

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/276204665>

Improving a Usability Inspection Technique Based on Quantitative and Qualitative Analysis

Conference Paper · October 2014

DOI: 10.1109/SBES.2014.23

CITATIONS

5

READS

133

2 authors:



Natasha Malveira Costa Valentim

Universidade Federal do Paraná

60 PUBLICATIONS 183 CITATIONS

[SEE PROFILE](#)



Tayana Conte

Federal University of Amazonas

283 PUBLICATIONS 2,263 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



ReComp [View project](#)



Risk Classification of End-User in Interaction with the Computing Environment [View project](#)

Improving a Usability Inspection Technique based on Quantitative and Qualitative Analysis

Natasha M. Costa Valentim

Grupo de Usabilidade e Engenharia de Software (USES)
Universidade Federal do Amazonas (UFAM)
Manaus, Brasil
natashavalentim@icomp.ufam.edu.br

Tayana Conte

Grupo de Usabilidade e Engenharia de Software (USES)
Universidade Federal do Amazonas (UFAM)
Manaus, Brasil
tayana@icomp.ufam.edu.br

Abstract— Usability is one of the most important factors that determine the quality of software product. Many usability inspection methods are gaining popularity as an effective alternative for addressing usability issues. However, usability evaluations are often avoided by developers and software companies due to their lack of knowledge and experience in the field. Usability evaluation can be expensive in terms of time and human resources; and sometimes companies' projects do not have budget to hire an expert inspector. Additionally, most of the evaluation methods can only be applied late in development, increasing correction costs. Aiming at solving this problem, we proposed a set of inspection techniques called MIT – Model Inspection Techniques for Usability Evaluation. This paper focuses on one of these techniques: the MIT2 technique that supports the inspection of mockups by software engineers. The goal of this paper is to present the evaluation and the evolution of MIT 2 (version 2), through the results of an empirical study. We discuss the quantitative and qualitative results of the result and their impact on improving the MIT2 technique. We analyzed the qualitative data using the procedures from the Grounded Theory (GT) method. Results indicated that our technique assists non-expert inspectors uncovering usability problems effectively; the MIT 2 (v2) was considered easy to use and useful by the participants of the study. Finally, the qualitative analysis points out relevant improvement opportunities.

Keywords— *usability evaluation; usability inspection; mockup; empirical studies.*

I. INTRODUÇÃO

Um dos atributos de qualidade mais importante para a aceitabilidade do software é a usabilidade [15]. Segundo a norma ISO/IEC 25010 [11], esta é definida como a “capacidade que o software tem de ser entendido, usado e aprendido, e também sua capacidade de agradar ao usuário, quando utilizado sob condições específicas”.

Alguns métodos de avaliação têm sido propostos para auxiliar na detecção de defeitos de usabilidade, porém muitas empresas de software não os têm utilizado [8]. Devido à falta de conhecimento sobre os métodos, várias empresas dependem de especialistas para realizar avaliações de usabilidade [8]. E a contratação destes especialistas aumenta os custos do projeto de desenvolvimento. Além disso, a maior parte dos métodos de avaliação só pode ser aplicada mais tarde no processo de desenvolvimento, aumentando os custos de correção [8]. Diante deste contexto, torna-se importante propor técnicas que

possam ser aplicadas pelos próprios engenheiros de software (não especialistas em usabilidade) para avaliar a usabilidade de artefatos nas fases iniciais dos projetos de desenvolvimento. Os benefícios de utilizar este tipo de técnica são: (i) apoiar os desenvolvedores a aprender sobre usabilidade e projeto de interação e (ii) diminuir os custos de avaliação de usabilidade.

Um dos artefatos construídos durante as fases iniciais do processo de desenvolvimento é o *mockup*. Segundo Luna *et al.* [13], *mockups* são artefatos usados para representar aspectos da interface do usuário servindo como esboços das aplicações. Destinam-se a ser desenvolvidos rapidamente para refletir os desejos dos clientes em termos de apresentação de forma mais substancial do que os requisitos expressos em linguagem escrita. No entanto, a qualidade destes *mockups* geralmente é avaliada tarde (quando se trata de um *mockup* funcional) ou até mesmo não é avaliada, impossibilitando a descoberta de problemas mais cedo.

Este artigo apresenta a versão 2 (v2) de uma técnica de inspeção, chamada MIT 2. Esta técnica tem o objetivo de avaliar a usabilidade através de *mockups* [25]. A MIT 2 (v2) faz parte de um conjunto de técnicas chamada *Model Inspection Techniques for Usability Evaluation* (MIT), composta por outras duas técnicas: MIT 1 (para avaliação de usabilidade em especificações de caso de uso) e MIT 3 (para avaliação de usabilidade em diagrama de atividades) [22, 23]. As MITs têm o propósito de reduzir os custos ao corrigir problemas de usabilidade de artefatos usados nas fases iniciais do processo de desenvolvimento. Elas possuem itens de verificação que guiam os inspetores a encontrar os problemas.

Para apoiar o desenvolvimento e validação da MIT 2 (v2), adotou-se a metodologia proposta por Shull *et al.* [19], de forma similar às pesquisas apresentadas em [3, 7], dividida em quatro etapas: (1) estudos de viabilidade: para determinar a viabilidade prática de aplicação da tecnologia; (2) estudos de observação: para aprimorar o entendimento dos pesquisadores em relação à aplicação da tecnologia e possibilitar seu refinamento; (3) estudo de caso em um ciclo de vida: para caracterizar a aplicação da tecnologia no contexto de um ciclo de vida de desenvolvimento; e (4) estudo de caso na indústria: para identificar se existem problemas de integração na aplicação da tecnologia proposta em um ambiente industrial. Como o propósito da MIT 2 (v2) é que seja facilmente adotável pela indústria e que os próprios engenheiros de software

possam utilizá-la para garantir a qualidade de seus *mockups*, executou-se a primeira etapa desta metodologia.

Desta forma, este artigo descreve a condução de um experimento controlado que foi realizado para verificar a viabilidade da técnica MIT 2 (v2). Este estudo mediu os indicadores de eficácia e eficiência da técnica para analisar se a MIT 2 (v2) é viável se comparada a sua técnica base, a Avaliação Heurística (AH) [17]. Ao medir estes indicadores, espera-se avaliar se a MIT 2 (v2) permite a identificação de mais problemas de usabilidade em *mockups* em um tempo razoável. Além disso, foi aplicado aos participantes um questionário para obtenção de dados qualitativos. A análise dos dados obtidos foi feita qualitativamente utilizando procedimentos baseados no método *Grounded Theory* (GT) [20], para extrair as dificuldades encontradas pelos inspetores e identificar oportunidades de melhoria.

O restante deste artigo está organizado da seguinte maneira: Seção II apresenta um breve referencial sobre métodos de avaliação de usabilidade. Seção III apresenta a técnica MIT 2 (v2) que tem o objetivo avaliar a usabilidade através de *mockups*. A Seção IV aborda sobre o estudo experimental e seus resultados quantitativos. A Seção V apresenta a percepção sobre a facilidade de uso e utilidade da MIT 2 (v2). A Seção VI apresenta os resultados qualitativos e mostra os problemas encontrados na MIT 2 (v2). A Seção VII apresentada as melhorias. Na Seção VIII se discute sobre as ameaças à validade. E por fim, a Seção IX apresenta as conclusões e trabalhos futuros.

