as áreas de desenvolvimento dos aplicativos e testes dos artigos resultantes, a distribuição das publicações ao longo dos anos, assim como os países de origem das pesquisas.

Também foram identificadas, agrupadas e sumarizadas em tópicos, a partir de trechos retirados dos estudos, as principais limitações e problemas apontados

pelos pesquisadores em relação aos testes de usabilidade realizados.

É importante que na fase de planejamento dos testes sejam escolhidos questionários de avaliação já estabelecidos, para evitar eventuais vieses nos resultados. Um questionário elaborado pela equipe desenvolvedora pode deixar de avaliar algum aspecto, mesmo que de forma inconsciente, que sabidamente não é bem atendido pela aplicação. Além disso, estes questionários estabelecidos apresentam escalas numéricas que facilitam as medições e o acompanhamento da evolução da usabilidade durante o desenvolvimento.

Para minimizar o problema de generalização dos resultados, se faz necessário desenvolver estratégias de captação levando em consideração os diferentes tipos de usuário. É preciso buscar usuários que frequentem ambientes onde o aplicativo será utilizado, ou que enfrentem situações reais onde o uso poderia ser benéfico. Os testes de usabilidade podem ser custosos e demorados e se não forem adequadamente considerados na fase de planejamento, podem representar ameaças ao cronograma e ao custo do projeto.

5 Conclusão

Baseado nas descobertas da revisão sistemática feita neste trabalho foi apresentado um estudo de como são realizados os testes de usabilidade de aplicativos móveis reportados na literatura. Foram analisados e filtrados 566 estudos de três bases de dados, resultando em 28 artigos usados para responder às questões de pesquisa.

Foram mapeados os métodos, processos e técnicas usadas nos testes onde se constatou que 57% dos estudos utilizam questionários como forma de avaliação, mas apenas metade destes relatou usar questionários padronizados de usabilidade para avaliação, que tornam os resultados mais confiáveis do que questionários genéricos. A maioria dos questionários padronizados utilizou o System Usability Scale (SUS) para suas avaliações, possivelmente por ele ser mais fácil e simples de ser aplicado. As outras técnicas mais usadas foram a *Think Aloud*, citada por 35% dos artigos selecionados, enquanto gravações de vídeo foram citadas por 28% e de áudio por 25%.

Foram levantadas possíveis limitações e problemas relacionados à execução dos testes. Os estudos apontaram dificuldades na generalização dos resultados relacionados a amostras de usuários que não correspondem ao público alvo, o que representa um risco para a aceitação e adesão ao uso do aplicativo. Além disso, alguns artigos relataram preocupação com a generalização devido ao número reduzido de usuários.

Nenhuma pesquisa, com exceção daqueles que propuseram e testaram suas próprias abordagens para testes de usabilidade, utilizou métodos de automatização. As coletas e análises foram feitas de forma manual, o que demanda bastante esforço por parte da equipe de testes como destacou o artigo [A2]. As propostas de automatização encontradas nesse estudo que utilizam processamento de logs apresentam limitações e não foram totalmente validadas em relação aos métodos mais comuns.

Os estudos também destacaram a importância de considerar a experiência prévia dos usuários com tecnologias semelhantes às avaliadas, e de testar utilizando uma variedade maior de sistemas e plataformas, tendo em vista a heterogeneidade dos dispositivos móveis e a necessidade de avaliar melhor as funcionalidades dos aplicativos. Estas limitações representam ameaças à validade dos resultados dos

testes e, conseqüentemente, à aceitação pelo usuário final, o que representa ameaças ao sucesso de um projeto de software.

5.1 Limitações de Pesquisa

Esse estudo procurou sintetizar o resultado de evidências encontradas nos artigos, entretanto todos eles tiveram caráter acadêmico. Embora as buscas nas bases de dados selecionadas tenham retornado uma boa quantidade de trabalhos com pesquisas relevantes e de qualidade, existem outras bases e publicações de conferências que podem ter mais evidências a serem coletadas. Além disso, alguns artigos não puderam ser acessados gratuitamente, fazendo com que algum artigo relevante possa deixar de ter deixado de ser analisado.

5.2 Trabalhos Futuros

A partir deste estudo, seus achados e limitações são propostos os seguintes direcionamentos de pesquisa:

- Realizar survey nas empresas para comparação com dados da literatura;
- Desenvolvimento de processos/ferramentas que diminuam essas limitações;
- Estender o mapeamento sistemático com outras bases de dados e conferências;
- Síntese dos benefícios semelhantes ao que se fez com as limitações;
- Desenvolvimento de estratégias para tratamento das limitações encontradas nesse trabalho.

6 Bibliografia

ABRAN, Alain et al. **Usability meanings and interpretations in ISO standards.** Software Quality Journal, v. 11, n. 4, p. 325-338, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 9241-11: Requisitos ergonômicos para trabalho de escritório com computadores: Parte 11—Orientação sobre usabilidade.** Rio de Janeiro, 2002.

CORRAL, Luis; SILLITTI, Alberto; SUCCI, Giancarlo. **Software assurance practices for mobile applications.** Computing, v. 97, n. 10, p. 1001-1022, 2015.

