

UX-Tips: A UX evaluation technique to support the identification of software application problems

Leonardo Marques
UFAM
Manaus, Amazonas, Brasil
lcm@icomp.ufam.edu.br

Patrícia Matsubara
UFAM
Manaus, Amazonas, Brasil
patriciagfm@icomp.ufam.edu.br

Walter Nakamura
UFAM
Manaus, Amazonas, Brasil
walter@icomp.ufam.edu.br

Igor Wiese
UTFPR
Campo Mourão, Paraná, Brasil
igor@utfpr.edu.br

Luciana Zaina
UFSCar
Sorocaba, São Paulo, Brasil
lzaina@ufscar.br

Tayana Conte
UFAM
Manaus, Amazonas, Brasil
tayana@icomp.ufam.edu.br

ABSTRACT

User Experience (UX) is a quality attribute that plays an important role in the quality of software applications, as it addresses users' feelings about a given product. Thus, evaluating the UX is essential to improve its quality. However, most UX evaluation techniques limit to indicate whether the UX was positive or negative, which makes it challenging to identify what must be corrected by the software development team. This paper presents UX-Tips (User eXperience for Interactive ProductS), a technique developed to deal with this limitation, whose main objective is to allow the identification of problems related to UX. We also present two empirical studies conducted to evaluate UX-Tips. In the first study, we assessed the feasibility of the technique concerning its primary goal. In the second study, we compared the performance of UX-Tips with a similar technique. The results showed that UX-Tips allowed the subjects to indicate whether the UX was negative, and the specific problems they detected. It also allowed more effective and efficient evaluations, in addition to being considered easier to use. In practice, UX-Tips enable the software development team to identify the problems that impair users' experience, which should be fixed in order to provide a good UX.

CCS CONCEPTS

- Software and its Engineering → Software Creation and Management → Software Verification and Validation

KEYWORDS

Software Quality, Verification and Validation, User Experience, Software evaluation

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

SBES 2019, September 23–27, 2019, Salvador, Brazil

© 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7651-8/19/09...\$15.00

<https://doi.org/10.1145/3350768.3350783>

ACM Reference format:

Leonardo Marques, Patrícia Matsubara, Walter Nakamura, Igor Wiese, Luciana Zaina and Tayana Conte. 2019. UX-Tips: A UX evaluation technique to support the identification of software application problems. In *Proceedings of XXXIII Brazilian Symposium on Software Engineering, Salvador, Brazil, September 23-27, 2019 (SBES 2019)*, 10 pages. <https://doi.org/10.1145/3350768.3350783>

1 Introdução

Usabilidade e UX são preocupações comuns ao projetar um produto [8]. Por muitos anos, avaliar se o usuário consegue alcançar os objetivos com eficácia e eficiência foi o suficiente para avaliar a qualidade de um software [10], tornando a usabilidade uma das principais preocupações ao projetá-lo. Com isso, a comunidade de Engenharia de Software (ES) tem investigado abordagens para integrar técnicas de usabilidade ao processo de desenvolvimento de software [7]. Há também propostas de novas técnicas relacionadas a esse atributo de qualidade, por exemplo para estimular a equipe de desenvolvimento a eliciar requisitos de usabilidade [17] e a avaliar a usabilidade do produto [5]. Entretanto, a UX tem atraído crescente interesse nos últimos anos [23]. Os conceitos de usabilidade e UX são relacionados entre si e, portanto, espera-se que a comunidade de ES também crie e avalie técnicas de UX mais adequadas ao contexto do desenvolvimento de software. Para fazer um paralelo com usabilidade, já há investigações sobre a elicitação de requisitos de UX também [24].

O termo UX foi primariamente introduzido por Donald Norman [21]. De maneira complementar à usabilidade, que foca principalmente na cognição e no desempenho do usuário, a UX provê uma visão mais holística, considerando não apenas aspectos orientados à tarefa, mas focando nos aspectos subjetivos, como afeto, sensações, emoções e valor da interação do usuário [3]. Assim, a UX se tornou um importante atributo de qualidade a ser considerado em um projeto de software [25].

Dada a sua importância, muitas pesquisas têm sido conduzidas na área, resultando em uma variedade de técnicas para avaliar a UX [31]. Contudo, poucos métodos são específicos para o contexto de aplicações de software. O Mapeamento Sistemático (MS) de Rivero e Conte [27] sobre técnicas de avaliação de UX de 2010 a 2015, por exemplo, apontou a existência de 227 técnicas, dentre as quais poucas eram específicas para avaliar aplicações de software (13,7% com foco em aplicações Web e 8,8% em aplicações móveis). Embora métodos genéricos possam ser utilizados para avaliar aplicações de software, eles podem não avaliar aspectos particulares desse tipo de produto, como por exemplo, aspectos da interface (organização, ícones) e gerenciamento de bateria.

Além disso, a maior parte das técnicas foca em capturar somente as percepções relacionadas às emoções dos usuários em relação ao produto avaliado, sem buscar identificar as razões que levaram a uma experiência positiva ou negativa [27]. Embora seja importante capturar a percepção dos usuários, ela não é suficiente para auxiliar a equipe de desenvolvimento a identificar problemas específicos que possam resultar em melhorias nas aplicações de software. Assim, torna-se essencial avaliar a UX de forma completa, verificando se um software provê maneiras eficazes e eficientes de manipulá-lo, ao mesmo tempo em que se avalia as emoções que ele evoca no usuário [10].

Nesse contexto, profissionais e acadêmicos têm procurado novas abordagens para o projeto de produtos interativos, que contemplem tanto as qualidades do produto como as experiências do uso [9]. Desta forma, com o objetivo de mitigar as limitações expostas anteriormente, este artigo apresenta uma nova técnica para avaliação de UX, a UX-Tips. O principal objetivo desta técnica é identificar os problemas que afetam a experiência, avaliando os aspectos subjetivos e específicos de aplicações de software, podendo ser utilizada tanto em inspeções (avaliações por especialistas) quanto em testes com usuários.

Para isto, foram realizados dois estudos experimentais para avaliar a UX-Tips. O primeiro, realizado com 25 participantes, verificou se a técnica permite a identificação dos problemas que causaram uma UX negativa. Além disso, coletou-se sugestões dos participantes para a melhoria e proposição de uma nova versão da UX-Tips. Assim, o segundo estudo, realizado com 68 participantes, comparou o desempenho da UX-Tips com outra técnica em termos de eficácia, eficiência e percepção de uso. Os resultados indicaram que a UX-Tips atende ao seu objetivo, permitindo verificar tanto se a UX foi negativa quanto os possíveis problemas relacionado à UX. Além disso, verificou-se que a UX-Tips possibilitou avaliações mais eficazes, eficientes e foi mais fácil de usar.