II. AVALIAÇÕES DE USABILIDADE

Um método de avaliação de usabilidade é um procedimento que é composto por um conjunto de atividades bem definidas para coletar dados de uso com relação à interação do usuário final com um produto de software [8]. Segundo Fernandez *et al.* [8] este tipo de método coleta como as propriedades específicas de um produto de software contribuem para a obtenção de um certo grau de usabilidade. Os métodos de avaliação de usabilidade podem ser divididos em duas categorias: (1) Inspeções de usabilidade – métodos de avaliação de usabilidade baseados na avaliação de especialistas; e (2) Testes de usabilidade – métodos de avaliação envolvendo a participação de usuários.

Os testes de usabilidade levam em consideração uma ampla gama de usuários finais. Uma vez que a maioria das aplicações é desenvolvida para diferentes perfis de usuários. Este tipo de método pode não ser eficaz em termos de custos, uma vez que requer uma grande quantidade de recursos [8], como dispositivos e equipamentos. Além disso, precisam de uma implementação total ou parcial da aplicação, o que significa que as avaliações de usabilidade são realizadas geralmente nos últimos estágios do processo de desenvolvimento [10].

Por outro lado, métodos de inspeção permitem que avaliações de usabilidade sejam realizadas em artefatos das fases iniciais como *mockups* ou modelos da interface do usuário [16]. Além disso, estes métodos requerem menos recursos do que os testes de usabilidade [15] e podem ser usados por não especialistas em usabilidade [8]. No entanto, segundo Fernandez *et al.* [8], a qualidade das inspeções de

usabilidade pode estar relacionada com a qualidade das diretrizes utilizadas (itens que auxiliam na verificação). Outro fator que influencia na qualidade é a experiência do avaliador, pois quanto mais inspeções ele realiza, mais facilidade ele tem para identificar os problemas. Além disso, a interação dos usuários finais não é levada em conta na inspeção.

Devido às vantagens de se realizar inspeções de usabilidade, diferentes técnicas de inspeção de usabilidade têm sido desenvolvidas e utilizadas [8]. Um dos métodos mais empregados é a Avaliação Heurística (AH), proposta por Nielsen [17]. Este método tem o objetivo de encontrar problemas de usabilidade através de uma análise de conformidade do sistema usando heurísticas. As dez heurísticas propostas por Nielsen são: Visibilidade do estado do sistema, Concordância entre o sistema e o mundo real, Controle e liberdade ao usuário, Consistência e padrões, Prevenção de Erros, Reconhecer ao invés de lembrar, Flexibilidade e eficiência de uso, Projeto minimalista e estético, Reconhecimento, diagnóstico e recuperação de erros, e Ajuda e Documentação. Durante o seu uso, cada avaliador examina o artefato e descreve os problemas encontrados, associando-os com as heurísticas violadas. A descrição das 3 primeiras heurísticas da AH é apresentada na Tabela 1.

Tabela 1: Parte da AH [17]

Heurística 1. Visibilidade do status do sistema
O sistema deve sempre manter os usuários informados sobre o que está acontecendo, através de uma realimentação apropriada dentro de um tempo razoável.
Heurística 2. Concordância entre o sistema e o mundo real
O sistema deve utilizar a linguagem do usuário, com palavras, frases e conceitos familiares a ele, ao invés de termos orientados ao sistema. Seguir convenções do mundo real, fazendo com que a informação apareça numa ordem natural e lógica.
Heurística 3. Controle e liberdade do usuário
O sistema deve dar apoio a funções como Undo e Redo ou funções que permitam ao usuário utilizar “saídas de emergência” em caso de escolhas de funções erradas ou para sair de um estado não esperado.

III. MODEL INSPECTION TECHNIQUE FOR USABILITY EVALUATION

As MITs (*Model Inspection Techniques for Usability Evaluation*) [22] são técnicas de leitura para inspeção de usabilidade em artefatos de análise e projeto. Segundo Travassos *et al.* [21], técnicas de leitura são um tipo de técnica de inspeção que contém uma série de passos para a análise individual de um produto de software de forma a alcançar a compreensão necessária para uma tarefa específica. As MITs orientam o inspetor durante a avaliação de usabilidade em especificações de Caso de Uso (MIT 1), em *mockups* (MIT 2) e em Diagrama de Atividades (MIT 3).

A MIT 2 (v2) possui itens de verificação que servem como guia para interpretar as heurísticas de Nielsen e que facilitam na identificação dos problemas de usabilidade, quando aplicadas a *mockups*. A versão completa da MIT 2 (v2) encontra-se disponível em um relatório técnico online [25]. A Tabela 2 apresenta os itens de verificação da MIT 2 (v2)

relacionados às mesmas heurísticas da Avaliação Heurística mostradas na Seção II.

Tabela 2: Parte da MIT 2 (v2) [25] contendo itens relacionados às heurísticas da AH mostradas na Seção II.

Heurística 2BA. Visibilidade do Status do Sistema	
Item de Verificação 2BA1	Verifique se há informações textuais ou nome nos Mockups que informa em que parte do sistema o usuário se encontra.
Item de Verificação 2BA2	Verifique se há algum texto que informa ao usuário o que foi realizado após uma persistência de dados (alteração, exclusão, etc).
Heurística 2BB. Concordância entre o sistema e o mundo real	
Item de Verificação 2BB1	Verifique se os símbolos, palavras, frases, nome de botões, links e textos informativos estão expressas em uma representação facilmente compreendida pelo usuário, ou seja, possuem uma linguagem familiar;
Item de Verificação 2BB2	Verifique se as opções do sistema estão apresentadas em uma ordem natural e lógica segundo os conceitos do domínio do problema.
Heurística 2BC. Controle e liberdade ao usuário	
Item de Verificação 2BC1	Verifique se o usuário tem as opções de desfazer ou refazer algo que ele tenha escolhido;
Item de Verificação 2BC2	Verifique se o usuário tem a opção de cancelar o que está realizando, ou se há opções similares que permitam ao usuário utilizar saídas em caso de escolhas erradas ou para sair de um estado ou local não esperado.

Na segunda versão da MIT 2, a técnica foi dividida em Alto e Baixo Detalhamento. Houve esta divisão porque no primeiro estudo da técnica, alguns inspetores observaram que havia itens de verificação que não se adequavam ao tipo de *mockup* sendo avaliado. Ou seja, os *mockups* que eles estavam utilizando não possuíam alto nível de detalhamento como mensagens do sistema (mensagens de erro, textos informativos e advertências). Assim, a MIT 2 (v2) – Alto Detalhamento é utilizada para inspecionar *mockups* que apresentem mensagens do sistema. E a MIT 2 (v2) - Baixo Detalhamento para *mockups* que não apresentem estas mensagens. A vantagem de se ter essa divisão é que dependendo do tipo de *mockup* que se utilizará (com menos ou mais detalhe), pode-se selecionar qual tipo da MIT 2 (v2) se deve usar.

IV. EXPERIMENTO CONTROLADO

A fim de avaliar a versão 2 da técnica antes de ser transferida para a indústria de software, foi executado um estudo de viabilidade, usando somente o tipo da técnica com baixo nível de detalhamento. A escolha deste tipo da técnica é devido ao fato de os *mockups* utilizados não apresentarem mensagens de erro, textos informativos e advertências do sistema.

A MIT 2 (v2) foi avaliada em comparação com a AH [17], que é um método de inspeção amplamente utilizado pela indústria [9] e é o método base da MIT 2 (v2). O objetivo deste experimento é apresentado na Tabela 3 usando o paradigma GQM [1].

Tabela 3: Objetivo do experimento controlado usando o paradigma GQM.