DINH, Hoang T. et al. **A survey of mobile cloud computing: architecture, applications, and approaches.** Wireless communications and mobile computing, v. 13, n. 18, p.1587-1611, 2013.

DYBÅ, Tore; DINGSØYR, Torgeir. **Empirical studies of agile software development: A systematic review.** Information and software technology, v. 50, n. 9, p. 833-859, 2008.

GALIN, Daniel. **Software quality assurance: from theory to implementation.** Pearson Education Limited, 2004.

HOLZINGER, Andreas. **Usability engineering methods for software developers.** Communications of the ACM, v. 48, n. 1, p. 71-74, 2005.

IEEE. **IEEE Standard for Software Quality Assurance Processes - Redline,** in IEEE Std 730-2014 (Revision of IEEE Std 730-2002) - Redline , vol., no., pp.1-231, June 13 2014, 2014.

JURISTO, Natalia; WINDL, Helmut; CONSTANTINE, Larry. **Guest Editors' Introduction: Introducing Usability.** IEEE software, v. 18, n. 1, p. 20-21, 2001.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering.** Keele University and Durham University Joint Report, Tech. Rep. EBSE 2007-001, 2007.

LOMOTÉY, Richard K.; DETERS, Ralph. **Architectural designs from mobile cloud computing to ubiquitous cloud computing-survey.** In: Services (SERVICES), 2014 IEEE World Congress. IEEE, 2014. p. 418-425, 2014.

MYERS, Glenford J. **The Art of Software Testing.** John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

NAIK, Kshirasagar; TRIPATHY, Priyadarshi. **Software testing and quality assurance: theory and practice.** John Wiley & Sons, Inc., Hoboken, New Jersey, 2008.

RADATZ, Jane; GERACI, Anne; KATKI, Freny. **IEEE standard glossary of software engineering terminology.** IEEE Std, v. 610121990, n. 121990, p. 3,1990.

WASSERMAN, Anthony I. **Software engineering issues for mobile application development.** In: Proceedings of the FSE/SDP workshop on Future of software engineering research. ACM, 2010. p. 397-400, 2010.

ZHANG, Dongsong; ADIPAT, Boonlit. **Challenges, methodologies, and issues in the usability testing of mobile applications.** International journal of human-computer interaction, v. 18, n. 3, p. 293-308, 2005.

Apêndice A - Protocolo de Revisão Sistemática



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO
CENTRO DE INFORMÁTICA
2017.2



Protocolo de Revisão Sistemática

Estudo Sistemático Sobre Testes de Usabilidade
Realizados em Aplicativos Móveis

Aluno: Valdi Ferreira do Nascimento Júnior (vfnj@cin.ufpe.br)

Orientador: Alexandre Vasconcelos (amlv@cin.ufpe.br)

Co-Orientador: Ivaldir de Farias Junior (ihfj@cin.ufpe.br)

Recife

2017

Objetivo

A Garantia da Qualidade de Software é um conjunto de ações planejadas e sistemáticas que visa à produção de evidências que estabeleçam a confiança de que seu produto atende os requisitos estabelecidos no grau necessário, esperado e desejado (IEEE, 2014). Esse processo de garantia e seus métodos, como testes e validação, variam de acordo com o tipo de software e seus diferentes ambientes. Os ambientes móveis e suas aplicações possuem características únicas de desempenho e usabilidade que são a razão da dificuldade na elaboração e execução da garantia da qualidade (FRANKIE e WEISE, 2011). Além disso, oferecem acesso a uma grande variedade de serviços e funcionalidades em diversas áreas (educação, e-commerce, cuidados de saúde, etc.) independente de local e hora (DINH et al., 2013). Isso permite que os usuários executem diversas atividades importantes através dos seus dispositivos. Sendo assim, alcançar um alto nível de satisfação do usuário representa uma condição crítica para o sucesso de uma aplicação. Portanto é essencial avaliar a usabilidade de uma aplicação móvel a fim de garantir que ela é prática, efetiva e fácil de usar, principalmente do ponto de vista do usuário (ZHANG e ADIPAT, 2005).

O objetivo geral desta pesquisa é realizar um estudo sistemático da literatura para investigar como estão sendo realizados os testes de usabilidade específicos de aplicações para ambientes móveis.

1. Questões de Pesquisa

Segundo Kitchenham e Charters (2007), a parte mais importante de uma revisão sistemática são as questões de pesquisa, pois elas guiam toda a metodologia da revisão sistemática.

Para alcançar os objetivos foi formulada a seguinte questão de pesquisa principal (QPP):

- QPP: Como estão sendo realizados os testes de usabilidade em aplicativos móveis?

As seguintes questões de pesquisa específicas (QPE) foram definidas para orientar a extração, análise e síntese dos resultados:

- QPE1: Quais são as práticas, métodos, processos e/ou técnicas reportadas

do processo de testes de usabilidade no desenvolvimento de aplicativos móveis?

- QPE2: Quais são os problemas e limitações dessas práticas?