2 Referencial Teórico

2.1 Usabilidade e UX

A usabilidade é um aspecto muito importante da qualidade, especialmente no caso de aplicativos interativos [15]. Há várias

definições para usabilidade. De acordo com a ISO 9241-11 [13], usabilidade é “*a medida que um sistema, produto ou serviço pode ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um determinado contexto de uso*”. Esta definição indica que, para um produto ser usável, os usuários devem ser capazes de usá-lo para atingir seus objetivos em um período de tempo aceitável e se sentirem satisfeitos com os resultados obtidos. Em Engenharia de Software (ES), a definição para usabilidade é dada no padrão ISO 25010 [12], que a descreve como “*a capacidade de um produto de software ser compreendido, aprendido, operado e atraente ao usuário, quando usado sob condições específicas*”. Essa definição enfatiza não apenas a relevância de uma interface intuitiva e com boa estética, mas também o potencial de um software atender as expectativas do usuário. Portanto, usabilidade é um atributo de qualidade focado na eficiência e realização de tarefas [1].

Por outro lado, o conceito de UX é comumente entendido como subjetivo, dependente de contexto e dinâmico [3]. Segundo a ISO 9241-210 [11], a UX é definida como “*As percepções e respostas de uma pessoa que resultam do uso e/ou uso antecipado de um produto, sistema ou serviço*”. Com o desenvolvimento de novas tecnologias, os usuários não estão procurando apenas realizar uma tarefa, mas também se divertir e se entreter [22]. Nesse sentido, a usabilidade não é suficiente para definir a qualidade de um software e conseguir a aceitação do usuário.

Hassenzahl [10] sugere que a UX de um produto interativo, como um software, pode ser avaliada considerando dois diferentes atributos: pragmáticos e hedônicos. Os atributos pragmáticos estão relacionados a fornecer meios eficazes e eficientes de manipular o software, enquanto os atributos hedônicos estão relacionados ao estímulo, identificação e aos sentimentos do usuário ao interagir com o software. Assim, Hassenzahl [10] descreve a UX como mais abrangente do que a usabilidade, que foca apenas nos aspectos pragmáticos. Esta abordagem é reforçada por Hassan e Galal-Edeen [8], que sugerem que a UX é holística e afirmam que a usabilidade pode ser vista como um subconjunto da UX. Neste artigo, o conceito de UX adotado está relacionado a considerar a usabilidade como parte da UX.

2.2 Trabalhos Relacionados

Com o surgimento e crescente interesse pelo conceito de UX na comunidade científica, tem crescido também a quantidade de pesquisas propondo métodos para avaliar e melhorar a UX. Vermeeren *et al.* [31] identificaram 96 métodos de avaliação de UX presentes na academia e na indústria. Os autores analisaram os métodos com base na fase de desenvolvimento do produto e no período de experiência estudado. Como conclusão, eles identificaram que é preciso desenvolver métodos com fundamentação teórica, considerando diferentes focos da experiência e o contexto para o qual o método é desenvolvido. Além disso, os autores destacaram a necessidade de métodos que

sejam praticáveis, isto é, que sejam fáceis de usar, analisar e que os resultados sejam úteis.

Rivero e Conte [27] realizaram um MS com o objetivo de caracterizar os métodos usados para avaliar a UX. Um dos resultados indica a baixa quantidade de técnicas de avaliação de UX específicas para contextos como o de software. Embora técnicas genéricas possam ser utilizadas para avaliar aplicações de software, podem não capturar aspectos específicos desse contexto, como gerenciamento de bateria, aspectos da interface (cor, ícones), entre outros.

Rajeshkumar e Omar [25] propuseram uma taxonomia dos métodos de avaliação de UX com o objetivo de auxiliar pesquisadores e profissionais na hora de escolher um método que mais se adeque às suas necessidades. Foram identificados 89 métodos presentes na literatura. Entre os parâmetros utilizados pelos autores para categorizar os métodos estão: (a) tipo de pesquisa/estudo, (b) fase do desenvolvimento que os métodos foram aplicados e (c) período da experiência avaliado. Os autores identificaram que a UX abrange todo o processo de desenvolvimento de software e, por isso, é importante combinar o interesse de diferentes *stakeholders*. Assim, é preciso considerar que apesar de ser importante capturar os sentimentos do usuário, é necessário identificar quais aspectos do software podem ter causado uma UX negativa para que a equipe de desenvolvimento do software possa identificar onde é necessário fazer adequações.

Portanto, é possível verificar que existem algumas lacunas em relação ao desenvolvimento de técnicas de UX, como necessidade de técnicas que sejam fáceis de usar e que produzam resultados úteis [31], que sejam para contextos específicos [27] e que atendam ao interesse de diferentes *stakeholders* [25]. Com o objetivo de preencher estas lacunas, desenvolveu-se a UX-Tips.

3 Apresentação da primeira versão da UX-Tips

Para fundamentar o desenvolvimento da UX-Tips, foram coletadas todas as dimensões presentes nos métodos identificados no trabalho de Vermeeren et al. [31] e feita uma análise integrativa. Esta análise teve por objetivo integrar em uma mesma dimensão todas as dimensões que tinham nomes diferentes, mas que, por definição, avaliavam os mesmos aspectos de uma aplicação. O intuito era obter um conjunto mínimo de dimensões, porém representativo, que pudesse servir de base para a UX-Tips. Todo o processo foi revisado por dois especialistas em UX.

A UX-Tips foi proposta para ser usada tanto em inspeções (avaliação por especialistas) quanto em testes com usuários. Na inspeção, o inspetor utiliza um conjunto de heurísticas para auxiliá-lo durante o processo de avaliação para possibilitar a identificação de problemas [14], sendo comumente adotado na indústria devido ao seu melhor custo/benefício em comparação a

testes envolvendo usuários [6]. Além disso, a maioria dos métodos de avaliação de UX foca na captura de emoções e sentimentos do usuário, sem identificar os problemas que levaram a uma experiência negativa [27], o que pode não deixar claro para as equipes de desenvolvimento de software o que precisa ser corrigido ou melhorado no produto. Nesse contexto, a UX-Tips possibilita a identificação de uma ampla gama de problemas que podem afetar a UX de aplicações de software, ao mesmo tempo em que permite ao avaliador relatar as emoções provocadas durante a interação com o produto.

Na inspeção, o inspetor deverá percorrer a interface do produto considerando os itens avaliativos propostos pela UX-Tips. Caso perceba que um dos itens da técnica não é atendido, o inspetor preenche um formulário, relatando o item em questão, a atividade que estava sendo executada e o ponto na interface em que o problema foi percebido, bem como uma descrição do problema do seu ponto de vista. Durante o teste de usuário, esse mesmo processo deve ser seguido. Contudo, o percurso pela interface é feito por um usuário, com o apoio de um moderador. O moderador é responsável por registrar os problemas encontrados.

