

Análise de satisfação de uso de sistemas colaborativos de apoio à Engenharia de Software: um estudo comparativo

Cíntia Santos de Resende Silva Costa
Universidade Federal de Viçosa - *Campus* Florestal
Florestal, Minas Gerais
cintia.s.costa@ufv.br

Gláucia Braga e Silva
Universidade Federal de Viçosa - *Campus* Florestal
Florestal, Minas Gerais
glaucia@ufv.br

ABSTRACT

Studies show the relevance of measuring the user satisfaction of groupware used in a software process. The user satisfaction may influence product efficiency, team motivation and project results. This research proposes a comparative approach to measure user satisfaction of collaborative tools used to support the software process based on the use of the following qualitative techniques: Thinking Out Loud and Questionnaire. The work presents the results of tests involving 8 volunteer students from an academic project. The students were divided into two groups of profiles according to their performance in the project. The results reveal that more productive students had different opinions about satisfaction for both techniques and a positive trend according to the Questionnaire results. On the other hand, the results from another group were very close for both techniques and there was a trend toward neutrality in terms of satisfaction.

PALAVRAS-CHAVE

Pensando em voz alta, Questionários, Groupware, Análise de sentimentos, SentiStrength

1 INTRODUÇÃO

Em um ambiente de desenvolvimento de *software*, as equipes trabalham de forma colaborativa, acessando repositórios de informações compartilhadas, em prol da produção de artefatos de *software*. Para automatizar tarefas do processo e viabilizar as colaborações entre os membros de uma equipe de *software*, diversas ferramentas computacionais de suporte são utilizadas. No entanto, ainda existem muitos desafios no que compete à usabilidade dessas ferramentas, visto que, embora elas sejam usadas por pessoas com uma grande afinidade computacional, problemas com baixa usabilidade podem ter impactos negativos na motivação e no desempenho do time. Pinelle, David e Gutwin [18] fizeram um levantamento sobre o estudo da evolução de *groupware*, conseguindo apontar a necessidade de, quanto desenvolvedor de ferramentas de apoio, tornar-se necessário saber como essas ferramentas interferem no trabalho das equipes que as utilizarão. Porém os autores perceberam uma deficiência de como se avaliar esta influência.

Dessa forma, torna-se evidente a necessidade de se avaliar questões de usabilidade de tais ferramentas para que elas se adequem às necessidades de seus usuários[20], embora esta não seja uma tarefa fácil [19]. Murgia *et al.* [15] mostram a importância de se avaliar a satisfação de uso na ótica dos desenvolvedores, quando sendo usuários de sistemas colaborativos de apoio à Engenharia de *Software*. Para Bevan [4], muitos desenvolvedores consideram que a satisfação de uso é uma resposta pessoal que não pode ser

quantificada, onde, nos testes de usabilidade somente se obtém um resultado qualitativo.

Este trabalho propõe uma avaliação comparativa da satisfação de uso de ferramentas colaborativas de apoio à Engenharia de *Software*, utilizando as técnicas qualitativas Pensando em Voz Alta e Questionários.

Os resultados obtidos neste trabalho buscam apoiar a tomada de decisão quanto à adoção ou substituição de ferramentas de apoio às equipes de desenvolvimento, bem como a necessidade de aprimoramento de uso de tais ferramentas, sempre levando em conta os impactos que a satisfação de uso pode vir a ter na motivação e no desempenho das equipes e também nos produtos de *software* produzidos por elas.

Este artigo está organizado da seguinte forma: a seção 2 apresenta a fundamentação teórica acerca da avaliação de satisfação de uso de sistemas colaborativos; na seção 3, são apresentados trabalhos relacionados; a seção 4 apresenta a abordagem de teste aplicada neste trabalho e a seção 5 discute os resultados obtidos com os testes; na seção 6, são abordadas algumas ameaças à validade; e a seção 7 apresenta as conclusões e trabalhos futuros.

2 SATISFAÇÃO DE USO NO CONTEXTO DE SISTEMAS COLABORATIVOS

Segundo a norma ISO/IEC 25010 [12], qualidade de software é "a capacidade do produto de *software* em satisfazer as necessidades implícitas e explícitas quando usado em condições específicas". Dentre os atributos de qualidade apontados pela norma, destaca-se a Usabilidade, que segundo a norma ISO/IEC 9241-11 [13], é o atributo responsável por mensurar e assegurar que o sistema seja capaz de ser facilmente utilizado pelo público alvo. A usabilidade abrange a efetividade de uso, a eficiência e a satisfação do usuário ao interagir com um sistema. Este trabalho endereça à satisfação de uso, um subatributo de usabilidade [12], que tem sido bastante estudado no âmbito da Engenharia de *Software* e da comunidade da área de interação humano computador (IHC).

Dentre os diversos sistemas para os quais se faz necessário avaliar a satisfação de uso, destacam-se os sistemas colaborativos, em especial, ao se considerar a influência que este atributo pode ter sobre os resultados do trabalho em grupo. Sistemas colaborativos ou *groupware* são grupos de ferramentas de *software* usados como facilitadores da execução de trabalhos em grupo. Segundo Nicolaci e Pimentel [16], os sistemas colaborativos devem criar espaços para serem habitados, possibilitar novas formas de trabalho e de interação social. Essas ferramentas têm como objetivo estimular e melhorar a colaboração, que neste trabalho é abordada a partir do Modelo 3C, proposto originalmente por Ellis, Gibbs e Rein [6] e, posteriormente, adaptado por Fuks *et al.* [9]. Esse modelo analisa o

suporte computacional à colaboração sob 3 dimensões relacionadas entre si: comunicação, coordenação e cooperação.

A dimensão da comunicação abrange toda e qualquer troca de informações entre os indivíduos. A cooperação, por sua vez, corresponde à operação conjunta em um espaço compartilhado. E por fim, a coordenação envolve a organização dos indivíduos e a gestão das tarefas executadas em todo o fluxo do trabalho colaborativo.