Analisar	Versão 2 da técnica MIT 2
Com o propósito de	Caracterizar
Com respeito à	Sua eficácia e eficiência, comparada a AH
Do ponto de vista	Dos pesquisadores em engenharia de software
No contexto de	Uma avaliação de usabilidade em <i>mockups</i> por estudantes de graduação

A. Hipóteses

O estudo foi planejado e conduzido a fim de testar as seguintes hipóteses (nula e alternativa, respectivamente):

- H01: Não há diferença entre as técnicas MIT 2 (v2) e AH com relação ao indicador de eficiência.
- HA1: Há uma diferença no indicador de eficiência quando comparadas as técnicas MIT 2 (v2) e AH.
- H02: Não há diferença entre as técnicas MIT 2 (v2) e AH com relação ao indicador de eficácia.
- HA2: Há uma diferença no indicador de eficácia quando comparadas as técnicas MIT 2 (v2) e AH.

B. Contexto

O estudo foi executado com *mockups* de um projeto real. O experimento foi conduzido com estudantes de graduação do curso de Ciência da Computação (segundo semestre de 2013) da Universidade Federal do Amazonas (UFAM). Os alunos já tinham frequentado, antes do estudo, uma disciplina introdutória de “Engenharia de Software” e estavam cursando uma disciplina de “Análise e Projeto de Sistemas”.

C. Variáveis de Seleção

As variáveis independentes foram as técnicas de avaliação de usabilidade (MIT 2-v2 e AH) e as variáveis dependentes foram os indicadores de eficiência e eficácia das técnicas. Eficiência e eficácia foram calculadas para cada participante como: (a) a razão entre o número de defeitos detectados e o tempo gasto no processo de inspeção; e (b) a razão entre o número de defeitos detectados e o número de defeitos existentes (conhecidos), respectivamente.

D. Seleção de Participantes

Dezoito participantes assinaram o formulário de consentimento e preencheram um formulário de caracterização que avaliava sua experiência em avaliação de usabilidade e em desenvolvimento de software. A caracterização foi usada para categorizar os participantes com: nenhuma, baixa ou média experiência. É considerado participante com alta experiência em avaliação de usabilidade o que participou de mais de 5 projeto/avaliação de usabilidade na indústria, com média experiência o que participou de 1 a 4 projeto/avaliação de usabilidade na indústria, com baixa experiência o que participou de pelo menos um projeto/ avaliação de usabilidade em sala de aula, com nenhuma experiência aquele que não possui nenhum conhecimento prévio sobre usabilidade ou que possui algumas noções de usabilidade adquiridas através de leituras/palestras, mas sem experiência prática. O outro critério

foi definido de forma análoga. A Tabela 3 (segunda e terceira colunas) apresenta a categorização dos participantes.

E. Projeto Experimental

Participantes foram divididos em dois grupos: o grupo da MIT 2 (v2) e AH. Ambos os grupos inspecionariam os mesmos *mockups*. Os participantes foram atribuídos para cada técnica usando projeto completamente randomizado. Cada grupo foi composto por nove participantes.

F. Instrumentação

Muitos artefatos foram definidos para apoiar o experimento, tais como: formulário de consentimento e de caracterização, especificação das técnicas MIT 2 (v2) e AH, instruções para a inspeção, uma planilha para a anotação das discrepâncias identificadas e um questionário pós-inspeção. Como objeto da inspeção, foram usados dois *mockups* que fazem parte da especificação de um sistema real de um Centro de Treinamento¹. O primeiro *mockup* apresenta uma lista com as possíveis turmas cadastradas, dando a opção de refinar a busca de turmas por curso ou município, além da opção de selecionar uma turma para alteração ou exclusão da mesma. Já o segundo *mockup* representa a tela de cadastro dessas turmas. Todos os artefatos deste estudo foram validados pelos autores deste artigo e pelos desenvolvedores do sistema.

G. Preparação

Todos os participantes receberam treinamento de 2 horas sobre princípios de usabilidade. Então, para cada grupo, foi feita uma apresentação de 15 minutos para esclarecer sobre o uso da técnica. Exemplos similares de como usar as técnicas foram mostrados para ambos os grupos. Este tempo foi suficiente para apresentar as técnicas e exemplificar seu uso.

H. Execução

No início do estudo, um pesquisador agiu como moderador, sendo responsável por passar as informações da avaliação para os inspetores. Então, os participantes foram divididos em dois grupos, que ficaram alocados em salas diferentes (um grupo utilizou a MIT 2-v2 e o outro utilizou a AH). Cada participante recebeu os artefatos descritos na Subseção IV.F. Durante a inspeção, cada participante preencheu uma tabela com os defeitos encontrados. Todos os participantes devolveram a planilha contendo os possíveis defeitos e o tempo total gasto na inspeção. Eles também entregaram o questionário pós-inspeção preenchido. A atividade de detecção de defeitos foi executada individualmente por cada inspetor. Durante esta atividade, os inspetores não receberam qualquer auxílio dos pesquisadores envolvidos no estudo. Ao todo, 9 inspetores utilizaram a técnica MIT 2 (v2) e 9 inspetores utilizaram a técnica AH.

I. Discriminação

Após a execução, a lista de discrepâncias individuais dos inspetores foi integrada em uma única lista, removendo a referência do inspetor que encontrou as discrepâncias e a

técnica que ele/ela tinha aplicado. Assim, esta lista foi avaliada por um grupo formado pelo engenheiro de software autor dos *mockups* e dois pesquisadores especialistas em usabilidade. Este grupo decidiu quais destas discrepâncias eram únicas e quais eram duplicatas (discrepâncias equivalentes apontadas por mais de um inspetor). Além disso, este grupo classificou as discrepâncias relatadas em defeitos reais ou em falso-positivos.

J. Resultados Quantitativos

A Tabela 4 mostra o resultado geral da avaliação de usabilidade em *mockups*. As colunas da Tabela 4 são descritas a seguir. A primeira coluna (#) mostra o código de cada participante (P01: Participante 1, P02: Participante 2, entre outros). A segunda coluna (AU) apresenta a experiência do usuário em Avaliação de Usabilidade (N: Nenhuma, B: Baixa, M: Média). A terceira coluna (DS) mostra a experiência do usuário em Desenvolvimento de Software. A quarta coluna (# Disc.) apresenta a quantidade de discrepâncias (problemas identificados) por inspetor. A quinta coluna (# FP) mostra a quantidade de falsos positivo (problemas identificados que os inspetores pensavam serem defeitos). A sexta coluna (# Def.) apresenta a quantidade de problemas que de fato são defeitos de usabilidade. A sétima coluna (Tempo (horas)) mostra quanto tempo cada inspetor levou para realizar a inspeção nos *mockups*. A oitava coluna (Def./ Hora) apresenta quantos defeitos por hora cada inspetor encontrou. E a nona e última coluna mostra a quantidade total de defeitos encontrados através das técnicas (incluindo duplicatas - discrepâncias equivalentes apontadas por mais de um inspetor).

Tabela 4: Resumo do resultado da inspeção por participante - MIT 2 (v2) x AH

#	AU	DS	# Disc.	# FP	# Def.	Tempo (horas)	Def./ Hora	Total Def.
P01	B	N	7	1	6	0,42	14,40	MIT2 (v2) = 50
P02	B	N	5	1	4	0,42	9,60	
P03	M	B	9	2	7	0,40	17,50	
P04	B	N	9	3	6	0,37	16,36	
P05	B	N	4	0	4	0,08	48,00	
P06	B	B	13	4	9	0,50	18,00	
P07	B	N	8	4	4	0,43	9,23	
P08	B	N	8	1	7	0,45	15,56	
P09	N	N	4	1	3	0,50	6,00	
P10	B	N	9	4	5	0,32	15,79	
P11	B	N	11	5	6	0,33	18,00	AH = 48
P12	B	N	9	4	5	0,50	10,00	
P13	B	N	6	1	5	0,38	13,04	
P14	B	N	4	0	4	0,35	11,43	
P15	B	B	10	2	8	0,35	22,86	
P16	M	N	5	0	5	0,23	21,43	
P17	B	N	8	4	4	0,42	9,60	
P18	B	N	7	1	6	0,30	20,00	

No geral, a inspeção resultou em um conjunto de 35 defeitos de usabilidade, incluindo os 6 defeitos inseridos nos *mockups*. Tabela 5 apresenta as médias de eficácia e eficiência.