2. Estratégia de Busca

O objetivo de uma revisão sistemática é encontrar o maior número possível de estudos primários relacionados com as questões de pesquisa. O rigor nesse tipo de estudo é o que o distingue dos tipos tradicionais de revisão. Sendo assim é fundamental desenvolver e seguir uma estratégia de busca, que deve ser formulada consultando fontes bibliográficas e pesquisadores de experiência relevante (KITCHENHAM e CHARTERS, 2007).

2.1. Termos de Busca

Os termos usados nas buscas feitas nas bases de dados eletrônicas são retirados da estrutura das questões de pesquisa. Os termos foram traduzidos para o inglês, por este ser o idioma das fontes de buscas selecionadas. Os termos de busca, seus sinônimos ou palavras relacionadas são apresentados no Quadro 1.

Quadro 1: Termos de busca, seus sinônimos ou palavras relacionadas.

Termos de Busca	Sinônimos e palavras relacionadas
Testes	test, testing
Usabilidade	usability
Aplicativos móveis	mobile software, mobile application, mobile app

Fonte: Próprio Autor

2.2. String de Busca

A *string* de busca é construída a partir da questão de pesquisa e dos termos de busca. Os sinônimos e palavras relacionadas de cada termo foram concatenados em listas através do operador booleano OR. Por fim a *string* de busca é finalizada com a união das listas usando o operador AND (KITCHENHAM e CHARTERS, 2007). A *string* resultante pode ser visualizada no Quadro 2.

Quadro 2: String de Busca

String de Busca
(test OR testing) AND (usability) AND ("mobile software" OR "mobile application" OR "mobile app")

Fonte: Próprio Autor

2.3. Fontes de Busca

A busca dos estudos foi realizada em bibliotecas digitais diferentes e de relevância na área a fim de aumentar o número de estudos primários relacionados selecionados (KITCHENHAM e CHARTERS, 2007). As fontes de busca selecionadas se encontram no Quadro 3.

Quadro 3: Fontes de Busca

Base de Dados	Website
ACM Digital Library	http://dl.acm.org/
IEEE Xplore	http://www.ieeexplore.ieee.org
Scopus	https://www.scopus.com/

Fonte: Próprio Autor

2.4. Critérios de seleção de estudos

O objetivo desses critérios é identificar os estudos primários que tenham ligação direta com a questão de pesquisa através de evidências. Embora possam ser refinados posteriormente, os critérios devem ser definidos na elaboração do protocolo a fim de reduzir a probabilidade de viés na pesquisa (KITCHENHAM e CHARTERS, 2007). Sendo assim foram elaborados os seguintes critérios de inclusão e exclusão:

2.4.1. Critérios de Inclusão

- IC1: Estudos primários;
- IC2: Estudos que apresentem dados empíricos;
- IC3: Pesquisas qualitativas ou quantitativas
- IC4: Trabalhos completos publicados em revistas revisadas por pares ou conferências;
- IC5: Estudos industriais e/ou acadêmicos

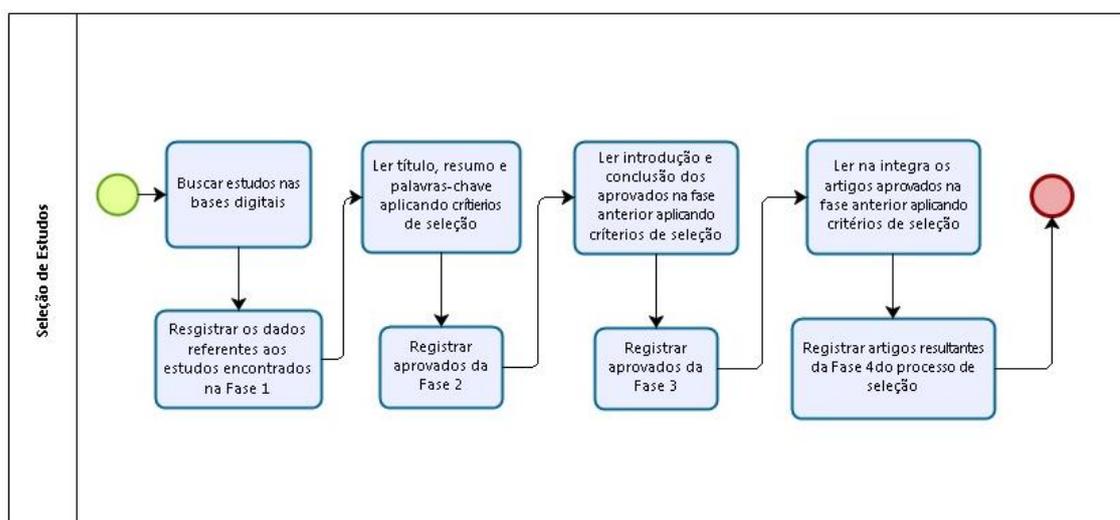
2.4.2. Critérios de exclusão

- EC1: Estudos duplicados ou repetidos;
- EC2: Estudos não escritos em inglês;
- EC3: Estudos que não tratem de testes de usabilidade em aplicativos de dispositivos móveis;
- EC4: Editoriais, prefácios, resumos de artigos, cursos, oficinas, painéis, entrevistas, notícias, opiniões de especialistas, sessões de pôsteres, apresentações, livros, dissertações e teses;
- EC5. Artigos que não estão disponíveis para download nos ambientes institucionais do CIn/UFPE.