As três dimensões do Modelo 3C podem ser facilmente mapeadas em um ambiente de *software* à medida que membros da equipe cooperam em prol do desenvolvimento de artefatos de *software*, realizando atividades de forma coordenada, se comunicando e acessando repositórios compartilhados, e cooperando entre si na execução de tarefas que resultam em novas versões de artefatos de *software*. Para o gerenciamento e suporte de equipes que trabalham colaborativamente, são utilizadas as chamadas ferramentas CASE¹ que podem ser classificadas como sistemas de *groupware*. [8]

Devido à sua importância para o trabalho colaborativo, no contexto do processo de *software*, deve-se levar em consideração a avaliação de usabilidade deste tipo de sistemas, em especial, quanto à satisfação de uso. Problemas de usabilidade em tais ferramentas podem influenciar negativamente os resultados do trabalho em si além de provocar experiências negativas nos usuários (membros da equipe), que embora não intencionalmente, podem replicar os problemas vivenciados nas interfaces construídas por eles no futuro.

2.1 Análise de satisfação de uso

A satisfação de uso de sistemas tem como estudo principal a experiência do usuário ao utilizar um dado sistema. Para se avaliar a satisfação de uso de um *software*, existem várias técnicas consolidadas na literatura, como: questionários, prototipagem rápida, avaliação de desempenho, observação, *hallway testing*, *rapid iterative testing and evaluation*, *think-aloud*, *Wizard of Oz*, *remote usability test* ou *co-discovery*, entre outras. Segundo Catecati *et al.* [5], essas técnicas podem ser divididas em duas classes, sendo elas: baseadas em observação e medidas realizadas sobre o usuário.

Com o intuito de explicitar o sentimento dos usuários ao interagirem com o sistema colaborativo, este trabalho se enquadra como sendo da classe de medidas realizadas sobre o usuário. Dentre as técnicas dessa classe, que conseguem apontar a satisfação de uso de forma mais concisa, destaca-se o Pensando em Voz Alta (*Think Aloud Protocol*) do qual derivam uma série de outros métodos cujos resultados são de natureza essencialmente qualitativa.

Da mesma forma, os questionários também se enquadram como medidas realizadas sobre usuário. Filardi *et al.* [7] complementam que a satisfação de uso pode ser medida subjetivamente a partir de questionários, fazendo assim um estudo da intensidade de sentimentos do usuário.

2.1.1 Pensando em voz alta (*Think Aloud Protocol*). Pensando em Voz Alta é uma técnica de teste que tem como base a psicologia cognitiva, inicialmente utilizada para testes de memória de curto prazo. Contudo, Nielsen [17] afirma que esse talvez seja o mais valioso teste de usabilidade, e que dependendo de suas variantes, podem ser extraídos diversos resultados diferentes.

O teste basicamente é feito a partir da aplicação de teste de ferramentas, onde os usuários irão verbalizar suas decisões, ações, opiniões sobre as tarefas e ferramentas e seus sentimentos em relação à ferramenta a ser avaliada.

Existem diversas variantes do Pensando em Voz Alta, sendo que para cada uma delas é possível extrair algum tipo de informação dos participantes. Catecati *et al.* [5] aponta algumas dessas variantes como Método *Coaching*, Aprendizagem por Co-descoberta, Método de Ensino, Testes para Medição de Desempenho e Protocolo de Questionamentos.

Para este trabalho, como o intuito de se extrair os sentimentos em torno da satisfação de uso de ferramentas colaborativas de apoio à Engenharia de *Software*, foi empregada a variante Protocolo de Questionamentos. Nessa variante, o usuário é incentivado a verbalizar seus pensamentos e o observador, caso necessário, estimula essa verbalização com perguntas diretas e específicas das tarefas, tentando assim, extrair o padrão mental do usuário em relação à tarefa e à ferramenta testada.

2.1.2 Questionários. Questionários são ferramentas rápidas e baratas de se obter informações de um grupo significativo de pessoas que fazem parte do experimento de teste. Segundo Filardi *et al.* [7], trata-se de uma técnica bem estabelecida de coleta de dados demográficos e de opiniões de usuários que pode ser utilizada isoladamente ou em conjunto com outros métodos de avaliação, a fim de se esclarecer ou aprofundar algum assunto.

Existem três tipos de questionários[7]:

- Fechados: Perguntas feitas de forma sucinta, contendo um conjunto prévio de respostas que o usuário pode dar como representação mais próxima do seu ponto de vista sobre a questão.
- Abertos: O usuário fica livre para formular suas respostas, de forma a explicitar ainda mais seu ponto de vista sobre a ferramenta testada.
- Mistos: Mistura dos outros dois tipos supracitados, devendo ser aplicado quando há necessidade de se obter informações mais completas do usuário.

Quanto ao formato adotado para o questionário segue alguns dos modelos populares em testes [1]:

- ASQ (*After Scenario Questionnaire*): Aborda o momento e forma como se fazer a aplicação de questionário, tendo em vista validar a confiabilidade e a sensibilidade, após a execução de testes.
- SEQ (*Single Ease Question*): Contém perguntas curtas, diretas e fáceis de serem respondidas, gerando conforto e facilidade ao se reponder o questionário.
- SUS (*System Usability Scale*): Escala que busca atribuir valores sobre a concordância ou discordância de cada questionamento, procurando assim alcançar a opinião de forma a abranger as possíveis opiniões de uma forma quantificada.