¹ Maiores informações relacionadas aos *mockups* avaliados não podem ser apresentadas por questões de confidencialidade.

Tabela 5: Eficácia e eficiência por técnica - MIT 2 (v2) x AH

Técnica	Total de Defeitos	Média de Defeitos	Eficácia (%)	Média de Tempo (min)	Eficiência (defeitos/hora)
MIT 2-v2	50	5,56	15,87%	23,78	14,02
AH	48	5,33	15,24%	21,22	15,08

Uma análise foi realizada usando o teste não paramétrico de Mann-Whitney [14], dado o tamanho limitado da amostra [6]. Um resumo dos resultados foi apresentado usando o gráfico de boxplot. As análises estatísticas foram executadas usando a ferramenta SPSS V. 19, com $\alpha = 0,10$. A escolha desta significância estatística foi motivada pelo pequeno tamanho da amostra usada neste experimento [18]. Figura 1 mostra o gráfico de boxplot com a distribuição da eficiência por técnica. Através da Figura 1 é possível observar que o grupo da MIT 2 (v2) teve quase a mesma eficiência que o grupo da AH. Quando as duas amostras são comparadas usando o teste de Mann-Whitney, não se encontra nenhuma diferença significativa entre os dois grupos ($p = 0,507$). Estes resultados apoiam a hipótese nula H01 que afirma que não há nenhuma diferença no indicador de eficiência de ambas as técnicas.

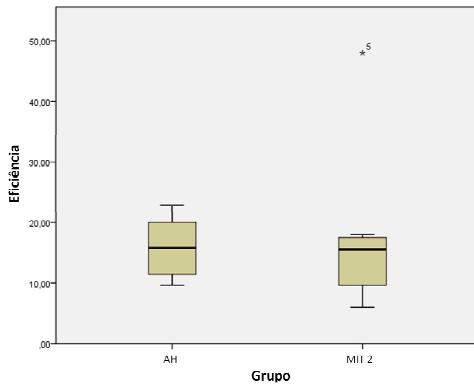


Figura 1. Boxplots para eficiência por participante - AH X MIT 2 (v2)

A mesma análise foi aplicada para determinar se havia alguma diferença significativa comparando o indicador de eficácia das duas técnicas na detecção de defeitos de usabilidade. O gráfico de boxplots com a distribuição da eficácia por técnica (ver Figura 2) mostra que o grupo da MIT 2 (v2) foi mais eficaz que o grupo da AH quando inspecionaram a usabilidade em *mockups*. Além disso, a mediana do grupo da MIT 2 (v2) está mais alta que a mediana do grupo da AH. No entanto, quando se compara as duas amostras usando o teste de Mann-Whitney, não se encontra diferença estatística significante entre os dois grupos ($p = 0,892$). Estes resultados apoiam a hipótese nula H02 que afirma que não há nenhuma diferença significativa no indicador de eficácia da MIT 2 (v2) e AH.

Uma análise de correlação foi realizada com as variáveis eficácia e eficiência, para determinar o relacionamento entre elas. O coeficiente de correlação de Spearman entre eficácia e eficiência foi de 0,597, com $p = 0,009$, mostrando correlação positiva. Portanto, participantes mais eficazes também foram mais eficientes.

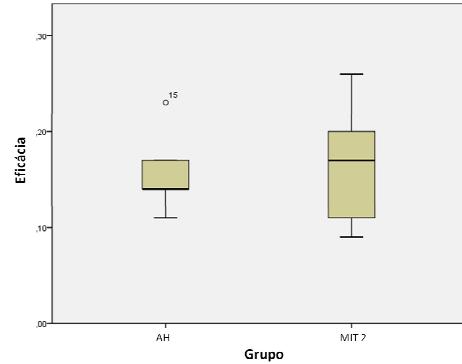


Figura 2. Boxplots para eficácia por participante – AH X MIT 2 (v2)

V. ANÁLISE DA PERCEPÇÃO SOBRE FACILIDADE DE USO E UTILIDADE DA MIT 2 (v2)

Após a análise quantitativa, foram analisados os questionários sobre a aceitação da tecnologia somente com respeito a MIT 2 (v2), pois desejava-se avaliar sua aceitabilidade. Estes questionários foram definidos com base nos indicadores do *Technology Acceptance Model* (TAM) [5]. Os dados coletados foram apresentados em formato de gráfico para fins de análise. Os indicadores definidos foram: (i) utilidade percebida, que define o grau que uma pessoa acredita que a tecnologia pode melhorar seu desempenho no trabalho, e (ii) facilidade de uso percebida, que define o grau que uma pessoa acredita que usar a tecnologia específica seria livre de esforço. A razão para focar nestes indicadores é que, de acordo com Davis [5], estes aspectos são fortemente correlacionados com a aceitação da tecnologia pelo usuário.

Participantes forneceram suas respostas em uma escala de seis pontos, baseados nos questionários aplicados por Lanubile *et al.* [12]. As possíveis respostas nos questionários são: concordo totalmente, concordo amplamente, concordo parcialmente, discordo parcialmente, discordo amplamente e discordo totalmente. Neste questionário, os inspetores responderam o seu grau de aceitação referente à utilidade e facilidade de uso.

A Figura 3 apresenta as percepções dos participantes com relação à facilidade de uso da MIT 2 (v2). O eixo X dos gráficos da Figura 3 refere-se às possíveis respostas do questionário pós-inspeção e o eixo Y refere-se à quantidade de participantes. Os códigos P01, P02 e os demais representam os participantes apresentados na Tabela 4. Pode-se notar que o participante P05, que possui baixa experiência em avaliação de usabilidade, discordou totalmente na questão “Eu entendia o que acontecia na minha interação com a MIT 2 (v2)”, mostrando que a MIT 2 (v2) em algum momento não foi fácil de entender. O participante P05 também discordou totalmente da questão “Foi fácil ganhar habilidade no uso da MIT 2 (v2)”. Já o participante P09, que não possuía experiência em avaliação de usabilidade, discordou parcialmente na questão “É fácil lembrar como usar a MIT 2 (v2) para realizar uma inspeção de usabilidade”, destacando a dificuldade de lembrar da MIT 2 (v2). No entanto, todos os inspetores concordaram com as outras questões, mostrando que no geral, a técnica MIT 2 (v2) foi considerada fácil de usar.