A primeira versão da UX-Tips foi proposta com 17 dimensões, todas baseadas nas publicações apresentadas no trabalho de Vermeeren *et al.* [31], e 42 itens avaliativos. A Tabela 1 apresenta duas dessas dimensões com seus respectivos itens de avaliação. A técnica completa pode ser consultada em [26].

Tabela 1 - Dimensões Facilidade de uso e Aprendizagem e Satisfação da UX-Tips v1

Dimensão Facilidade de uso e Aprendizagem	
Item	Descrição
FUA1	É fácil usar a aplicação.
FUA2	É fácil reconhecer como usar a aplicação.
FUA3	É fácil reconhecer as funcionalidades da aplicação.
FUA4	A aplicação fornece dicas ou guias de como usá-la.
Dimensão Satisfação	
STF1	A aplicação atende às expectativas do usuário.
STF2	Eu gosto de usar a aplicação de um modo geral.

3.1 Estudo de viabilidade com a UX-Tips v1

Com o objetivo de avaliar a UX-Tips, realizou-se um estudo de viabilidade que consistiu na avaliação das versões móveis de duas aplicações, o Google Keep^{TM1} e o TripAdvisor^{TM2}. A escolha das aplicações foi baseada nos seguintes critérios: (i) são disponibilizadas gratuitamente, (ii) são amplamente usadas, com mais de 100 milhões de downloads na *PlayStore*, (iii) possuem aplicativos para dispositivos Android e iOS e (iv) tem focos diferenciados, permitindo maior diferença de escopo na avaliação. Este estudo teve como objetivo verificar se a UX-Tips possibilita identificar os problemas de UX das aplicações avaliadas.

¹ Link: <https://keep.google.com/>

² Link: <https://www.tripadvisor.com.br/>

3.1.1 Planejamento

Este estudo foi realizado com 25 participantes voluntários da disciplina Interação Humano-Computador (IHC) do curso de Ciências da Computação da Universidade Federal do Amazonas. Os participantes estavam cursando a etapa final da disciplina, e, portanto, já tinham conhecimento prévio sobre UX e como realizar avaliações de UX. Eles já haviam realizado um trabalho em que avaliaram a UX de uma determinada aplicação Web. Além disso, todos receberam um treinamento prévio sobre como proceder para realizar a avaliação de UX utilizando a técnica UX-Tips. Os participantes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) indicando que estavam de acordo com o estudo. Os outros artefatos utilizados foram [26]: (i) a técnica UX-Tips, (ii) dois roteiros de atividades a serem executadas nas aplicações avaliadas e (iii) um questionário de *feedback*.

3.1.2 Execução

A execução do estudo deu-se em duas etapas. Na primeira, cada participante deveria instalar o aplicativo Google Keep. Cada participante recebeu um roteiro de atividades a serem executadas na aplicação e a técnica UX-Tips para realizar a avaliação de UX. Durante o estudo em sala de aula, os participantes atuaram como inspetores e utilizaram a UX-Tips para verificar quais problemas afetaram a UX durante a interação, sem a ajuda dos pesquisadores. Cada participante preencheu e entregou uma tabela com os problemas encontrados em suas respectivas inspeções.

Na segunda etapa, realizou-se testes com usuários. Considerando que os participantes eram do curso de computação e já usavam o aplicativo Google Keep, eles foram considerados como usuários experientes (usam aplicativos regularmente). Portanto, eles deveriam selecionar um usuário com perfil inexperiente (não usa aplicativos regularmente) para o teste. Em relação ao TripAdvisor, os participantes deveriam realizar dois testes de usuários, sendo um com perfil experiente e um com perfil inexperiente. Todos os usuários deveriam ser externos à turma.

Na avaliação do TripAdvisor, os participantes do estudo atuaram como moderadores, verificando junto aos usuários quais foram as dificuldades encontradas ao interagir com as aplicações. Após a realização da avaliação de UX, os moderadores responderam ao questionário de feedback sobre o uso da UX-Tips.

3.1.3 Análise

Após a execução, as tabelas de discrepâncias individuais foram integradas em uma única lista. Neste estudo, discrepância é todo relato fornecido pelos participantes em suas avaliações. Em seguida, realizou-se uma análise que verificou quais discrepâncias eram problemas reais (que devem ser corrigidos na aplicação), quais eram falso-positivos (não representavam problemas reais na aplicação) e quais eram duplicadas (diferentes discrepâncias que representam o mesmo problema). Um pesquisador elaborou uma tabela incluindo todas as discrepâncias, indicando a classificação

de cada uma. A tabela então foi revisada por dois outros pesquisadores com experiência em avaliações de UX.

3.2 Resultados do estudo de viabilidade

Nesta seção, os resultados foram divididos em três subseções: (i) resultados sobre o aplicativo Google Keep™, (ii) resultados sobre o TripAdvisor™, e (iii) resultados sobre a técnica UX-Tips.

3.2.1 Resultados do Google Keep™

Foi encontrado um total de 74 problemas únicos, com média de 14 problemas por participante. A quantidade mínima de problemas encontrado por um único participante foi 5 e a máxima foi 24. Os principais problemas encontrados foram (i) a dificuldade em reconhecer que função desempenha determinado ícone na tela da aplicação (47 relatos); (ii) a dificuldade em verificar como salvar uma nota, pois a aplicação não apresenta claramente esta funcionalidade (43 relatos); (iii) a falta de inovação, tornando a aplicação comum (26 relatos); (iv) a falta de dicas ou guias sobre como utilizar os recursos disponíveis (21 relatos), o que pode influenciar na aprendizagem de uso da aplicação; (v) a falta de uma experiência prazerosa de uso (20 relatos).

Também foi feita uma análise qualitativa dos dados com os relatos fornecidos pelos participantes para cada problema. Quando observadas as descrições na Tabela 2, é possível verificar que os problemas apontados influenciaram na experiência de uso. Na Tabela 2 estão presentes relatos tanto da inspeção (ex.: P11) quanto dos testes com usuários (ex.: P08). Os relatos dos usuários dos testes foram relacionados com os respectivos moderadores.