Para realização deste trabalho, serão aplicados questionários do tipo fechado, tendo como base métodos descritos por Lewis [14] que utilizam o formato ASQ e uma escala SUS de 1 à 7, voltados a investigar a satisfação de uso, em conjunto com alguns métodos descritos por Bevan [4], analisando assim, os aspectos do prazer

¹ *Computer-Aided Software Engineering*: Todo *software* que auxilia atividades de Engenharia de *Software*, da análise de requisitos e modelagem até a programação e testes.

de uso, conforto, confiança e amabilidade, conforme recomendado pela norma ISO/IEC 25010 [12].

3 TRABALHOS RELACIONADOS

Santos *et al.* [20] fazem um levantamento sobre as formas utilizadas para se avaliar aplicações colaborativas, mostrando o quanto o assunto é vasto e importante e que embora sejam sempre feitos muitos estudos sobre os métodos de avaliação, são feitas avaliações em sistema monousuários. O trabalho destaca que em geral os métodos utilizados nas avaliações são combinações de métodos já existentes e com foco em se avaliar sistemas monousuário.

Gutwin *et al.* [10] fizeram um estudo sobre os métodos de avaliação voltados a usabilidade de *groupware*, mostrando a possibilidade de se fazerem mudanças aos métodos existentes, a fim de adaptá-los à avaliação de sistemas colaborativos. No trabalho, os autores afirmam a importância da usabilidade em sistemas colaborativos, já que nestes ambientes, caso o grupo se sinta incapaz de se comunicar com eficácia e eficiência ou seja incapaz de coordenar suas ações de forma simples, o desempenho e a satisfação deste grupo será afetada de alguma maneira.

Murgia [15] faz um estudo apontando a necessidade de se entender mais a fundo as emoções dos desenvolvedores durante o processo de desenvolvimento de *software*, apontando a necessidade de uma ferramenta automatizada que faça a mineração de emoções.

É neste contexto, de se extrair os sentimentos dos usuários de sistemas colaborativos, usando ferramentas de análise de sentimentos, que este trabalho busca uma ligação entre a satisfação de uso e os sentimentos dos usuários durante o uso de tais ferramentas.

4 TESTES REALIZADOS

Esta seção apresenta uma abordagem de testes de satisfação de uso em sistemas colaborativos de apoio à Engenharia de *Software*, descrevendo as técnicas aplicadas e os resultados obtidos com os testes. Apesar de, usualmente, esse tipo de teste ser feito com base em questionários, para este trabalho foi aplicada uma abordagem combinada destes com a técnica Pensando em Voz alta, com intuito de se obter a verbalização, contendo as emoções e sentimentos vivenciados pelos usuários. Para dar suporte à técnica Pensando em Voz Alta, utilizou-se da análise de sentimentos, aplicada sobre textos gerados a partir da verbalização dos usuários durante o teste. Os textos foram então submetidos a um método de análise de sentimentos, com a intenção de facilitar a análise das emoções e sentimentos dos usuários para com as ferramentas CASE testadas.

4.1 Ferramentas sob testes

Os testes foram aplicados em um cenário acadêmico, referente a um projeto interdisciplinar que vem sendo aplicado, desde 2015, no curso de Ciência da Computação da Universidade Federal de Viçosa - *Campus* Florestal. Esse projeto envolve alunos das disciplinas de Programação Orientada a Objetos (POO), Banco de Dados (BD), Engenharia de *Software* (ES) e Arquitetura de *Software* (AS) para proporcionar-lhes, ainda que no universo acadêmico, experiências mais próximas à realidade da indústria de *software*, no que diz respeito ao desenvolvimento colaborativo de um produto de *software*.

Este trabalho utilizará os dados da última edição do projeto interdisciplinar (2017/2), com a finalidade de extrair informações de seus participantes no que compete à satisfação de uso das ferramentas colaborativas adotadas. Serão avaliadas três ferramentas CASE, visando abranger as três dimensões do modelo 3C, sendo o *Slack*² avaliado no âmbito da Comunicação, o *Trello*³, no que compete à Coordenação e o *GitLab*⁴ para a dimensão da Cooperação.

4.1.1 Slack. Voltado a dar suporte à Comunicação, o *Slack* é uma ferramenta que possui funcionalidades parecidas com a de um *chat*, podendo realizar chamadas de vídeo, altamente customizável e sendo sempre citado por sua facilidade ao ser utilizado por equipes de desenvolvimento.

4.1.2 Trello. Basicamente um quadro *Kaban*⁵ virtual, utilizado para Coordenação de tarefas. Possui funcionalidades que evidenciam as tarefas a serem feitas, os responsáveis pelas mesmas, comentários relacionados e os prazos.

4.1.3 GitLab. Similar ao *Git*, visa apoiar a Cooperação da equipe dentro da equipe de desenvolvimento, gerenciando o repositório, permitindo que os desenvolvedores hospedem os códigos em seus servidores.

4.2 Procedimento de testes

Para condução dos testes de satisfação de uso aplicados neste trabalho, os seguintes passos foram realizados: Seleção das personas (usuários representativos), Preparação das tarefas dos testes, Preparação do ambiente de teste, Execução dos testes, e Processamento dos dados dos testes.

4.2.1 Seleção das personas. Segundo Nielsen [17], um número razoável de personas para estes tipos de testes, de forma a obter dados fiéis, seria de 5 a 8 personas. Sendo assim, foram escolhidos 2 alunos de cada disciplina, respeitando a ideia de variedade de perfil segundo o rendimento de cada disciplina, totalizando assim 8 personas, que serão, a partir de agora citados como persona X, sendo X índice de 1 a 8.

Para recrutar as personas, foram contactados os professores responsáveis pelas quatro disciplinas envolvidas no projeto interdisciplinar para que eles indicassem 2 alunos por disciplina, selecionados com base em seu rendimento e classificados em dois grupos, um com aqueles alunos com melhor aproveitamento (G1) e outro com aproveitamento mediano (G2). Para facilitar a distribuição das personas e a análise dos resultados, o grupo G1 abrangeu as personas de índices ímpares e o grupo G2, aquelas com índices pares (Tabela1).