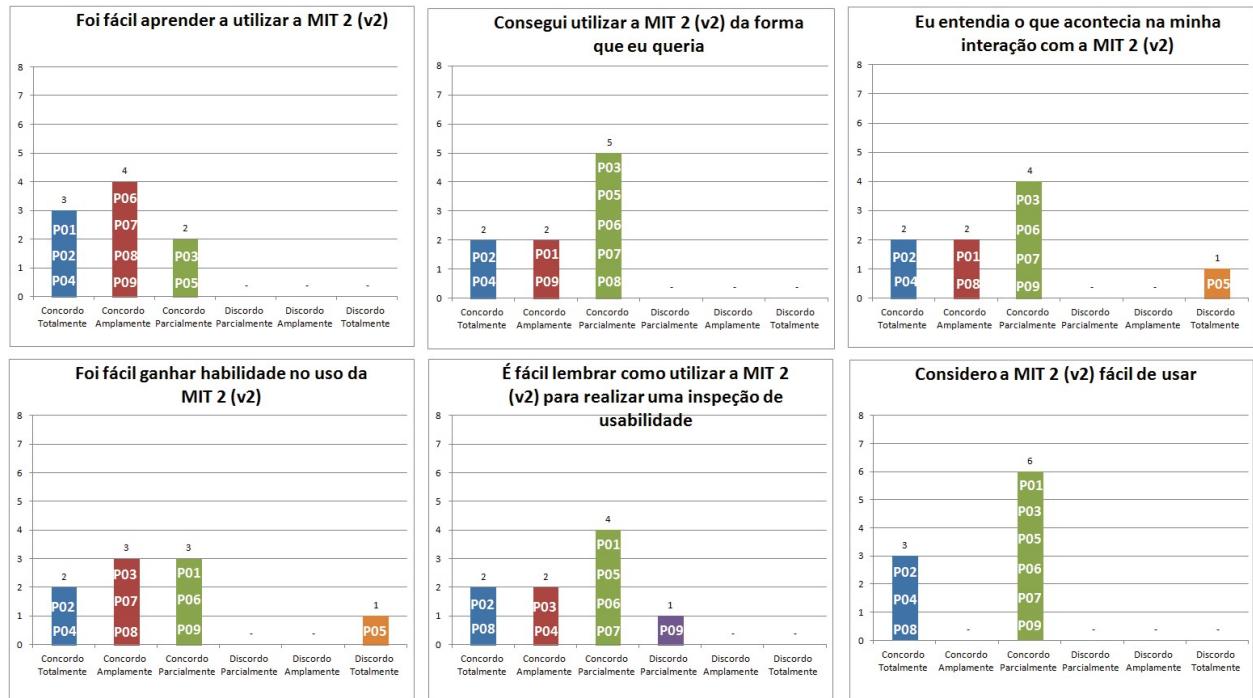


Figura 3. Percepção dos participantes sobre a Facilidade de Uso da MIT 2 (v2)

Figura 4 apresenta as percepções dos participantes com relação à utilidade da MIT 2 (v2). Verificou-se que somente o participante P01, que possui baixa experiência em avaliação de usabilidade, discordou parcialmente na questão “A MIT 2 (v2) me permitiu detectar defeitos mais rápido”, indicando que em algum ponto o uso da técnica consumiu muito tempo. Porém, todos os inspetores concordaram com as outras questões, reforçando que a MIT 2 (v2) é útil no processo de inspeção.

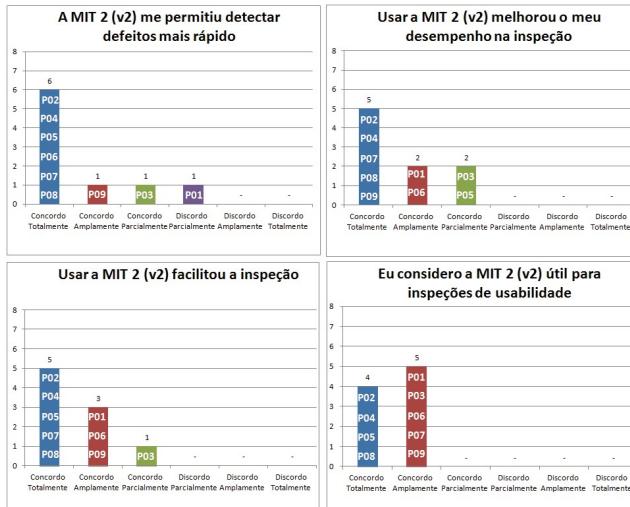


Figura 4. Percepção dos participantes sobre a Utilidade da MIT 2 (v2)

VI. ANÁLISE QUALITATIVA

Além da análise por meio do modelo TAM, foi feita uma análise específica dos dados qualitativos (comentários adicionais dos inspetores) contidos nos questionários, tendo por base o método *Grounded Theory* (GT) [20]. Como o objetivo da análise qualitativa era melhorar a técnica proposta (MIT 2-v2), só foi feita a análise dos dados da MIT 2 (v2).

Os dados qualitativos extraídos dos questionários foram analisados utilizando um subconjunto das fases do processo de codificação sugerido por Strauss e Corbin [20] para o método GT – as codificações aberta (1^a fase) e axial (2^a fase). Ao analisar os dados destes questionários, criou-se códigos (conceitos relevantes para entendimento da percepção sobre a técnica) relacionados a trechos (citações dos participantes) - codificação aberta (1^a fase). Após isso, os códigos foram agrupados de acordo com as suas propriedades, formando conceitos que representam categorias e subcategorias. Finalmente, estas foram relacionadas entre si - codificação axial (2^a fase). Os procedimentos de GT visam uma análise mais aprofundada, através da comparação e da análise da relação entre estes conceitos. O objetivo dessa análise neste estudo foi entender qual a percepção dos inspetores sobre a experiência de uso da MIT 2 (v2). Como não se pretendia criar uma teoria a esse respeito, não foi realizada a codificação seletiva (3^a fase do método GT). As etapas de codificação aberta e axial foram suficientes para entender o motivo de alguns problemas. Os conceitos referentes ao método GT são apresentados detalhadamente em [4].

Verificou-se que o conteúdo dos questionários teve maior direcionamento em relação à estrutura, ao uso e às possíveis

melhorias na MIT 2 (v2). O processo de codificação produziu no total 17 códigos que foram associados a 2 categorias. O questionário pós-inspeção possuía questões sobre a adequação e facilidade de uso da técnica, o que levou à identificação dessas duas categorias: Estrutura da MIT 2 (v2) e Feedback do Uso da MIT 2 (v2). Além disso, dentro da categoria Feedback de Uso da MIT 2 (v2) foi criada uma subcategoria chamada Sugestões de Melhorias – onde são apresentadas sugestões de modificações na MIT 2 (v2) com o objetivo de melhorar o ganho com a sua utilização.

Entre os códigos criados na categoria “Estrutura da MIT 2 (v2)” (Figura 5), três deles apontam evidências de inadequação como os códigos “O item 2BA2 é inútil se não houver mensagens de erros nos mockups” e “2BF1 e 2BF2 tem quase a mesma função para campos de análise diferentes (um para nomes e outro para figuras)”. O primeiro código foi citado

pelo participante P08, que possui baixa experiência em avaliação de usabilidade. Ele menciona diretamente um item de verificação que não convém à MIT 2 (v2) - Baixo Nível de Detalhamento – uma vez que os *mockups* com pouco detalhamento não apresentam as mensagens do que foi realizado após uma persistência de dados (alteração, exclusão, entre outros). Pode-se notar que mesmo com pouca experiência, um participante da graduação conseguiu identificar um problema na técnica. Além disso, cinco códigos da categoria “Estrutura da MIT 2 (v2)” apontam dificuldades, como os códigos: “Não está claro o sentido da palavra “terminologia” no item 2BD1” citado pelo participante P07 e “É necessário deixar mais claro o que é consistente (com a realidade ou com o sistema)” citado pelo participante P01. Estes códigos demonstram indícios das dificuldades que inspetores tiveram com alguns termos usados da técnica.