Tabela 2 - Descrição dos Problemas no Google Keep

Problema: Aplicação não é legal de ser usada.	
P08	"Devido a algumas dificuldades, o usuário não achou legal usar a aplicação, inclusive mencionou que usa uma melhor"
P11	"Nem sempre é legal de usar, pois tive dificuldade em uma tarefa que não costumo usar"
Problema: Não foi prazeroso usar a aplicação.	
P01	"Não achei prazeroso, na verdade fiquei bastante chateada por não conseguir fazer o que a tarefa pedia logo que comecei a usar"
P11	"Quando fiz as tarefas que tinha que fazer, não foi prazeroso, pois demorei para achar"
Problema: Não há dicas sobre como utilizar a aplicação.	
P07	"É estressante não saber o que fazer"
P17	"Não saber exatamente comandos ou o que fazer deixa a interação estressante"

3.2.2 Resultados do TripAdvisor™

Os resultados geraram um conjunto de 68 problemas únicos, com média de 9,76 problemas por participante. A quantidade mínima de problemas encontrado por um único participante foi 2 e a máxima 27. Os principais problemas encontrados pelos

participantes foram (i) como escolher um restaurante perto de uma atração (32 relatos); (ii) a aplicação não fornece dicas ou guias de uso inicial, apresentando suas funcionalidades (29 relatos); (iii) a falta de características inovadoras na aplicação que permitissem melhorar a experiência (12 relatos); (iv) não há formas claras para consultar o preço de determinados produtos em determinados estabelecimentos (11 relatos); e (v) algumas funcionalidades da aplicação não são tão intuitivas (9 relatos).

Também foi feita uma análise qualitativa das respostas dos participantes para cada problema reportado. Com relação ao problema mais citado, a dificuldade em discernir qual ícone permitia ir para os restaurantes próximos fez com que o participante P05 não se sentisse feliz. Mais exemplos de como os problemas afetaram a UX são mostrados na Tabela 3.

Tabela 3 - Descrição dos Problemas no TripAdvisor

Problema: Escolher um restaurante.	
P05	“Não me senti feliz ao ter que fazer um caminho maior para visualizar o restaurante perto de um lugar [atração]”
Problema: Não há dicas ou guias na aplicação.	
P21	“O usuário ficou aborrecido quando não conseguia realizar operações na aplicação”
P18	“Fiquei inquieto, não sabendo usar o aplicativo”
Problema: Não apresentar características inovadoras.	
P19	“O aplicativo é monótono e não me propôs nenhum tipo de diversão para usá-lo”
P22	“A aplicação não tem nada de inovador e o usuário quis ver outras coisas”

3.2.3 Resultados da análise de viabilidade

Para verificar a percepção dos participantes em relação à técnica UX-Tips e identificar oportunidades de melhorias, foram analisadas as respostas no questionário de *Feedback*. De 25 participantes do estudo, 15 responderam. Portanto, os dados da Tabela 4 estão relacionados a estes participantes.

Tabela 4 - Visão geral das respostas do Feedback

Id e resumo da questão	Pos	Neg	Neu
Q1: A técnica é apropriada para avaliar UX?	9 (60%)	5 (33%)	1 (7%)
Q2: A técnica possibilitava a descrição dos problemas?	8 (53%)	3 (20%)	4 (27%)

Como a Tabela 4 ilustra, embora a maioria dos participantes tenha considerado a técnica apropriada para avaliar a UX, possibilitando a descrição dos problemas, algumas avaliações foram negativas ou neutras, indicando oportunidades de melhoria. A última pergunta do questionário de *feedback* contribuiu para uma melhor compreensão de tais oportunidades, pois solicitava aos participantes que fornecessem sugestões de melhorias que poderiam ser aplicadas na UX-Tips. Uma das mudanças sugeridas

envolvia a redução da quantidade de itens da técnica, considerada excessiva. A outra mudança era a remoção de itens que eram similares, ou seja, pareciam avaliar o mesmo aspecto. Tais itens tornavam a avaliação cansativa e confusa, como ilustra a citação de um dos participantes: “A técnica apresenta muitas perguntas redundantes, o que faz com que o usuário fique frustrado e se sinta menos incentivado a responder as outras perguntas” – P02. Desta forma, a UX-Tips passou por uma reavaliação que considerou a possibilidade de haver itens e dimensões sobrepostas.

4 Segunda versão da UX-Tips

4.1 Redesign e refinamento da UX-Tips v2

Para propor a nova versão, o processo de reformulação consistiu de duas etapas. Primeiramente, foi feita uma análise para verificar se diferentes dimensões poderiam ser agrupadas sem perder a completude. Essa análise foi baseada na definição das dimensões, ponderando se diferentes dimensões focavam no mesmo aspecto. Com isso, esperava-se resolver o problema da quantidade de itens, relatado na Seção 3.2.3. Na segunda etapa, o processo se repetiu. Entretanto, a análise foi feita nos itens de cada dimensão, utilizando uma matriz de confusão. O objetivo era lidar com itens que pareciam avaliar o mesmo aspecto da aplicação.

Na matriz de confusão criada, em uma dimensão foram colocados os problemas únicos obtidos no estudo de viabilidade e na outra os problemas duplicados relacionados a cada problema único. Em ambas, os problemas foram enumerados com o item da técnica UX-Tips em que foram classificados. Assim, seria possível obter quantos e em quais itens diferentes um problema único foi classificado. Considerando que se um mesmo problema foi classificado com dimensões/itens diferentes, provavelmente a descrição destas dimensões/itens estaria confusa, levando o usuário/avaliador a classificar de diferentes formas. A Figura 1 apresenta um fragmento da matriz de confusão. Para efeitos de análise, criou-se códigos para identificar cada problema, sendo “K” para Keep e “T” para TripAdvisor.

		Duplicados	Estética					Usabilidade		Fac. de uso e Aprendizagem				Esforço Cognitivo	
			EST1	EST2	EST3	EST4	EST5	USB1	USB2	FUA1	FUA2	FUA3	FUA4	CG01	CG02
Usabilidade	USB1	PP15-006K (Menus diferentes)						1							
		PP6-003T (Interface confusa)			2			2	1	3				1	
	USB2	PU3-004K (ícone não reconhecido)	7	4	2			1	3	4	3	13	2	2	13

Figura 1 - Parte da Matriz Confusão usada para análise

Na Figura 1, o problema do Keep™ “PU3-004K” foi classificado com o item USB2 (“Usabilidade”) da técnica UX-Tips,

na coluna de problemas únicos. Porém, houve outras ocorrências do mesmo problema classificadas com outros itens e dimensões, com destaques para: (i) EST1 com 7 ocorrências; (ii) FUA3 com 13 ocorrências; e (iii) CGO2 com 13 ocorrências. Esse é um exemplo de uma possível sobreposição de itens/dimensões. Isto foi investigado na segunda etapa da análise.