4.2.2 Preparação das tarefas dos testes. No que compete ao Pensando em Voz Alta, como cada uma das ferramentas será avaliada de forma independente, foi elaborada uma tarefa de teste, com um cenário de uso representativo, para cada uma delas. Levando-se em consideração que o intuito do teste é extrair informações que contivessem o sentimento das personas ao utilizar as ferramentas e não a facilidade de uso, foram elaboradas tarefas básicas das mesmas.

²<https://slack.com/>

³<https://trello.com/login>

⁴<https://about.gitlab.com/>

⁵Ferramenta de gestão, originária do Toyotismo que permite saber o que, quando e quanto produzir.

Grupos	Disciplinas			
	POO	BD	ES	AS
G1	Persona 1	Persona 5	Persona 7	Persona 3
G2	Persona 8	Persona 6	Persona 2	Persona 4

Tabela 1: Divisão de personas por grupos

Foram criados três fluxogramas (Figuras 1, 2 e 3), um para cada ferramenta testada, bem como um roteiro de texto corrido, para orientar as personas durante os testes. Além disso, para assegurar um limite de tempo durante a execução de teste, ficou estipulado, que caso necessário, cada participante teria 3 tentativas para executar as tarefas do teste para cada ferramenta.

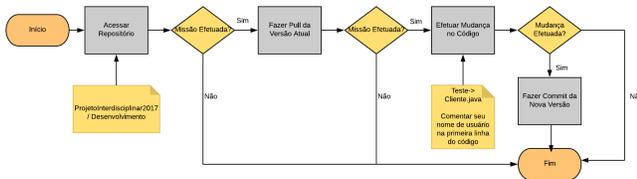


Figura 1: Cenário de teste - GitLab

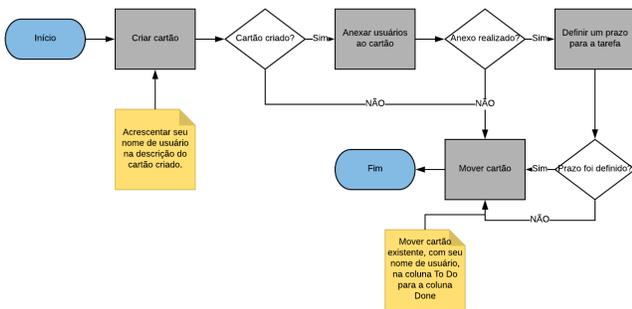


Figura 2: Cenário de teste - Trello

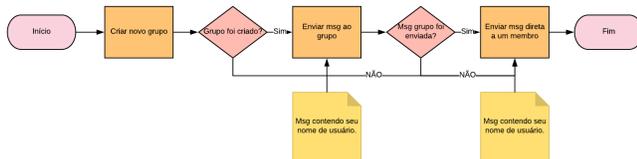


Figura 3: Cenário de teste Slack

Por fim, para viabilizar a execução dos testes segundo os cenários planejados, foram criados perfis nas três ferramentas a serem testadas e inseridas algumas informações para se criar um ambiente

"real" para o teste. Para o *GitLab*, replicou-se o repositório utilizado no projeto interdisciplinar.

Com relação ao Questionário, foram utilizadas 10 perguntas de múltipla escolha (Tabela 2), elaboradas em um questionário fechado por meio do *Google Forms*⁶. O mesmo questionário foi aplicado para cada ferramenta testada e objetivou extrair, de forma mais explícita, o sentimento das personas durante a realização dos testes com o Pensando em Voz Alta. As perguntas foram extraídas dos questionários de satisfação de uso baseado no trabalho de Lewis [14], adaptando a escala SUS utilizada, de 1 a 7 para 1 a 5, conforme sugerido por Albert *et al.* [1].

4.2.3 *Preparação do ambiente de teste.* Para aplicação do teste com Pensando em Voz Alta, foi utilizada uma sala da Universidade Federal de Viçosa - *campus Florestal*, equipada com um computador (teste individual) conectado a uma rede cabeada, para se garantir a estabilidade da conexão à internet, já que as três ferramentas são acessíveis pela *Web*. De forma a evitar atrasos na realização dos testes, cada pessoa já iniciava o teste logada em cada uma das ferramentas. Por fim, para realizar o registro das emoções/sentimentos das personas durante a realização do teste Pensando em Voz Alta, foram utilizados um celular para gravação dos áudios e uma ferramenta de gravação de áudio *Kazam Screencaster*, para ser utilizada como *backup*, instalada no computador utilizado nos testes.

Já para a aplicação do Questionário, não houve preparação do ambiente, visto que, por se tratar de um questionário disponibilizado na *Web*, cada pessoa pôde respondê-lo sem acompanhamento.

4.2.4 *Execução dos Testes.* No início dos testes das ferramentas, as personas foram orientadas a fazerem uso dos fluxogramas (Figuras 1, 2 e 3) e verbalizarem suas emoções durante o teste, conforme a técnica Pensando em Voz Alta, apresentada na seção 2.1.1. Esta verbalização foi gravada em áudio para posterior processamento. No ambiente de testes, havia apenas o aplicador de testes e uma pessoa por vez. As personas foram orientadas, previamente, sobre a necessidade da verbalizar suas emoções durante o uso das ferramentas. Porém, em alguns casos, houve a intervenção do aplicador dos testes que estimulava essas verbalizações por meio de perguntas às personas.

Conforme dito anteriormente, o Questionário foi respondido de forma livre pelas personas, sem acompanhamento, e disponibilizado às mesmas pela *Web*, por meio do *Google Forms*.