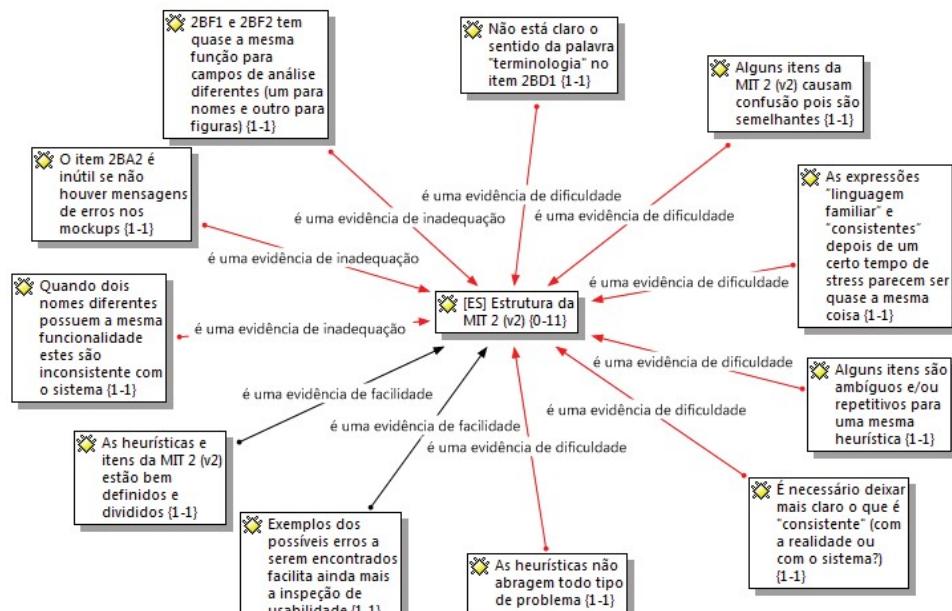


Figura 5. Categoria Estrutura da MIT 2 (v2)

A Figura 6 apresenta a categoria “Feedback do uso da MIT 2 (v2)”. Esta categoria possui seis códigos que apontam dificuldades, tais como: “Dificuldade de entender alguns itens da MIT 2 (v2) para procurar problemas” e “Dificuldade de imaginar links observados por pessoas experientes”. Através do primeiro código, citado pelo participante P07, é possível observar que este inspetor teve dificuldades em tentar entender alguns itens de verificação da MIT 2 (v2), levando a supor que estes itens ainda não estão claros o suficiente. O participante P07 possui baixa experiência em usabilidade e isto nos leva a notar que a técnica precisa estar mais clara para inspetores menos experientes. Além disso, no segundo código (citado pelo participante P01) é afirmado que se tem dificuldades de aplicar um determinado item, pois o mesmo também dificulta o entendimento. Quando se percebe que a técnica ainda possui itens difíceis de entender para inspetores menos experientes, novos esforços precisam ser empenhados a fim de que a

melhore, pois uma técnica difícil de entender, possivelmente não será usada.

A subcategoria “Sugestões de Melhoria” também contém 6 códigos associados que descrevem melhorias na MIT 2 (v2) e sugerem novas características (Figura 6). Os códigos mencionam questões referentes à diminuição do tempo de inspeção, tais como: “Sugestão de juntar alguns itens de verificação para diminuir o tempo de leitura da MIT 2 (v2)” e “Sugestão de juntar os itens de verificação 2BF1 e 2BF2”, citados pelo participante P01. Além disso, os códigos sugerem inspecionar a interação entre *mockups* (“Sugestão de inspecionar a interação entre mockups” e “Sugestão de a MIT 2 (v2) verificar caso deje de preencher um campo num formulário, o sistema deve avisar sobre o erro”). Ambos os códigos foram citados pelo participante P08. Isto mostra que mesmo participantes com baixa experiência em avaliação de usabilidade sugerem melhorias interessantes para a técnica.

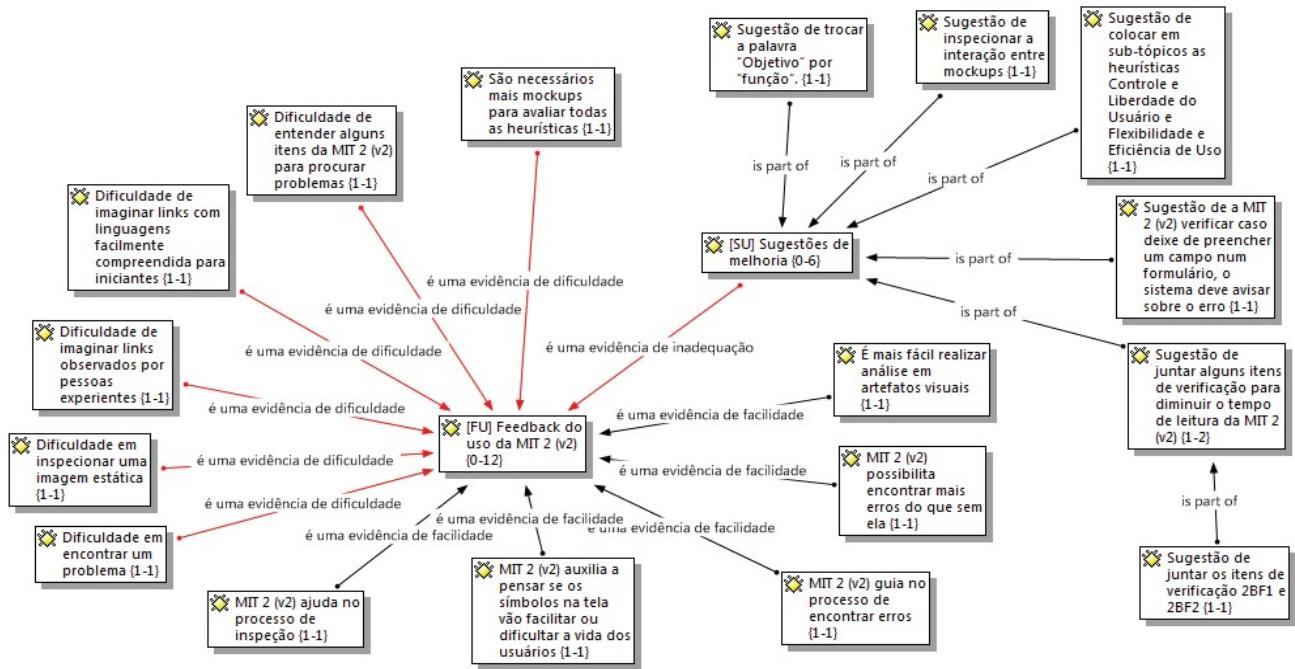


Figura 6. Categorias Feedback do uso da MIT 2 (v2) e Sugestões de Melhoria

VII. MELHORIAS NA MIT 2 (v2)

A análise qualitativa permitiu a identificação das causas para o resultado negativo obtido com a análise quantitativa. Através da análise desses dados, alterações foram feitas e uma terceira versão (v3) da MIT 2 foi elaborada. A MIT 2 (v3) está disponível em [24].

Um dos inspetores afirmou que “É necessário deixar mais claro o que é “consistente” (com a realidade ou com o sistema?)” no item de verificação 2BD1. Além disso, um inspetor comentou que “Não está claro o sentido da palavra “terminologia” no item 2BD1”. Para solucionar o primeiro problema, acrescentaram-se os termos “com o sistema representado”, para deixar mais claro com que a terminologia deve estar com consistente. Para resolver o segundo problema, adicionou-se o termo “nomenclatura” ao lado da palavra “terminologia”. Estas alterações foram realizadas no item de verificação 2BD1 (Figura 7) que pertence à heurística 2BD - Consistência e Padrões.

2BD1	Verifique se a terminologia (nomenclatura), gráficos e símbolos estão consistentes com o sistema representado;
-------------	--

Figura 7. Item de verificação 2BD1 da MIT 2 (v3)

Outro inspetor observou que “O item 2BA2 é inútil se não houver mensagens de erros nos mockups”. Através deste comentário foi constatado que de fato este item de verificação (Figura 8) não era para estar contido na MIT 2 (v2) – Baixo Detalhamento, pois esta versão da técnica não leva em consideração mensagens de erro, textos informativos e advertências. Portanto, este item de verificação foi retirado da MIT 2 (v2) – Baixo Detalhamento.