A segunda etapa consistiu em investigar se a dimensão e o item que foram usados para classificar um mesmo problema estavam relacionados. Para tanto, foi analisado se a descrição dos itens estava parecida e se diferentes dimensões estariam avaliando as mesmas características. Esta etapa envolveu seis iterações, onde em cada iteração houve mudanças na técnica UX-Tips. Estas mudanças estavam relacionadas à melhoria na descrição dos itens, realocação de itens dentro das dimensões ou a junção de algumas dimensões consideradas sobrepostas, como por exemplo a junção da dimensão emoção com atratividade e a realocação de alguns itens da dimensão estética para a dimensão facilidade de uso e aprendizagem. Para apoiar esta análise, além do autor desta pesquisa, outros três pesquisadores com experiência em UX, revisaram e sugeriram mudanças na técnica. Durante as seis iterações para propor a nova versão da técnica foram feitas reuniões para discussão sobre cada mudança realizada.

Tabela 5 - Facilidade de uso e Aprendizagem da UX-Tips v2

Dimensão Facilidade de uso e Aprendizagem	
Item	Descrição
FUA 1	A interface da aplicação é consistente (ou seja, mesmos itens da interface representam as mesmas coisas).
FUA2	O conteúdo da aplicação (textos, imagens, informações, ícones) são apresentados de forma visível e compreensível.
FUA3	As funcionalidades da aplicação fazem o que elas aparentam fazer.
FUA4	A aplicação é fácil suficiente para realizar as atividades sem dificuldades.
FUA5	A aplicação disponibiliza de maneira visível dicas ou guias de como usá-la.

Enquanto a primeira versão da UX-Tips tinha 17 dimensões e 42 itens avaliativos como descrito na Seção 3, após as mudanças realizadas na técnica a nova versão foi proposta com 13 dimensões e 29 itens avaliativos. Isso corresponde a quatro dimensões e 13 itens avaliativos a menos do que na primeira versão da técnica. A Tabela 5 apresenta a dimensão Facilidade de uso e Aprendizagem da UX-Tips, que sofreu mais alterações. A versão 2 da UX-Tips pode ser consultada em [26].

4.2 Estudo Comparativo com a UX-Tips v2

Para avaliar a nova versão da UX-Tips, foi realizado um estudo comparativo com outra técnica da literatura: a Inspeção Heurística

Integrativa (IHI) [29]. A IHI foi selecionada a fim de tornar a comparação mais justa, pois apresenta características semelhantes à UX-Tips, como: (i) ser uma técnica de inspeção, com itens em um formato de checklist, e (ii) considerar que UX não é apenas a avaliação dos aspectos emocionais, mas de todos os aspectos relevantes para a experiência, como sugerido por Hassenzahl [10].

O estudo envolveu 68 participantes divididos em dois grupos de mesmo tamanho, sendo que cada grupo usou uma técnica. Do total de participantes, 54 são da Universidade Tecnológica Federal do Paraná (UTFPR), e na época do estudo cursavam a disciplina de Interação Humano-Computador (IHC), e 14 são de uma empresa júnior denominada Haken e sediada na UTFPR de Campo Mourão. Os integrantes já tinham cursado a disciplina de IHC.

4.2.1 Planejamento

Todos os participantes responderam a um questionário de caracterização antes da realização do estudo, que foi usado para balancear os grupos e evitar viés de seleção. A caracterização considerou o nível de experiência com desenvolvimento de software na indústria e em avaliações de usabilidade e de UX.

Neste estudo, como o objetivo era comparar as técnicas, foi selecionado apenas o aplicativo TripAdvisor™ para ser avaliado. Os materiais usados são praticamente os mesmos do estudo anterior (ver Seção 3.2.1). A técnica repassada aos participantes dependia do grupo em que foram alocados. Adicionalmente, eles receberam também uma tabela para reportar os problemas encontrados. Depois da avaliação, todos os participantes foram convidados a preencher um formulário de *feedback* [26].

4.2.2 Execução

O estudo foi executado em três sessões, em três dias diferentes. Antes da avaliação, os participantes foram treinados nos conceitos de usabilidade e UX, bem como nas diferenças entre essas duas abordagens. Dois exemplos práticos sobre como fazer avaliação de UX usando as duas técnicas também foram apresentados, sem identificar cada técnica individualmente, para evitar qualquer viés. Em seguida, os participantes realizaram a avaliação.

Tabela 6 - Desempenho por Técnica

Técnica	Discr.	FPs	Dupl.	PU	Tempo total
UX-Tips	320	6	208	69	30h25min
IHI	285	41	158	34	30h08min

Depois da avaliação, todas as discrepâncias foram reunidas em uma lista única, como recomendado por Fernandez *et al.* [5]. Como mostra a Tabela 6, as discrepâncias (Discr.) foram classificadas como (i) falso-positivos (FP), ou seja, que foram reportadas como problemas, mas não eram; (ii) duplicadas (Dupl.),

ou seja, que indicavam o mesmo problema; ou (iii) problemas reais. Um total de 120 problemas únicos (PU) foram identificados. Desses, 17 foram encontrados por ambas, 69 foram identificados apenas pelo grupo da UX-Tips, e 34 apenas pelo grupo da IHI.

4.2.3 Análise

Uma análise quantitativa sobre a eficiência e eficácia das técnicas foi feita, considerando as definições dadas a essas métricas em estudos anteriores sobre avaliações de usabilidade ou UX [5][30]. A eficiência foi calculada como a taxa entre o total de problemas encontrados por todos os participantes pelo tempo total gasto para a realização do processo de inspeção. A eficácia foi calculada como a média da razão entre a quantidade de problemas encontrados individualmente pelo total de problemas conhecidos da aplicação. As seguintes hipóteses foram testadas:

- **H01:** Não há diferença em termos de eficiência entre as técnicas UX-Tips e IHI;
- **HA1:** Há uma diferença em termos de eficiência entre as técnicas UX-Tips e IHI;
- **H02:** Não há diferença em termos de eficácia entre as técnicas UX-Tips e IHI;
- **HA2:** Há uma diferença em termos de eficácia entre as técnicas UX-Tips e IHI.

Para avaliar as hipóteses, testes estatísticos foram conduzidos com o uso do IBM SPSS v24³. A eficácia e eficiência foram avaliadas por meio do teste de normalidade *Kolmogorov-Smirnov* (tamanho da amostra > 50), a fim de identificar qual seria o teste estatístico correto a utilizar para cada métrica [32]. Caso a distribuição fosse normal ($p\text{-value} > 0,05$) para os dados dos dois grupos, então o teste *t de Student* era aplicado. Caso contrário, o teste não paramétrico de *Mann-Whitney* era aplicado. O tamanho do efeito foi calculado usando o *d de Cohen* [2] (intervalo de confiança de 95%) e a diferença entre as médias dos grupos. Tanto o *d de Cohen* quanto a diferença entre as médias são consideradas como métricas de significância prática dos resultados [18].