4.2.5 *Processamento dos dados dos testes.* Para processar os dados obtidos com a técnica Pensando em Voz Alta, foi utilizada análise de sentimentos, com o intuito de se avaliar as emoções dos usuários no que diz respeito à satisfação em relação à ferramenta sob teste. Foi adotada uma análise em nível de sentenças⁷, pois foram analisadas sentenças referentes às emoções dos usuários durante a utilização das ferramentas. Para automatizar a análise de sentimentos, foi utilizado o *SentiStrength*, um método bem conhecido e difundido na área que utiliza aprendizado de máquina para fazer um comparativo entre técnicas supervisionadas e não supervisionadas [2].

⁶ <https://www.google.com/forms/about/>

⁷ Frases curtas, expressando uma única opinião pertencente a um indivíduo em análise [3].

	1	2	3	4	5
Questões	Discordo fortemente	Discordo	Neutro	Concordo	Concordo fortemente
1) No geral, estou satisfeito com o quão fácil é usar a ferramenta. (Sei como utilizar a ferramenta e sua interface facilita o uso.)					
2) Fui capaz de completar todas as tarefas da ferramenta.					
3) Fui capaz de completar as tarefas da ferramenta rapidamente. (Tive facilidade e a interface é muito intuitiva para realizar as tarefas.)					
4) Eu me sinto confortável usando a ferramenta. (As informações da ferramenta são harmoniosas, o que faz com que não haja estresse ao utilizá-la.)					
5) A ferramenta possui um feedback que me ajuda a corrigir problemas. (Os feedbacks e interface auxiliam muito no uso da ferramenta.)					
6) Sempre que eu cometo um erro usando a ferramenta, eu me recupero com facilidade e rapidez. (Consegui encontrar as opções corretas na interface da ferramenta.)					
7) As informações fornecidas pela ferramenta são claras e fáceis de entender. (Possui interface clara e intuitiva.)					
8) A organização das informações nas telas da ferramenta é clara. (Todas as informações necessárias para seu uso se encontram disponíveis em sua interface.)					
9) A interface da ferramenta é agradável. (A ferramenta apresenta características agradáveis, deixando a tarefa prazerosa.)					
10) No geral, estou satisfeito com a ferramenta.					

Tabela 2: Questionário aplicado às personas

Dentre os resultados obtidos em uma análise de sentimentos, se encontra a polaridade, que tem como resultado a emoção (positiva, negativa ou neutra), a intensidade de sentimento e o próprio sentimento (felicidade, tristeza, raiva, etc). Dessa forma, é possível medir a qualidade afetiva das ferramentas sob avaliação, ou seja, dimensionar o potencial ou estímulo das mesmas em causar mudanças no estado afetivo dos usuários [11].

Sendo assim, os áudios gravados durante o teste com o Pensando em Voz Alta foram convertidos em texto com o intuito de submeter os registros à análise de sentimentos. Para isso, foi usada a ferramenta *voicemeeter*⁸ em conjunto com a ferramenta *web Voice Dictation*⁹, sendo necessário, em alguns trechos, adequar o texto manualmente, para assim, poder separar os registros por ferramenta e fazer retirada de falas do aplicador do teste, que por vezes teve que intervir com perguntas curtas e constavam nas gravações feitas. Após essa pré-formatação do texto, foram gerados 3 textos, um para cada uma das ferramentas, que foram submetidos a análise de sentimentos no nível de sentença, por meio da ferramenta *web IFeel*¹⁰. Essa ferramenta faz a análise de sentimentos de um texto de entrada e fornece, como saída, a polaridade (positivo, negativo ou neutro) e o *score* de 18 métodos diferentes. No entanto, neste trabalho, apenas as polaridades do método *SentiStrength* foram utilizadas.

Com relação aos dados obtidos com as respostas do Questionário, para facilitar a avaliação comparativa com os resultados do Pensando em Voz Alta, processados pela análise de sentimentos,

foi necessário contabilizar o número de respostas positivas, negativas e neutras, de cada um dos questionários, para cada ferramenta. A partir dessa contagem, foi atribuída uma polaridade para cada persona por ferramenta, como é possível visualizar na Tabela 3. Para as respostas na escala de 1 a 5, contidas nos questionários, foi realizado um mapeamento para a polaridade (-1, 0 e 1) da seguinte forma:

- Respostas 1 e 2: Contabilizadas como negativas = -1.
- Resposta 3: Contabilizada como neutra = 0.
- Respostas 4 e 5: Contabilizadas como positivas = 1.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS DOS TESTES

Após a execução de todos os passos descritos na seção anterior, obteve-se como resultado: um conjunto de polaridades proveniente da análise de sentimentos sobre os testes com o Pensando em Voz Alta; e um conjunto de respostas referentes à aplicação do Questionário, devidamente polarizadas para facilitar a análise comparativa.

Diante dos resultados, com suas respectivas polaridades (positivas, negativas e neutras), foram elaborados gráficos correspondentes aos resultados de cada um dos dois grupos de personas (G1 e G2), segundo os testes de cada ferramenta. Para fins de construção dos gráficos, foi criada uma correspondência numérica, onde a polaridade positiva tem valor igual a 1, a polaridade neutra terá valor igual a 0 e a polaridade negativa tem valor igual a -1. Vale ressaltar

⁸<https://www.vb-audio.com/Voicemeeter/>

⁹<https://dictation.io/>

¹⁰<http://blackbird.dcc.ufmg.br:1210/?analysingLine=true>

	GitLab				Trello				Slack			
	Positivo	Neutro	Negativo	Respostas polarizadas	Positivo	Neutro	Negativo	Respostas polarizadas	Positivo	Neutro	Negativo	Respostas polarizadas
Persona 1	8	2	0	1	10	0	0	1	7	2	1	1
Persona 2	0	6	4	0	4	5	1	0	5	4	1	1
Persona 3	2	5	3	0	8	2	0	1	3	2	5	-1
Persona 4	0	2	8	-1	1	9	0	0	1	7	2	0
Persona 5	1	0	9	-1	10	0	0	1	7	3	0	1
Persona 6	1	4	5	-1	7	3	0	1	9	1	0	1
Persona 7	1	4	5	-1	10	0	0	1	9	1	0	1
Persona 8	0	0	10	-1	0	10	0	0	10	0	0	1

Tabela 3: Polarização dos questionários

que existem gráficos contendo somente a escala positiva, diante da ausência de resultados negativos.