2.5 Processo de seleção dos estudos

O processo de seleção dos estudos possui múltiplas fases e é realizado por dois pesquisadores com o intuito de diminuir o viés na aplicação dos critérios de inclusão e exclusão. O processo, ilustrado na Figura 1, foi realizado nas seguintes fases:

Figura 1: Processo de seleção de estudos



Fonte: Elaboração própria

Fase 1: A busca é realizada nas bases digitais selecionadas para esse estudos utilizando a *string* construída a partir da questão de pesquisa. Os dados principais de identificação dos artigos serão exportados das bibliotecas e registrados utilizando o recurso de compartilhamento e construção de planilhas do Google Drive (<https://drive.google.com>). Os estudos duplicados serão removidos de acordo com

os metadados (título, autores e ano de publicação).

Fase 2: **Os títulos, resumo e palavras-chaves** dos estudos remanescentes da Fase 1 são lidos e os critérios de exclusão e inclusão são aplicados. Em caso de dúvidas sobre a exclusão ou inclusão do estudo, o mesmo será incluído para a próxima fase. Os estudos aprovados nessa fase serão registrados e compartilhados usando recursos de planilhas do Google Drive (<https://drive.google.com>).

Fase 3: **A introdução e conclusão dos estudos** aprovados na Fase 2 são lidos para que sejam aplicados os critérios de seleção. Os estudos que forem selecionados na Fase 3 são catalogados como nas fases anteriores.

Fase 4: **Nesta fase é feita a leitura completa** dos artigos resultantes da fase anterior. Eles são submetidos aos critérios de inclusão e exclusão para obtenção dos artigos primários finais do processo de seleção. Os estudos resultantes desse processo serão registrados e compartilhados usando recursos de planilhas do Google Drive (<https://drive.google.com>).

2.6. Avaliação de Qualidade

Os critérios de qualidade são definidos com o objetivo de oferecer uma avaliação mais detalhada para seleção dos estudos, investigar se níveis de qualidade de explanação oferecem resultados diferentes, ponderar importância de estudos individuais, orientar a interpretação dos achados e propor recomendações de pesquisa futuras (KITCHENHAM e CHARTERS, 2007). Os critérios para esse processo estão descritos no Formulário A.

Formulário A

Avaliação da Qualidade		ID:
Item	Critérios	Avaliação
Q1	Existe uma clara declaração dos objetivos da pesquisa?	
Q2	A tecnologia (aplicativo) avaliada é claramente definida?	
Q3	A estratégia de pesquisa foi adequada aos objetivos da pesquisa?	
Q4	Os métodos de coleta de dados foram adequadamente descritos?	

Q5	O trabalho é bem/adequadamente referenciado (apresenta trabalhos relacionados/semelhantes e baseia-se em modelos e teorias da literatura)?	
Q6	O estudo relata de forma clara e não ambígua os resultados?	
Q7	Existe um método ou um conjunto de métodos descrito para a realização dos testes?	
Q8	Os participantes ou unidades de observação estão adequadamente descritos?	
Q9	Os objetivos ou questões do estudo são alcançados?	
Q10	O estudo possui valor para a academia ou para a indústria?	

Fonte: Elaboração própria baseada em Kitchenham e Charters (2007).

A avaliação é feita usando a escala de 3 valores proposta por Likert (1932) onde são atribuídas respostas gradativas para avaliar o quanto o estudo está de acordo com cada um dos critérios de qualidade descritos no Formulário A:

- Não atende (0): deve ser atribuído caso não exista nada no trabalho que atenda ao critério avaliado;
- Dúbio (0.5): deve ser atribuído na falta de clareza do critério avaliado;
- Atende (1): deve ser atribuído se o critério avaliado for atendido.

Os artigos serão classificados em 4 faixas de qualidade de acordo com o somatório alcançado pela a avaliação dos itens de avaliação de qualidade:

Tabela 1 - Níveis de Qualidade.

Baixa	Média	Alta	Muito Alta
$0 \leq N \leq 3$	$3 < N \leq 7$	$7 < N \leq 8,5$	$8,5 < N \leq 10$

Fonte: Elaborada pelo autor.

3. Extração de dados

Para que os registros das informações obtidas dos estudos primários sejam feitos de maneira mais precisa, devem ser desenvolvidos formulários aos quais os estudos são submetidos. Esses formulários são elaborados com o intuito de evitar vieses na pesquisa. Entretanto seu objetivo principal é a coleta das informações relacionadas às questões de pesquisa (KITCHENHAM e CHARTERS, 2007).