Também foi feita uma análise qualitativa dos dados, que incluiu (i) as descrições dos problemas encontrados, usadas para complementar a avaliação de UX; e (ii) as respostas dadas ao questionário de *feedback*, usadas para avaliar a percepção dos participantes sobre a técnica usada. Tais percepções foram classificadas como positivas, negativas ou neutras.

4.3 Resultados do Estudo Comparativo

4.3.1 Análise quantitativa

Conforme a Tabela 6, o número de discrepâncias reportadas pelo grupo da UX-Tips foi maior do que do grupo da IHI. Portanto, a UX-Tips possibilitou aos participantes encontrar mais problemas.

Além disso, participantes que usaram a UX-Tips também reportaram um maior número de problemas únicos, o que pode sugerir que, apesar das duas técnicas terem quantidades parecidas de itens (IHI tem 28 itens), as dimensões propostas pela UX-Tips levam à busca de problemas fora do escopo das dimensões da IHI.

Além disso, a quantidade de FPs encontrados com a UX-Tips foi menor – seis casos contra 41 reportados com o uso da IHI. Isso pode sugerir que a UX-Tips possibilitou aqueles que a usaram serem mais precisos ao reportar problemas. Precisão em inspeções pode ser muito importante antes e depois das avaliações, pois economiza-se o tempo que seria gasto para reportar e compreender questões que não são problemas. Houve também um número maior de discrepâncias duplicadas no grupo que usou a UX-Tips, indicando que a técnica levou participantes diferentes a encontrarem mais frequentemente o mesmo problema, dando maior credibilidade para os resultados da avaliação.

Como descrito na Seção 4.2.3, foram testadas hipóteses sobre a eficiência e eficácia das técnicas. A Figura 2 mostra os *boxplots* referentes a essas duas métricas para cada técnica. Os *boxplots* sugerem que os participantes que usaram a UX-Tips foram mais eficazes e eficientes do que os que usaram a IHI.

Foi verificado que os dados de eficiência seguiam a distribuição normal ($p\text{-value} = 0,162$ para UX-Tips e $p\text{-value} = 0,200$ para IHI), enquanto os dados de eficácia não seguiam ($p\text{-value} = 0,014$ para UX-Tips e $p\text{-value} = 0,000$ para IHI). Desta forma, para os dados de eficiência foi aplicado o teste paramétrico *t de Student* [16] para amostras independentes, enquanto para os dados de eficácia foi aplicado o teste *Mann-Whitney* [19].

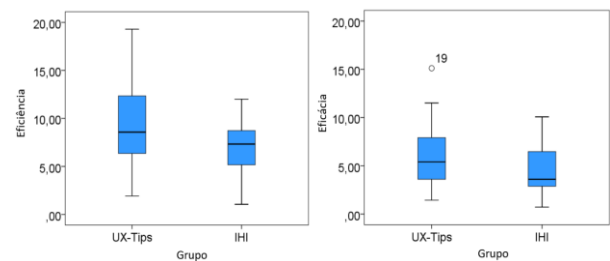


Figura 2 - Boxplots de eficiência e eficácia.

Quando as duas amostras foram comparadas em relação à eficiência, diferenças estatísticas significativas foram encontradas ($p\text{-value} = 0,003$), o que leva à rejeição da hipótese nula H01 e à aceitação da hipótese alternativa HA1. Assim, há uma diferença significativa em termos de eficiência entre as técnicas UX-Tips e IHI, sugerindo que nesse experimento a UX-Tips foi mais eficiente na descoberta de problemas de UX do que a IHI. O tamanho do efeito foi médio ($d = 0,76$), considerando a classificação definida por Cohen [2]. Além disso, a diferença das médias foi de 2,74 problemas/hora, isto é, os participantes que usaram a UX-Tips

³ Link: <https://www.ibm.com/br-pt/products/spss-statistics>

encontraram de dois a três problemas a mais do que os que usaram a técnica IHI para cada hora dedicada à avaliação de UX.

Em relação à eficácia, também foram encontradas diferenças estatísticas significativas ($p\text{-value} = 0,044$), levando à rejeição da hipótese nula $H02$. Ou seja, há uma diferença em termos de eficácia entre as técnicas UX-Tips e IHI (HA2). Isso sugere que o grupo que usou a UX-Tips encontrou mais problemas do conjunto total de problemas existentes do que o grupo que usou a IHI. O tamanho do efeito foi médio ($d = 0,52$) e a diferença de médias foi de 1,48%. Esse último resultado deve ser considerado à luz do fato de que a eficácia média dos participantes que usaram a UX-Tips foi de 5,94% contra 4,46% dos que usaram a IHI. Um dos participantes teve eficácia superior aos outros (15,11%). Apesar de indicar que não tinha experiência na indústria e experiência média em relação à usabilidade e UX, todos os relatos da sua avaliação indicavam um problema real na aplicação. Não foram identificadas características que distinguissem tal participante em relação aos demais. Segundo o professor, trata-se de um estudante muito esforçado em todos os trabalhos da disciplina.

4.3.2 Análise qualitativa

Dado o relato dos participantes sobre problemas, notou-se que os que usaram a técnica IHI dificilmente forneciam detalhes de como o problema afetou a experiência. Em geral, eles apenas apontaram o problema, sem detalhar como se sentiram. Isso pode ser um indício de que a descrição dos itens da IHI não leva o avaliador a analisar os atributos hedônicos, ao contrário da técnica UX-Tips. Desta forma, a UX-Tips pode contribuir com maiores informações para que as equipes de desenvolvimento possam identificar as causas para uma UX negativa.

Tabela 7 - Resumo quantitativo da percepção dos usuários.

Questão	Pos.	Neg.	Neu.	Técnica
Q1: A técnica é apropriada para avaliar a UX?	30 (91%)	1 (3%)	2 (6%)	UX-Tips
	22 (76%)	3 (10%)	4 (14%)	IHI
Q2: A técnica possibilitava a descrição dos problemas?	23 (70%)	4 (12%)	6 (18%)	UX-Tips
	21 (72%)	5 (17%)	3 (10%)	IHI
Q3: A técnica é fácil de usar?	22 (67%)	7 (21%)	4 (12%)	UX-Tips
	14 (48%)	12 (41%)	3 (10%)	IHI

A segunda parte da análise qualitativa foi relacionada à análise da percepção dos participantes no uso das técnicas a partir das respostas ao questionário de *feedback*. Porém, um dos participantes que usou a UX-Tips e cinco dos que usaram a IHI não responderam a este questionário. A Tabela 7 apresenta alguns dados quantitativos sobre a percepção dos participantes. As respostas foram classificadas como positivas (Pos.), negativas

(Neg.) ou neutras (Neu.). Os resultados foram similares para as duas técnicas nas duas primeiras questões. Contudo, em relação à facilidade de uso, os participantes que usaram a técnica UX-Tips tiveram uma opinião mais homogênea e positiva. Os participantes que usaram a técnica IHI ficaram mais divididos nesse quesito, levantando dúvidas sobre o quão fácil realmente é usar a técnica. A Tabela 8 apresenta exemplos de respostas dos participantes.