Inicialmente, na Figura 4, é possível visualizar os resultados obtidos para a ferramenta *GitLab*, para os grupos G1 (Figura 4(a)) e G2 (Figura 4(b)). No gráfico da Figura 4(a), é possível visualizar os resultados obtidos para as personas do grupo G1, onde apenas a persona 7 obteve resultados diferentes para as técnicas utilizadas. Além disso, houve diferentes opiniões, em termos da satisfação de uso, entre as personas desse grupo, com uma pequena tendência para a polaridade negativa, em função dos resultados do Questionário.

Já para os resultados do grupo G2, o gráfico da Figura 4(b), também para o *GitLab*, aponta todos os resultados idênticos para as duas técnicas aplicadas. Também é possível observar, que 3 das 4 personas mostraram-se neutras com relação à ferramenta. Isso parece indicar uma indiferença da persona para com a ferramenta. Uma possível explicação para esses resultados seria o fato de que as personas do grupo G2 tiveram um rendimento inferior no projeto interdisciplinar e, consequentemente, em termos de utilização das ferramentas, conforme apontado pelos professores responsáveis.

No que compete aos resultados para a ferramenta *Trello*, pode-se observar o gráfico da Figura 5. Para o grupo G1 (Figura 5(a)), apenas a persona 1 teve as duas técnicas com resultados iguais. É interessante dar atenção especial aos dados das personas 5 e 7, que apresentaram resultados contraditórios em termos de seu sentimento com relação à ferramenta. Segundo a técnica Pensando em Voz Alta, essas personas parecem ser afetadas negativamente pela ferramenta, ao passo que, pelos resultados do Questionário, tivemos uma avaliação positiva. Considerando que as personas desse grupo tiveram um bom desempenho no projeto interdisciplinar, incluindo a facilidade em usar a ferramenta, existem opiniões controversas em termos de satisfação de uso e os resultados não são conclusivos.

Em contrapartida, conforme mostra a Figura 5(b), para grupo G2, obteve-se novamente resultados idênticos para as duas técnicas, novamente com 3 das 4 personas apresentando neutralidade em termos da satisfação de uso.

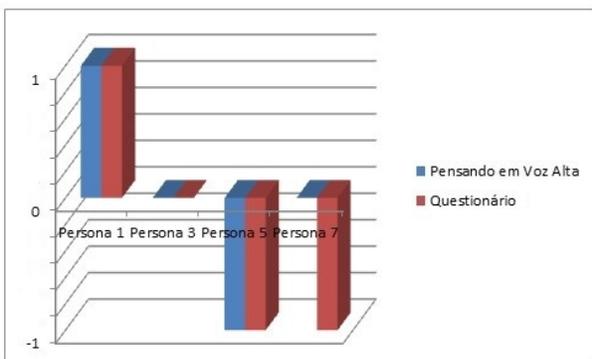
Por fim, sobre a análise da ferramenta *Slack* do grupo G1, conforme o gráfico da Figura 6(a), a persona 1 obteve resultados inversos, sendo positivo no Questionário e negativo com a técnica Pensando em Voz Alta. Uma possível interpretação para esse resultado, pode ser que embora a persona reconheça a relevância de ferramenta, de alguma forma, esta não lhe desperta emoções agradáveis. Ao mesmo tempo que para a persona 2, tem-se neutralidade,

no Pensando em Voz Alta, e polaridade negativa, no Questionário, o que poderia indicar que ela não tenha se envolvido com a ferramenta durante o uso, já que tem uma opinião negativa com relação à mesma, conforme apontado pelo Questionário. No entanto, de forma geral, para as duas técnicas, houve uma tendência para a polaridade positiva para esta ferramenta.

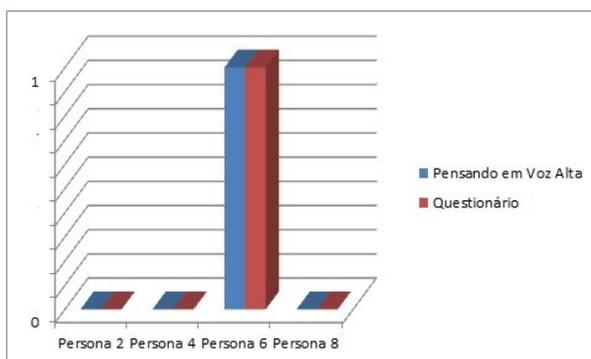
Já no grupo G2 (Gráfico 6(b)), todas as personas apresentaram neutralidade em relação ao Pensando em Voz Alta e 3 delas apresentaram polaridade positiva em relação à satisfação de uso, segundo a avaliação do Questionário. Ressalta-se que a persona 2 teve reações neutras segundo os dados obtidos com as duas técnicas. Esse resultado parece reforçar o baixo envolvimento dessas personas com as ferramentas até mesmo na fase de teste.