Formulário B

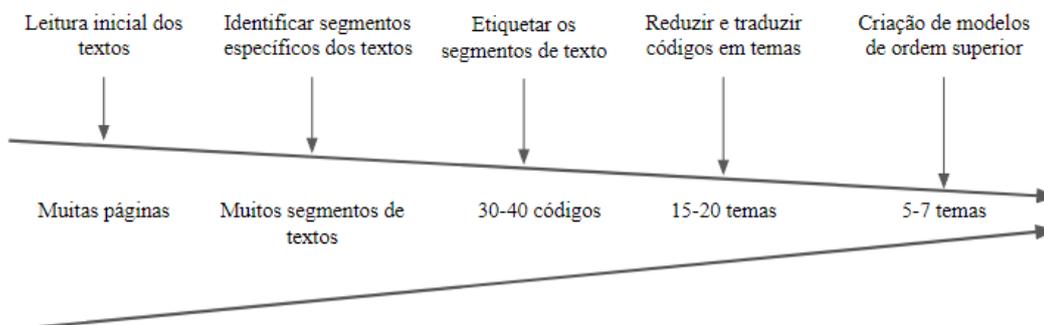
Formulário de Extração de dados		
ID:	Ano:	País:
Título:		
Autores:		
Fonte:		
Evidências de Contexto		
Tipo de estudo		Método de pesquisa
Questões de Pesquisa:		
QP1: Quais são as práticas, métodos, processos e/ou técnicas reportadas do processo de testes de usabilidade no desenvolvimento de aplicativos móveis?		
QP2: Quais são os benefícios e limitações dessas práticas?		

Fonte: Elaborada pelo autor com base em Dyba e Dingsor (2008)

4. Síntese de dados

Nessa fase os dados resultantes extraídos dos estudos primários selecionados são categorizados e organizados. O método de síntese temática proposto por Cruzes e Dyba (2011) é utilizado por oferecer uma maneira bem organizada de descrever uma grande quantidade de evidências potencialmente diversas. A Figura 2 provê uma visão geral do processo, que possui as seguintes etapas:

Figura 2: Processo de Síntese Temática



Fonte: Cruzes e Dyba (2011)

Extração de dados: Os dados e textos essenciais são retirados dos estudos primários de forma explícita e consistente de acordo com a estratégia definida dando ênfase à extração de dados de publicação (título, autores, ano de publicação e etc.), dados de descrição de contextos (assuntos, tecnologias, contexto e etc.) e dados descobertos (resultados, eventos, comportamentos e etc.).

Codificação dos dados: É o processo de examinar e organizar os dados encontrados em todos os estudos da revisão. Permite agrupar informações similares em categorias que compartilhem características semelhantes, baseadas em uma teoria ou ideia descritiva. Para isso, segmentos de textos são etiquetados usando um processo de codificação.

Códigos são rótulos atribuídos a unidades do texto com significado. Podem ser anexados a palavras, a uma ou mais frases ou mesmo a um parágrafo inteiro, conectados ou não a um cenário (MILES e HUBERMAN, 1994). A codificação vai além de catalogar os trechos relevantes dos textos, pois exige clareza do contexto em que as descobertas foram feitas. Codificar permite que o pesquisador organize e agrupe informações que possuam características semelhantes nas mesmas categorias.

Transformar códigos em temas: Os códigos são agrupados em conjuntos menores com mais significado com o objetivo de reduzir a grande quantidade de código em unidades menores para facilitar a interpretação do pesquisador e auxiliar o processo de construção de um mapa cognitivo. Inicialmente os códigos são analisados para identificar as diferentes maneiras de reuni-los em temas mais abrangentes. Conforme essa fase avança os dados codificados podem ser

reorganizados, reclassificados em diferentes códigos e até mesmo em novos códigos. Essa fase se encerra quando se satura a possibilidade de criação de temas. Tabelas, mapas de árvore, redes temáticas e mapas mentais são recursos que podem ser usados para organizar os códigos e transformá-los em temas.

Criação de um modelo de ordem superior: Os temas que surgiram dos passos anteriores devem agora ser explorados e interpretados buscando estabelecer temas superiores e suas relações entre si. Ao desenvolver esses temas de ordem maior, as estruturas criadas na fase anterior podem ter seus ramos ou caminhos revistos a fim de descrevê-los e contextualizá-los buscando encontrar relações.

Referências

CRUZES, Daniela S.; DYBA, Tore. **Recommended steps for thematic synthesis in software engineering.** In: Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on. IEEE, 2011. p. 275-284, 2011.

DINH, Hoang T. et al. **A survey of mobile cloud computing: architecture, applications, and approaches.** Wireless communications an mobile computing, v. 13, n. 18, p. 1587-1611, 2013.

DYBÅ, Tore; DINGSØYR, Torgeir. **Empirical studies of agile software development: A systematic review.** Information and software technology, v. 50, n. 9, p. 833-859, 2008.

FRANKE, Dominik; WEISE, Carsten. **Providing a software quality framework for testing of mobile applications.** In: *Software Testing, Verification and Validation (ICST)*, 2011 IEEE Fourth International Conference on. IEEE, 2011. p. 431-434, 2011.

IEEE. **IEEE Standard for Software Quality Assurance Processes - Redline,** in IEEE Std 730-2014 (Revision of IEEE Std 730-2002) - Redline , vol., no., pp.1-231, June 13 2014, 2014.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering.** Keele University and Durham University Joint Report, Tech. Rep. EBSE 2007-001, 2007.