2BA2	Verifique se há algum texto que informa ao usuário o que foi realizado após uma persistência de dados (alteração, exclusão, etc.).
-------------	--

Figura 8. Item de verificação 2BA2 da MIT 2 (v2) – Baixo Detalhamento

Através do código “Dificuldade de imaginar links observados por pessoas experientes”, um inspetor expressou sua dificuldade em tentar entender como seriam links abreviados para usuários experientes presente no item de verificação 2BG6 (Figura 9). Levando em consideração que este item causa dificuldades de entendimento e que o mesmo já está contido no item de verificação 2BG7 (Figura 9), o item de verificação 2BG6 foi retirado tanto da MIT 2 (v2) – Baixo Detalhamento quanto da MIT 2 (v2) – Alto Detalhamento.

2BG6	Verifique se há nomes de links abreviados para usuários mais experientes;
2BG7	Verifique se há como as funcionalidades do sistema serem realizadas de forma mais rápida e eficiente para usuários mais experientes.

Figura 9. Item de verificação 2BG6 da MIT 2 (v2) – Baixo e Alto Detalhamento

Um dos inspetores comentou que os itens de verificação “2BF1 e 2BF2 tem quase a mesma função para campos de análise diferentes (um para nomes e outro para figuras)” e deu a “Sugestão de juntar os itens de verificação 2BF1 e 2BF2”. Para melhorar a técnica no ponto citado, uniram-se os itens de verificação 2BF1 e 2BF2 em um único item (Figura 10), ajudando assim a reduzir o tempo de leitura da técnica. A mesma alteração foi feita na MIT 2 (v2) – Alto Detalhamento.

2BF1	Verifique se os nomes das opções, campos, botões e links são informados e se os ícones/figuras minimizam o esforço físico e cognitivo do usuário de se lembrar;
-------------	---

Figura 10. Item de verificação 2BF1 da MIT 2 (v3)

Alguns códigos como “*Dificuldade de entender alguns itens da MIT 2 (v2) para procurar problemas*”, “*Alguns itens são ambíguos e/ou repetitivos para uma mesma heurística*”, e “*Sugestão de juntar alguns itens de verificação para diminuir o tempo de leitura da MIT 2 (v2)*” foram atendidos com as melhorias já relatadas.

Códigos como “*Sugestão de inspecionar a interação entre mockups*” e “*Sugestão de a MIT 2 (v2) verificar caso deixa de preencher um campo num formulário, o sistema deve avisar sobre o erro*”, ainda serão melhor analisados e possivelmente serão atendidos em versões futuras da técnica.

Outros problemas foram relatados, como “*Dificuldade em encontrar um problema*”, “*Dificuldade em inspecionar uma imagem estática*”, “*Alguns itens da MIT 2 (v2) causam confusão pois são semelhantes*”, “*As heurísticas não abrangem todo tipo de problema*”. Buscou-se fazer uma verificação abrangente na técnica para atender a estes pontos.

VIII. AMEAÇAS À VALIDADE

Como em todos os estudos, existem ameaças que podem afetar a validade dos resultados. Nesta seção, discutem-se estas ameaças, categorizando-as usando a mesma abordagem de Wohlin *et al.* [26]: interna, externa, conclusão e constructo.

A. Validade Interna

Neste experimento, consideraram-se quatro principais ameaças que representam um risco para uma interpretação inapropriada dos resultados: (1) efeitos do treinamento, (2) classificação da experiência, (3) medição de tempo e (4) influência do moderador. Pode ter havido um efeito do treinamento se o treinamento da técnica AH tivesse menor qualidade do que o treinamento da MIT 2 (v2). Controlaram-se os efeitos do treinamento preparando treinamentos equivalentes para ambos os grupos e com os mesmos exemplos de detecção de discrepâncias. Além disso, com relação à classificação da experiência dos participantes, esta foi baseada em uma autoclassificação dos participantes. Eles foram classificados de acordo com o número e tipo de experiências anteriores em avaliação de usabilidade e desenvolvimento de software. Com relação à medição de tempo, foi solicitado aos participantes serem tão precisos quanto possível, e o moderador também checou o tempo anotado por cada participante quando entregava a planilha de discrepâncias. E por fim, para diminuir a ameaça com relação à influência do moderador nos resultados do estudo, uma equipe de especialistas fez a análise das discrepâncias encontradas, julgando se eram defeitos de usabilidade ou não, sem a interferência do moderador.

B. Validade Externa

Quatro questões foram consideradas: (1) participantes foram estudantes de graduação; (2) o estudo foi conduzido em um ambiente acadêmico; (3) a validade do artefato avaliado como um artefato representativo; (4) o pesquisador inseriu alguns defeitos nos *mockups*; e (5) participantes com necessidade de treinamento. Com relação à questão 1, poucos participantes tinham experiência na indústria uma vez que eles eram somente estudantes de graduação. De acordo com Carver *et al.* [2], estudantes que não tem experiência na indústria

podem ter habilidades semelhantes aos inspetores menos experientes. Com relação à questão 2, os artefatos inspecionados (*mockups*) fazem parte do projeto de um sistema real. No entanto, não é possível afirmar que o artefato usado no estudo representa todos os tipos de *mockups* (questão 3). Com relação à questão 4, todos os problemas de usabilidade inseridos foram encontrados por ambos os grupos de participantes. Além disso, o número de defeitos encontrados pelos grupos foi muito maior que o número de defeitos inseridos pelo moderador. E por fim, com relação à necessidade do participante ter um treinamento, o ideal seria que não houvesse necessidade de treinamento. No entanto, o curto tempo de treinamento possibilita que a técnica seja usada por desenvolvedores sem muita experiência em avaliação de usabilidade.

C. Validade de Conclusão

Neste estudo, o principal problema é o tamanho e homogeneidade da amostra. A quantidade de participantes não é a ideal do ponto de vista estatístico e, além disso, os participantes são todos estudantes de uma mesma instituição. Tamanho reduzido da amostra é um problema conhecido em estudos da Engenharia de Software [9, 18]. Porém, este é um estudo inicial para verificar a viabilidade da técnica. Futuramente, pretende-se realizar novos estudos com mais participantes e que sejam da indústria. Devido a estes fatos, há limitação nos resultados, sendo estes considerados indícios e não conclusivos.

D. Validade de Constructo

As medidas de eficiência e eficácia são frequentemente usadas em estudos que investigam técnicas de detecção de defeitos [7, 9].

IX. CONCLUSÕES E TRABALHOS FUTUROS

O objetivo deste artigo foi apresentar a avaliação e evolução da técnica MIT 2 (v2), através dos resultados quantitativos e qualitativos de um estudo experimental.

A análise quantitativa do experimento mostrou que a MIT 2 (v2) teve quase a mesma eficiência que a AH [17]. No entanto, nenhuma diferença estatística significante foi encontrada. Com relação ao indicador de eficácia, o grupo que usou MIT 2 (v2) teve melhor desempenho do que o grupo que usou a AH. Porém, quando se comparou as duas amostras usando o teste de Mann-Whitney, não foi encontrada nenhuma diferença significante entre os dois grupos. Como mencionado na seção de ameaças à validade, devido ao tamanho pequeno da amostra deste estudo, não se pode considerar estes resultados conclusivos, sendo necessário repetir este estudo com uma amostra maior e mais heterogênea.