Diante dos resultados apresentados, observa-se uma maior discrepância entre o nível de satisfação de uso entre as personas do grupo G1. Sabendo-se que, segundo as indicações dos docentes responsáveis, as personas desse grupo apresentaram bom rendimento nas disciplinas e mostraram-se familiarizados com o uso das ferramentas testadas, pode-se apontar que ainda que um usuário tenha facilidade em usar a ferramenta, nem sempre ele se mostra satisfeito com o uso da mesma. Também é possível ponderar que talvez ele não tenha clareza de seu sentimento para com a ferramenta em termos da satisfação de uso porque o resultado do Pensando em Voz Alta mostra-se contraditório para algumas personas em relação à opinião declarada no Questionário, conforme análises para as ferramentas *Trello* e *Slack*. Para a ferramenta *GitLab*, os resultados mostraram contraditórios entre as técnicas e dentro dos resultados da mesma técnica. Dessa forma, novos testes, envolvendo um número maior de personas mostram-se necessários.

Com relação ao grupo G2, obteve-se resultados idênticos nas duas técnicas para as ferramentas *GitLab* e *Trello*. No caso da ferramenta *Slack*, obteve-se um resultado distinto para as personas 2, 6 e 8, que no Questionário, avaliaram positivamente a ferramenta, mesmo sem ter demonstrado reação positiva segundo o Pensando em Voz Alta. Interessante notar que em nenhuma das três ferramentas, para esse grupo, houve notas negativas, sendo em sua maioria, notas neutras. Isso pode indicar que as personas não chegaram a se envolver nas tarefas durante a execução dos testes ou não se sentiram seguras em avaliar negativamente as ferramentas, dado seu nível de experiência com as mesmas. Vale salientar ainda que algumas das personas do grupo G2 abortaram partes dos testes em virtude de desconhecimento das ferramentas, o que pode ter

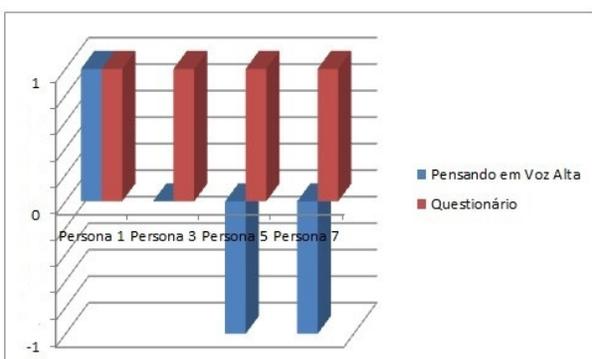


(a) Satisfação de Uso - G1

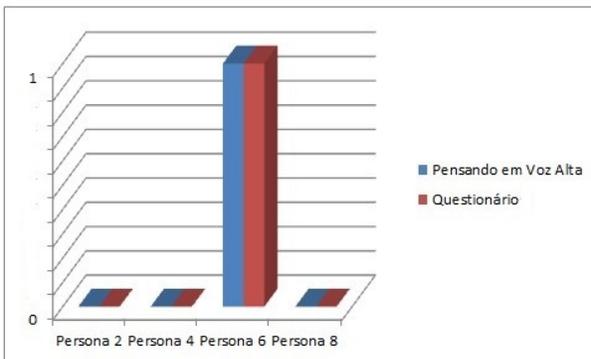


(b) Satisfação de Uso - G2

Figura 4: Testes da Ferramenta GitLab

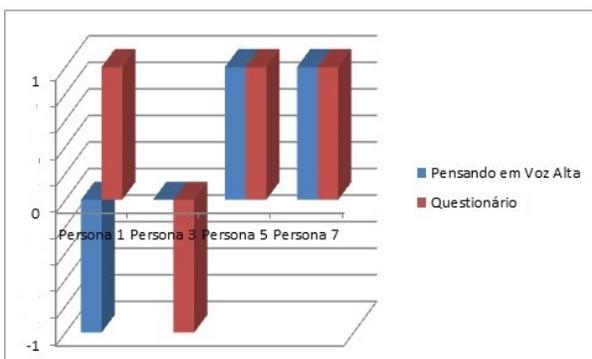


(a) Satisfação de Uso - G1

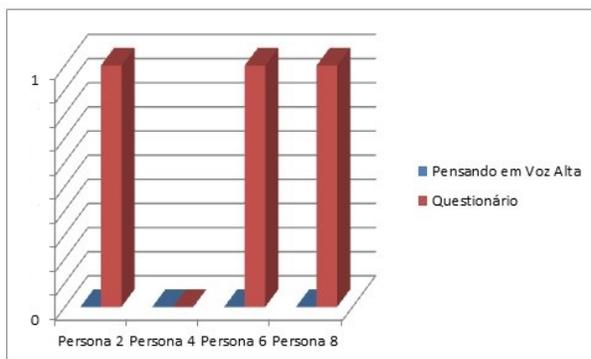


(b) Satisfação de Uso - G2

Figura 5: Testes da Ferramenta Trello



(a) Satisfação de Uso - G1



(b) Satisfação de Uso - G2

Figura 6: Testes da Ferramenta Slack

levado ao grande número de polaridades neutras nos resultados. Por fim, em virtude da neutralidade de emoções apontada pelos testes com personas do grupo G2, faz-se necessário novos testes

com um número maior de personas e até mesmo outros questionários, para que se possa avaliar o grau de influência da escolha

das ferramentas com o baixo envolvimento dos alunos durante o projeto interdisciplinar.

6 AMEAÇAS À VALIDADE

Considerando-se o alto custo de aplicação, em termos de tempo, do teste com a técnica Pensando em Voz Alta, já que se trata de um teste individual monitorado, foi utilizado um número reduzido de personas (8). Esse limitador dificulta uma análise mais conclusiva sobre os resultados dos testes, em especial, para os resultados referentes aos testes com o grupo G1 de personas, que se mostraram contraditórios.

A escolha do estudo de caso envolvendo o projeto interdisciplinar, embora tenha facilitado a aplicação dos testes, diz respeito a um ambiente acadêmico em que são simulados cenários baseados em situações reais. No entanto, o nível de experiência e envolvimento dos usuários (personas), que são alunos das disciplinas envolvidas, pode não corresponder fielmente a resultados que seriam obtidos em situações reais.