LIKERT, Rensis. **A technique for the measurement of attitudes.** Archives of psychology, 1932.

MILES, Matthew B.; HUBERMAN, A. Michael. **Qualitative data analysis: An expanded sourcebook.** sage, 1994.

ZHANG, Dongsong; ADIPAT, Boonlit. **Challenges, methodologies, and issues in the usability testing of mobile applications.** International journal of human-computer interaction, v. 18, n. 3, p. 293-308, 2005.

Apêndice B – Artigos Selecionados

Artigos Selecionados				
ID	Ano	Fonte	Autor(ers)	Título
A1	2008	IEEE	M. Fetaji ¹ ; Z. Dika; B. Fetaji	Usability testing and evaluation of a mobile software solution: A case study
A2	2015	IEEE	M. Borys; M. Milosz	Mobile application usability testing in quasi-real conditions
A3	2015	IEEE	S. Z. A. Shah; I. A. Khan; I. Maqsood; T. A. Khan; Y. Khan	First-Aid Application for Illiterates and Its Usability Evaluation
A4	2015	IEEE	H. M. Az-zahra; A. Pinandito; H. Tolle	Usability evaluation of mobile application in culinary recommendation system
A5	2014	IEEE	A. Hussain; N. A. Mutalib; A. Zaino	A usability testing on JFakih Learning Games for hearing impairment children
A6	2017	Scopus	Hussain A., Mkpojiogu E.O.C., Abubakar H., Hassan H.M.	The usability evaluation of Mudah.my on mobile device
A7	2017	Scopus	Hussain A., Mkpojiogu E.O.C., Musa J., Mortada S.	A user experience evaluation of Amazon Kindle mobile application
A8	2017	Scopus	van Leer E., Pfister R.C., Zhou X.	An iOS-based Cepstral Peak Prominence Application: Feasibility for Patient Practice of Resonant Voice
A9	2016	Scopus	Paulino D., Amaral D., Amaral M., Reis A., Barroso J., Rocha T.	"Professor piano": A music application for people with intellectual disabilities
A10	2016	Scopus	Yoo S., Jung S.Y., Kim S., Kim E., Lee K.-H., Chung E., Hwang H.	A personalized mobile patient guide system for a patient-centered smart hospital: Lessons learned from a usability test and satisfaction survey in a tertiary university hospital
A11	2016	Scopus	Nitsch M., Dimopoulos C.N., Flaschberger E., Saffran K., Kruger J.F., Garlock L., Wilfley D.E., Taylor C.B., Jones M.	A Guided Online and Mobile Self-Help Program for Individuals With Eating Disorders: An Iterative Engagement and Usability Study
A12	2015	Scopus	Wei Q., Chang Z., Cheng Q.	Usability study of the mobile library App: an example from Chongqing University
A13	2015	Scopus	Vedanthan R., Blank E., Tuikong N., Kamano J., Misoi L., Tulieng D., Hutchinson C., Ascheim D.D., Kimaiyo S., Fuster V., Were M.C.	Usability and feasibility of a tablet-based decision-support and integrated Record-keeping (DESIRE) tool in the nurse management of hypertension in rural western Kenya

A14	2014	Scopus	De La Vega R., Roset R., Castarlenas E., Sánchez-Rodríguez E., Solé E., Miró J.	Development and testing of painometer: A smartphone app to assess pain intensity
A15	2014	Scopus	O'Malley G., Dowdall G., Burls A., Perry I.J., Curran N.	Exploring the usability of a mobile app for adolescent obesity management
A16	2014	Scopus	Mirkovic J., Kaufman D.R., Ruland C.M.	Supporting cancer patients in illness management: Usability evaluation of a mobile app
A17	2013	Scopus	Mintmire E., Ling C., Shehab R.	Usability evaluation of a paratransit mobile application
A18	2013	Scopus	Ben-Zeev D., Kaiser S.M., Brenner C.J., Begale M., Duffecy J., Mohr D.C.	Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia.
A19	2013	Scopus	Buller D.B., Berwick M., Shane J., Kane I., Lantz K., Buller M.K.	User-centered development of a smart phone mobile application delivering personalized real-time advice on sun protection
A20	2013	Scopus	Ma X., Yan B., Chen G., Zhang C., Huang K., Drury J., Wang L.	Design and implementation of a toolkit for usability testing of mobile apps
A21	2009	Scopus	Kenteris M., Gavalas D., Economou D.	An innovative mobile electronic tourist guide application
A22	2017	Scopus	Hussain A., Mkpjojiogu E.O.C., Almazini H., Almazini H.	Assessing the usability of Shazam mobile app
A23	2017	Scopus	Ribeiro D., Machado J., Ribeiro J., Vasconcelos M.J.M., Vieira E.F., De Barros A.C.	SousChef: Mobile meal recommender system for older adults
A24	2014	Scopus	Reynoldson C., Stones C., Allsop M., Gardner P., Bennett M.I., Closs S.J., Jones R., Knapp P.	Assessing the quality and usability of smartphone apps for pain self-management
A25	2005	ACM	Anne Kaikkonen and Mihael Cankar and Titti Kallio and Anu Kankainen	Usability Testing of Mobile Applications: A Comparison Between Laboratory and Field Testing
A26	2016	ACM	Jing Xu and Xiang Ding and Ke Huang and Guanling Chen	Unsupervised Detection of Abnormal Moments for Usability Testing of Mobile Apps
A27	2016	ACM	Jackson Feijó Filho; Filho and Wilson Prata and Juan Oliveira	Affective-ready, Contextual and Automated Usability Test for Mobile Software
A28	2016	ACM	Alba de Castro and José A. Macías	SUSApp: A Mobile App for Measuring and Comparing Questionnaire-Based Usability Assessments