Tabela 8 - Respostas positivas e negativas sobre as técnicas.

Relato	
Q1	Positiva - UX-Tips: "Sim, pois a técnica possui vários tipos de avaliações e todos bem organizados facilitando a avaliação." (P47)
	Negativa - UX-Tips: "Foi um pouco complicado avaliar o usuário com a UX-Tips, pois a maioria dos fatores foi respondido com sim, tirando o Fator Facilidade de uso e Aprendizagem, que é o mais fácil de ser utilizado, pois, em geral, usuários costumam ter certo tipo de dificuldade ao utilizar aplicativos." (P03)
	Positiva - IHI: "Sim, permitiui [avaliar a UX com a técnica], principalmente os tópicos de acessibilidade e usabilidade aonde foram encontrados a maioria dos erros." (P15)
Q2	Negativa - IHI: "Não, certas partes pareceram muito específicas, alguns desnecessários, e talvez fosse melhor outras definições." (P27)
	Positiva - UX-Tips: "Sim, todos os erros que eu encontrei, eu consegui apontar itens da técnica." - (P40)
	Negativa - UX-Tips: "Não consegui indicar a falta de informações do aplicativo." (P50)
Q3	Positiva - IHI: "Sim, considero que conseguir indicar os problemas." - (P55)
	Negativa - IHI: "Não foi possível identificar o nível de satisfação do usuário. Além de não poder identificar se o usuário conseguiu completar a tarefa passada." (P22)
	Positiva - UX-Tips: "Foi fácil. As descrições estavam bem didáticas a respeito do que fazer." (P36)
Q3	Negativa - UX-Tips: "Não, pois não explicava direito as técnicas." (P53)
	Positiva - IHI: "Foi fácil de usar e entender a técnica." (P55)
	Negativa - IHI: "Não, achei difícil a compreensão." (P56)

4.4.3 Ameaças à validade

Nesse estudo foram identificadas três ameaças à **validade interna**. Uma delas é o efeito de treinamento, que poderia ser de qualidade inferior entre as técnicas. Essa ameaça foi controlada com a criação de um treinamento único, incluindo exemplos das duas técnicas, mas sem nomeá-las. Outra ameaça é que a classificação dos participantes por nível de experiência foi baseada

em auto relato, o que pode levar a imprecisões. Finalmente, houve também a ameaça relacionada à medição de tempo para o cálculo da eficiência. Para lidar com ela, foi solicitado aos participantes que fossem precisos. O moderador das sessões também verificou o tempo na entrega do material preenchido após a sessão.

Em relação às ameaças à **validade externa**, a principal questão é a representatividade da amostra, visto que a maioria dos participantes eram da mesma instituição. Para reduzir o efeito desta ameaça, foram recrutados também membros de uma empresa júnior. Contudo, os resultados do estudo podem não ser generalizados para outros tipos de software e técnicas.

Uma das principais ameaças à **validade da conclusão** é relacionada à coleta de dados, que foi tratada com a aplicação do mesmo procedimento em cada sessão individual do experimento e com o uso das mesmas fórmulas no cálculo das medidas.

Em relação a ameaças à **validade do constructo**, pode haver questões em relação às métricas utilizadas para a análise quantitativa. Neste caso, usou-se métricas típicas em experimentos de inspeção e avaliações de usabilidade [5][18].

5 Discussão

Um dos objetivos propostos no desenvolvimento da técnica UX-Tips é que, em vez de apenas classificar a UX de um aplicativo como positiva ou negativa, a técnica contribuisse com a identificação de como os problemas de UX encontrados afetavam a experiência do usuário. Com isso, espera-se fornecer maiores esclarecimentos práticos sobre problemas pontuais para as equipes de desenvolvimento de software sobre os problemas de UX. Considerando os resultados da comparação, foi possível notar que esse objetivo da técnica foi satisfeito.

Com isso, já na primeira versão, a técnica permitiu a captura de problemas relacionados a propriedades pragmáticas (relacionadas ao uso) e hedônicas (emoções) dos aplicativos avaliados, dois aspectos que são relevantes na UX [10]. Contudo, o estudo de viabilidade demonstrou que havia oportunidades de melhoria para a técnica. A principal queixa dos participantes, relacionada à quantidade excessiva de itens da técnica, foi sanada na segunda versão. No geral, a proporção de avaliações positivas aumentou de uma versão da técnica para a outra. Desta forma, o *redesign* da técnica permitiu manter a identificação de problemas de UX, porém com um menor número de dimensões e itens.

Para a indústria, é importante identificar uma quantidade maior de problemas, com menor tempo e menores custos. Dados os resultados quantitativos do estudo comparativo como um todo, os participantes que usaram a UX-Tips conseguiram ter um desempenho melhor em comparação com aqueles que usaram a IHI em todas as métricas analisadas. Embora a diferença na eficácia entre as técnicas seja pequena, a UX-Tips permitiu aos participantes encontrarem mais defeitos em menos tempo. Além

disso, os participantes a consideraram mais fácil de usar, o que é fundamental para aumentar as chances de adoção em ambientes reais de desenvolvimento. Adicionalmente, a flexibilidade de ser empregada tanto em inspeções quanto em testes com usuários pode possibilitar a identificação de uma maior variedade de problemas, sendo útil para que a equipe de desenvolvimento identifique as melhorias que podem ser feitas na aplicação.

Em relação ao desenvolvimento de novas técnicas de avaliação de UX, o estudo comparativo dá indícios de que a descrição dos itens deve ser direta. O alto número de FP encontrados pelo grupo que usou a técnica IHI é um indicativo desse fato, pois ao analisar a descrição dos problemas classificados como FP, ficou claro como os itens da técnica poderiam levar a interpretações incorretas. Isso também pode dificultar a avaliação feita por pessoas menos experientes.

Os aspectos práticos do uso da técnica, como recursos e habilidades requeridas, aplicabilidade dos resultados e facilidade de uso, também são importantes [31]. Além de ter sido considerada mais fácil de usar do que a técnica IHI, a UX-Tips também evocou mais descrições sobre a experiência do avaliador com o uso do aplicativo, o que torna seus resultados mais esclarecedores em termos do que realmente precisa ser modificado e melhorado nos aplicativos. Tais características podem ter influenciado para que a UX-Tips tenha sido melhor avaliada.