Através da análise da percepção sobre a facilidade de uso e utilidade da técnica é possível observar que, no geral, muitos inspetores concordaram com as questões relacionadas à estas percepções, ou seja, julgaram a MIT 2 (v2) útil e fácil de usar. Alguns inspetores citaram que a MIT 2 (v2) ajuda no processo de inspeção e possibilita encontrar mais defeitos do que sem ela. Estes resultados mostram evidências da facilidade de uso ao aplicar a técnica. O fato de a MIT 2 (v2) ter uma boa aceitação dos inspetores do experimento pode indicar que esta

técnica é adequada para inspetores com baixo conhecimento em usabilidade, uma vez que participantes foram alunos de graduação. Isto é um indício de que é viável treinar engenheiros de software para realizarem inspeções de usabilidade em projetos de desenvolvimento.

A análise qualitativa permitiu a identificação de problemas ao se utilizar a MIT 2 (v2) no estudo de viabilidade em questão, como: termos que não estavam claros, item de verificação que não era adequado para *mockups* com baixo detalhamento, itens de verificação que podiam ser unidos, dentre outros. Estes resultados qualitativos levaram à proposta de uma nova versão da técnica (MIT 2 - v3), contendo melhorias como: itens de verificação mais claros e junção de itens para a redução do tempo de leitura da técnica.

Como trabalhos futuros, pretende-se executar novos estudos empíricos, de forma a garantir a qualidade da técnica para sua futura transferência para a indústria. Um estudo de observação será o próximo passo visto que ele ajuda a aprimorar o entendimento dos pesquisadores em relação à aplicação da tecnologia e responder a questão “Os passos do processo fazem sentido?”.

Espera-se que os resultados apresentados neste artigo sejam úteis para a promoção e melhoria da prática atual e da pesquisa de usabilidade. Espera-se, também, que a técnica proposta ajude na avaliação de *mockups* realizados na fase inicial do processo de desenvolvimento, melhorando a sua qualidade a um custo baixo.

AGRADECIMENTOS

Os autores agradecem ao apoio financeiro da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e da FAPEAM (Fundação de Amparo à Pesquisa do Estado do Amazonas) através do Edital N° 016/2013 PROTI – PESQUISA, processo N° 062.00578/2014, aos participantes do estudo de viabilidade e aos pesquisadores do USES-UFAM pelas contribuições na execução desta pesquisa.

REFERÊNCIAS

- [1] Basili, V., Rombach, H., “The TAME Project: Towards Improvement-Oriented Software Environments”, In: IEEE Transactions on Software Engineering, v. 14, 1988, pp. 758 – 773.
- [2] Carver, J., Jaccheri, L., Morasca, S., Shull, F. “Issues in Using Students in Empirical Studies in Software Engineering Education”. In 9th International Symposium on Software Metrics (METRICS), 2003, pp. 239 – 249.
- [3] Conte, T., Massolar, J., Mendes, E., Travassos, G. “Web Usability Inspection Technique Based on Design Perspectives”, In Simpósio Brasileiro de Engenharia de Software (SBES), 2007, v. 1, pp. 394-410.
- [4] Conte, T., Cabral, R., Travassos, G. H. “Applying Grounded Theory in Qualitative Analysis of an Observational Study in Software Engineering – An Experience Report”. In INFOCOMP Journal of Computer Science, v.2, n.1, 2009, p. 58-69.
- [5] Davis, F. “Perceived usefulness, perceived ease of use, and user acceptance of information technology”. In MIS Quarterly, v. 13, n. 3, 1989, pp. 319 – 339.
- [6] Dyba T., Kampenes V., Sjoberg D. “A systematic review of statistical power in software engineering experiments”. In Information and Software Technology, v. 48, n. 8, 2006, pp. 745-755.
- [7] Fernandes, P., Conte, T., Bonifácio, B. “WE-QT: A Web Usability Inspection Technique to Support Novice Inspectors”. In 26th Brazilian Symposium on Software Engineering (SBES), 2012, v. 1. pp. 11-20.
- [8] Fernandez, A., Insfran, E., Abrahão, S. “Usability evaluation methods for the web: a systematic mapping study”. In Information and Software Technology, v. 53 (8), 2011, pp. 789-817.
- [9] Fernandez, A., Abrahão, S., Insfran, E., Matera, M. “Further analysis on the validation of a usability inspection method for model-driven web development”. In International Symposium on Empirical software engineering and measurement (ESEM), September, 2012, pp. 153-156.
- [10] Hornbæk, K., Høegh, R. T., Pedersen, M. B., Stage, J. “Use Case Evaluation (UCE): A Method for Early Usability Evaluation in Software Development”. In 11th International Conference on Human-Computer Interaction, 2007, pp. 578-591.
- [11] ISO, International Organization for Standardization, ISO/IEC 25010, Systems and software engineering -- SQuaRE - Software product Quality Requirements and Evaluation -- System and Software Quality Models, 2011.
- [12] Lanubile, F., Mallardo, T., Calefato, F. “Tool support for Geographically Dispersed Inspection Teams”. In Software Process Improvement and Practice, v. 8, 2003, pp. 217 – 231.
- [13] Luna, E. R., Panach, J. I., Grigera, J., et al. “Incorporating usability requirements in a test/model-driven web engineering approach”. In: Journal of Web Engineering, 2010, v. 9, issue 2, pp. 132 - 156.
- [14] Mann, H. B., Whitney, D. R. “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other”. In Annals of Mathematical Statistics 18, 1947, pp. 50 – 60.
- [15] Matera, M., Costabile, M. F., Garzotto, F., Paolini, P. “SUE Inspection: An Effective Method for Systematic Usability Evaluation of Hypermedia”. In IEEE Transactions on Systems, Man and Cybernetics, Part A, 32, 2002, pp. 93 – 103.
- [16] Matera, M., Rizzo, F., Carugui, G. T. “Web Usability: Principles and Evaluation Methods”. In Mendes, E., Mosley, N. (eds), Web Engineering, Chapter 5, New York, Springer Verlag, 2006, pp. 143-180.
- [17] Nielsen, J., “Heuristic evaluation”. In Usability Inspection Methods (Eds. Nielsen and Mack), John Wiley & Sons, New York, 1994.
- [18] Rivero, L., Conte, T. “Using an Empirical Study to Evaluate the Feasibility of a New Usability Inspection Technique for Paper Based Prototypes of Web Applications”. In: 26th Brazilian Symposium on Software Engineering (SBES), 2012, v. 1. pp. 81-90.
- [19] Shull, F., Carver, J., Travassos, G. H., “An empirical methodology for introducing software processes.” ACM SIGSOFT Software Engineering Notes, 2001, v. 26, n. 5, pp. 288-296.
- [20] Strauss, A., Corbin, J. “Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory”. 2 ed. SAGE Publications, London, 1998.
- [21] Travassos, G. H., Shull, F., Carver, J., Basili, V. “Reading Techniques for OO Design Inspections”. University of Maryland, 2012.
- [22] Valentim, N. M. C., Oliveira, K. M., Conte, T. “Definindo uma Abordagem para Inspeção de Usabilidade em Modelos de Projeto por meio de Experimentação”. In XI Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais (IHC), 2012, pp. 165 – 174.
- [23] Valentim, N. M. C., Da Silva, T. S., Silveira, M. S., Conte, T. “Estudo Comparativo entre técnicas de Inspeção de Usabilidade sobre Diagramas de Atividades. In XII Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais (IHC), 2013, pp. 92 – 101.
- [24] Valentim, N. M. C., Conte, T., “Relatório Técnico: Versão 3 da MIT 2”, Relatório de número 004, 2014. Available at: <http://uses.icomp.ufam.edu.br/>.
- [25] Valentim, N. M. C., Conte, T., “Relatório Técnico: Versão 2 da MIT 2”, Relatório de número 005, 2014. Available at: <http://uses.icomp.ufam.edu.br/>.
- [26] Wöhl, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., Wessl, A. “Experimentation in software engineering: an introduction”. Kluwer Academic Publishers, 2000.