Apêndice C – Avaliação dos Artigos

Avaliação da Qualidade												
ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	TOTAL	Avaliação
A1	1	1	1	0,5	0,5	1	1	0,5	1	1	8,5	Alta
A2	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A3	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A4	1	1	1	1	0,5	1	1	1	1	0,5	9	Muito Alta
A5	1	1	0,5	0,5	1	1	1	0,5	0,5	1	8	Alta
A6	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A7	1	0,5	1	1	0,5	1	1	1	1	0,5	8,5	Alta
A8	1	1	0,5	1	0,5	1	1	1	1	1	9	Muito Alta
A9	1	1	0,5	1	1	1	1	1	1	0,5	9	Muito Alta
A10	1	1	1	1	0,5	1	0,5	1	1	1	9	Muito Alta
A11	1	0,5	1	1	1	1	1	1	1	1	9,5	Muito Alta
A12	1	1	1	1	1	1	1	1	0,5	1	9,5	Muito Alta
A13	1	1	0,5	0	0	1	0,5	0,5	0,5	0,5	5,5	Média
A14	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A15	1	1	1	0,5	1	1	1	1	0,5	1	9	Muito Alta
A16	1	1	1	1	0,5	1	0,5	1	1	1	9	Muito Alta
A17	1	1	1	1	1	1	1	0,5	1	1	9,5	Muito Alta
A18	1	1	1	0,5	1	1	1	0,5	1	1	9	Muito Alta
A19	1	1	0,5	0,5	0,5	0,5	1	0,5	1	1	7,5	Alta
A20	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A21	1	1	1	1	1	0,5	0,5	1	0,5	0,5	8	Alta
A22	1	1	1	1	1	1	0,5	0,5	1	1	9	Alta
A23	1	1	1	0,5	1	0,5	0,5	0	0,5	0,5	6,5	Média
A24	1	1	1	1	1	1	1	0,5	1	1	9,5	Muito Alta
A25	1	1	1	1	1	1	1	1	1	1	10	Muito Alta
A26	1	1	0,5	1	1	1	1	0,5	1	1	9	Muito Alta
A27	1	1	1	0	1	1	1	0,5	1	1	8,5	Muito Alta
A28	1	1	1	0,5	1	1	1	0,5	1	1	9	Muito Alta

Apêndice D – Contexto dos Artigos Selecionados

ID	Tipo de Estudo	Tipo de Metodologia
A1	Empírico	Estudo de caso/ Experimento
A2	Empírico	Estudo de Caso/ Experimento
A3	Empírico	Experimento
A4	Empírico	Experimento
A5	Empírico	Estudo Comparativo - Experimento
A6	Empírico	Experimento
A7	Empírico	Experimento
A8	Empírico	Experimento
A9	Empírico	Quase-Experimento
A10	Empírico	Experimento
A11	Empírico	Experimento
A12	Empírico	Experimento
A13	Empírico	Experimento
A14	Empírico	Experimento
A15	Empírico	Experimento
A16	Empírico	Experimento
A17	Empírico	Experimento
A18	Empírico	Experimento
A19	Empírico	Experimento
A20	Empírico	Experimento
A21	Empírico	Experimento
A22	Empírico	Experimento
A23	Empírico	Experimento
A24	Empírico	Estudo de Campo/Experimento
A25	Empírico	Experimento
A26	Empírico	Experimento
A27	Empírico	Experimento
A28	Empírico	Experimento