6 Conclusões

Neste artigo foi apresentada a técnica de avaliação de UX denominada UX-Tips, bem como dois estudos para avaliá-la. O estudo de viabilidade permitiu concluir que a técnica atende ao seu principal objetivo, pois favorece uma avaliação que aponta os problemas que podem ter levado o usuário a uma UX negativa. O estudo comparativo permitiu concluir que a UX-Tips é uma técnica competitiva frente a outra semelhante da literatura, pois possibilitou uma avaliação mais eficaz e mais eficiente, além de ter sido considerada mais fácil de usar. Em geral, os resultados da UX-Tips se mostraram úteis para identificar os problemas e entender os motivos de uma UX negativa, contribuindo para que a equipe de projeto saiba que melhorias podem ser feitas na aplicação para que ela forneça uma boa UX.

O principal trabalho futuro a ser realizado é a condução de estudos na indústria com a técnica proposta, para avaliar a aplicabilidade da técnica por parte de profissionais em projetos de software reais. Outra possibilidade é a customização das dimensões da técnica, o que pode contribuir ainda mais com a eficiência dos avaliadores, já que dimensões não aplicáveis, dado o tipo de aplicativo, poderiam ser excluídas da avaliação.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) -

Código de Financiamento 001 e do processo 175956/2013, CNPq processos 311494/2017-0 e 423149/2016-4. Agradecemos ao grupo USES pelo apoio e aos alunos que participaram desta pesquisa.

REFERÊNCIAS

- [1] J. A. Bargas-Avila and K. Hornbæk (2011). “Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience”. In Proceedings of the SIGCHI conference on human factors in computing systems, pp. 2689-2698.
- [2] H. Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Hillsdale, New Jersey.
- [3] E.L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, J. Kort (2009) Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 719–728.
- [4] E. L. C. Law, and S. Abrahão (2014). Interplay between User Experience (UX) evaluation and system development. *International Journal of Human-Computer Studies*, 72(6), pp. 523-525.
- [5] A. Fernandez, S. Abrahão, E. Insfran, and M. Matera (2012). “Further analysis on the validation of a usability inspection method for model-driven web development”. In Proceedings of International symposium on Empirical software engineering and measurement, Lund, Sweden, pp. 153- 156.
- [6] A. Fernandez, S. Abrahão, E. Insfran, M. Matera (2013). Usability Inspection in Model-driven Web Development: Empirical Validation in WebML”. In Proceedings of 16th International Conference on Model Driven Engineering Languages and Systems, pp. 740-756.
- [7] X. Ferre, N. Juristo and A. M. Moreno (2005). Framework for Integrating Usability Practices into the Software Process. In: Bomarius F., Komi-Sirviö S. (eds) *Product Focused Software Process Improvement*. PROFES 2005. Lecture Notes in Computer Science, vol 3547. Springer, Berlin, Heidelberg.
- [8] H. M. Hassan and G. H. Galal-Edeen, G. H. (2017). From usability to user experience. In 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), pp. 216-222.
- [9] M. Hassenzahl, S. Diefenbach and A. Göritz (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22, 353–362.
- [10] M. Hassenzahl. 2018. The thing and I: understanding the relationship between user and product. *Funology 2*, Springer, Cham.
- [11] ISO 9241-210 (2010). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.
- [12] ISO/IEC 25010 (2011). Systems and software engineering - SquaRE - Software product Quality Requirements and Evaluation: System and Software Quality Models).
- [13] ISO 9241-11 (2017). Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts.
- [14] G. H. J. Johannessen and K. Hornbæk (2014). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology*, 33(2), pp. 195-206.
- [15] N. Aquino, J. Vanderdonckt, N. Condori-Fernández, Ó. Dieste, and Ó. Pastor. (2010). Usability evaluation of multi-device/platform user interfaces generated by model-driven engineering. In Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (p. 30). ACM.
- [16] N. Juristo, A. M. Moreno. 2001. *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers.
- [17] N. Juristo, A. M. Moreno and M.-I. Sanchez-Segura. (2007). Guidelines for eliciting usability functionalities. *IEEE Transactions on Software Engineering*, 33(11), pp. 744-758.
- [18] V. Kampenes, T. Dybå, J. E. Hannay, and D. I. K. Sjøberg (2007). Systematic review: A systematic review of effect size in software engineering experiments. *Inf. Softw. Technol.* 49(11-12), pp. 1073-1086.
- [19] H. B. Mann, D. R. Whitney. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. In *Annals of Mathematical Statistics*, 18(1), pp. 50-60.
- [20] J. Nielsen and R. Molich (1990). Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 249-256.
- [21] D. Norman, J. Miller, and A. Henderson (1995). What you see, some of what's in the future, and how we go about doing it: HI at Apple Computer. In Conference companion on Human factors in computing systems, p. 155.
- [22] H. Petrie, and N. Bevan. (2009). *The Evaluation of Accessibility, Usability, and User Experience*. The universal access handbook, CRC Press, Hoboken, NJ, USA.
- [23] I. Pettersson, F. Lachner, A. K. Frison, A. Riener, and A. Butz (2018). A Bermuda Triangle? In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 461.
- [24] E. G. Pinheiro, L. A. Lopes, T. U. Conte, L. A. M. Zaina (2018). The contribution of non-technical stakeholders on the specification of UX requirements: an experimental study using the proto-persona technique. In Proceedings of the Brazilian Symposium on Software Engineering, pp. 92-101.
- [25] S. Rajeshkumar, R. Omar and M. Mahmud (2013). Taxonomies of user experience (UX) evaluation methods. In 2013 International Conference on Research and Innovation in Information Systems (ICRIIS), pp. 533-538.
- [26] L. C. Marques, P. G. F. Matsubara, W. T. Nakamura, I. S. Wiese, L. A. M. Zaina, T. U. Conte. (2019). Relatório Técnico. Disponível em: <https://doi.org/10.6084/m9.figshare.8111234>
- [27] L. Rivero, and T. Conte (2017). A systematic mapping study on research contributions on UX evaluation technologies”. In Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems.
- [28] J. B. da Silva Júnior, and A. H. Kronbauer (2018). A Study of Hedonic Experience Related to UX Capture Techniques. In Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems p. 4.
- [29] J. I. G. Urrutia, E. Brangier and L. Cessat (2017). “Is a Holistic Criteria-Based Approach Possible in User Experience?”. In International Conference of Design, User Experience, and Usability, pp. 395-409.
- [30] N. M. C. Valentim, J. Rabelo, A. C. Oran, T. Conte, T. and S. Marczak (2015). “A controlled experiment with usability inspection techniques applied to use case specifications: comparing the MIT I and the UCE techniques”. In 2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS), pp. 206-215.
- [31] A. P. Vermeeren, L.-C. E. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila (2010). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, pp. 521-530.
- [32] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, B. and A. Wesslén (2012). *Experimentation in software engineering*. Springer Science & Business Media.