Como a satisfação de uso é um atributo muito próximo à facilidade de uso, a realização de testes nesta dimensão poderia facilitar a interpretação dos resultados, em especial, por validar a classificação das personas nos dois grupos de perfis utilizados (G1 e G2) e até mesmo auxiliar na validação dos resultados contraditórios e/ou neutros.

7 CONSIDERAÇÕES FINAIS

Após a realização deste trabalho, pode-se perceber a dificuldade de se medir a satisfação de uso, principalmente de sistemas colaborativos de apoio à Engenharia *Software*. Foi utilizada um estudo comparativo dos resultados da aplicação de duas técnicas qualitativas voltadas à avaliação de satisfação de uso em um cenário acadêmico.

Os testes envolveram oito pessoas, categorizadas em dois grupos de perfis (G1 e G2), que apresentaram resultados distintos, onde usuários mais produtivos apresentaram divergência de opiniões para as duas técnicas e uma tendência positiva nas respostas dos questionários. Já para usuários menos produtivos, os resultados entre as técnicas são aproximados e houve uma tendência à neutralidade. Contudo, houve uma tendência para a polaridade positiva na resposta dos questionários, no âmbito das ferramentas *Trello* e *Slack*. Já com relação ao *GitLab*, os resultados não foram conclusivos. Para o grupo G2, os resultados foram mais harmoniosos entre as duas técnicas, porém com uma tendência muito forte para a neutralidade em relação à satisfação de uso. Apenas para a ferramenta *Slack*, obteve-se uma polaridade indicativa de sentimento positivo, quando analisados os questionários.

Para trabalhos futuros, pretende-se aplicar o teste com um número maior de participantes e em um cenário real. Além disso, pretende-se ampliar os testes, incluindo testes de facilidade de uso, com o intuito de se avaliar as influências desta sobre a satisfação dos usuários.

REFERÊNCIAS

[1] William Albert and Thomas Tullis. 2013. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes.

- [2] Matheus Araújo, Pollyanna Gonçalves, Fabricio Benevenuto, and M Cha. 2013. Métodos para análise de sentimentos no twitter. In *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*.
- [3] Fabricio Benevenuto, Filipe Ribeiro, and Matheus Araújo. 2015. Métodos para Análise de Sentimentos em mídias sociais.
- [4] Nigel Bevan. 2010. Extending the concept of satisfaction in ISO standards. In *Proceedings of the KEER 2010 International Conference on Kansei Engineering and Emotion Research*.
- [5] Tiago Catecati, Fernanda Gomes Faust, GAL Roepke, FS Araújo, D Albertazzi, ARG Rammirez, and MGG Ferreira. 2011. Métodos para a avaliação da usabilidade no design de produtos. *DA Pesquisa* 4 (2011), 564–581.
- [6] C. A. Ellis, S. J. Gibbs, and G. L. Rein. 1991. Groupware. *Commun. ACM* 34, 1 (1991), 39–58.
- [7] Ana Lúcia Filardi and Agma Juci Machado Traina. 2008. Montando questionários para medir a satisfação do usuário: avaliação de interface de um sistema que utiliza técnicas de recuperação de imagens por conteúdo. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*. Sociedade Brasileira de Computação, 176–185.
- [8] Hugo Fuks and Mariano Pimentel. 2011. *Sistemas colaborativos*. Elsevier Brasil.
- [9] Hugo Fuks, Alberto B. Raposo, Marco A. Gerosa, Mariano Pimentel, and Carlos J.P. Lucena. 2007. The 3C Collaboration Model. In *The encyclopedia of e-Collaboration*. Information Science Reference, 637–644.
- [10] Carl Gutwin and Saul Greenberg. 2000. The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000.(WET ICE 2000)*. *Proceedings. IEEE 9th International Workshops on*. IEEE, 98–103.
- [11] Elaine Hayashi, Vânia Neris, Cecília Baranauskas, Maria Cecília Martins, Lara Piccolo, and Rosely Costa. 2009. Avaliando a qualidade afetiva de sistemas computacionais interativos no cenário brasileiro. *Usabilidade, Acessibilidade e Inteligibilidade Aplicadas em Interfaces para Alfabetos, Idosos e Pessoas com Deficiência* (2009), 55.
- [12] ISO 25010 2011. *ISO/IEC 25010 System and software quality models*. Standard. International Organization for Standardization.
- [13] ISO 9241-11 2018. *ISO/IEC 9241-11 Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. Standard. International Organization for Standardization.
- [14] James R Lewis. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7, 1 (1995), 57–78.
- [15] Alessandro Murgia, Parastou Tourani, Bram Adams, and Marco Ortu. 2014. Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In *Proceedings of the 11th working conference on mining software repositories*. ACM, 262–271.
- [16] Ana Maria Nicolaci-da Costa and Mariano Pimentel. 2011. Sistemas colaborativos para uma nova sociedade e um novo ser humano. *Sistemas colaborativos*. PIMENTEL, M.; FUKS, H.(Orgs.). Rio de Janeiro: Elsevier (2011).
- [17] Jakob Nielsen. 1994. *Usability engineering*. Elsevier.
- [18] David Pinelle and Carl Gutwin. 2000. A review of groupware evaluations. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000.(WET ICE 2000)*. *Proceedings. IEEE 9th International Workshops on*. IEEE, 86–91.
- [19] Raquel Oliveira Prates and Alberto Barbosa Raposo. 2006. Desafios para testes de usuários em sistemas colaborativos-lições de um estudo de caso. In *Proceedings of VII Brazilian symposium on Human factors in computing systems*. ACM, 9–12.
- [20] Natália Sales Santos, Raquel Oliveira Prates, et al. 2012. Caracterização das adaptações em métodos de avaliação para aplicações colaborativas. In *Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems*. Brazilian Computer Society, 297–300.