Apêndice E – Referências dos Artigos Selecionados

ID	Referências
A1	FETAJI, Majlinda; DIKA, Zamir; FETAJI, Bekim. Usability testing and evaluation of a mobile software solution: a case study. In: Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on. IEEE, 2008. p. 501-506.
A2	BORYS, Magdalena; MILOSZ, Marek. Mobile application usability testing in quasi-real conditions. In: Human System Interactions (HSI), 2015 8th International Conference on. IEEE, 2015. p. 381-387.
A3	SHAH, Syed Zafar Ali et al. First-Aid Application for Illiterates and Its Usability Evaluation. In: Frontiers of Information Technology (FIT), 2015 13th International Conference on. IEEE, 2015. p. 125-131.
A4	AZ-ZAHRA, Hanifah Muslimah; PINANDITO, Aryo; TOLLE, Herman. Usability evaluation of mobile application in culinary recommendation system. In: Wireless and Mobile (APWiMob), 2015 IEEE Asia Pacific Conference on. IEEE, 2015. p. 89-94.
A5	HUSSAIN, Azham; MUTALIB, Norasyikin Abdul; ZAINO, Azida. A usability testing on JFakih Learning Games for hearing impairment children. In: Information and Communication Technology for the Muslim World (ICT4M), 2014 The 5th International Conference on. IEEE, 2014. p. 1-4.
A6	HUSSAIN, Azham et al. The usability evaluation of Mudah. my on mobile device. In: AIP Conference Proceedings. AIP Publishing, 2017. p. 020058.
A7	HUSSAIN, Azham et al. A user experience evaluation of Amazon Kindle mobile application. In: AIP Conference Proceedings. AIP Publishing, 2017. p. 020060.
A8	VAN LEER, Eva; PFISTER, Robert C.; ZHOU, Xuefu. An iOS-based cepstral peak prominence application: feasibility for patient practice of resonant voice. Journal of Voice, v. 31, n. 1, p. 131. e9-131. e16, 2017.
A9	PAULINO, Dennis et al. Professor Piano: a music application for people with intellectual disabilities. In: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. ACM, 2016. p. 269-274.
A10	YOO, Sooyoung et al. A personalized mobile patient guide system for a patient-centered smart hospital: Lessons learned from a usability test and satisfaction survey in a tertiary university hospital. International journal of medical informatics, v. 91, p. 20-30, 2016.
A11	NITSCH, Martina et al. A guided online and mobile self-help program for individuals with eating disorders: An iterative engagement and usability study. Journal of medical Internet research, v. 18, n. 1, 2016.
A12	WEI, Qunyi; CHANG, Zhaoxin; CHENG, Qin. Usability study of the mobile library App: an example from Chongqing University. Library Hi Tech, v. 33, n. 3, p. 340-355, 2015.
A13	VEDANTHAN, Rajesh et al. Usability and feasibility of a tablet-based Decision-Support and Integrated Record-keeping (DESIRE) tool in the nurse management of hypertension in rural western Kenya. International journal of medical informatics, v. 84, n. 3, p. 207-219, 2015.
A14	DE LA VEGA, Rocío et al. Development and testing of painometer: A smartphone app to assess pain intensity. The Journal of Pain, v. 15, n. 10, p. 1001-1007, 2014.
A15	O'MALLEY, G. et al. Exploring the usability of a mobile app for adolescent obesity management. Physiotherapy, v. 102, p. e44-e45, 2016.
A16	MIRKOVIC, Jelena; KAUFMAN, David R.; RULAND, Cornelia M. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR mHealth and uHealth, v. 2, n. 3, 2014.
A17	MINTMIRE, Elizabeth; LING, Chen; SHEHAB, Randa. Usability evaluation of a paratransit mobile

	application. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage CA: Los Angeles, CA: SAGE Publications, 2013. p. 1318-1322.
A18	BEN-ZEEV, Dror et al. Development and usability testing of FOCUS: A smartphone system for self-management of schizophrenia. <i>Psychiatric rehabilitation journal</i> , v. 36, n. 4, p. 289, 2013.
A19	BULLER, David B. et al. User-centered development of a smart phone mobile application delivering personalized real-time advice on sun protection. <i>Translational behavioral medicine</i> , v. 3, n. 3, p. 326-334, 2013.
A20	MA, Xiaoxiao et al. Design and implementation of a toolkit for usability testing of mobile apps. <i>Mobile Networks and Applications</i> , v. 18, n. 1, p. 81-97, 2013.
A21	KENTERIS, Michael; GAVALAS, Damianos; ECONOMOU, Daphne. An innovative mobile electronic tourist guide application. <i>Personal and ubiquitous computing</i> , v. 13, n. 2, p. 103-118, 2009.
A22	HUSSAIN, Azham et al. Assessing the usability of Shazam mobile app. In: <i>AIP Conference Proceedings</i> . AIP Publishing, 2017. p. 020057.
A23	RIBEIRO, David et al. SousChef: Mobile meal recommender system for older adults. In: <i>Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health. ICT4AWE</i> , 2017. p. 36-45
A24	REYNOLDSON, Charmian et al. Assessing the quality and usability of smartphone apps for pain self-management. <i>Pain medicine</i> , v. 15, n. 6, p. 898-909, 2014.
A25	KAIKKONEN, Anne et al. Usability testing of mobile applications: A comparison between laboratory and field testing. <i>Journal of Usability studies</i> , v. 1, n. 1, p. 4-16, 2005.
A26	XU, Jing et al. Unsupervised Detection of Abnormal Moments for Usability Testing of Mobile Apps. In: <i>Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems</i> . ACM, 2016. p. 3247-3254.
A27	PRATA, Wilson et al. Affective-ready, contextual and automated usability test for mobile software. In: <i>Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct</i> . ACM, 2016. p. 638-644.
A28	DE CASTRO, Alba; MACÍAS, José A. SUSApp: A Mobile App for Measuring and Comparing Questionnaire-Based Usability Assessments. In: <i>Proceedings of the XVII International Conference on Human Computer Interaction</i> . ACM, 2016. p. 19.