

Uma Avaliação do Uso de Gamificação para Apoiar o Ensino e Aprendizagem de Testes Exploratórios

Igor Ernesto Ferreira Costa

Programa de Pós-Graduação em Ciência da Computação
- Universidade Federal do Pará (UFPA)
Belém - Pará - Brasil
igor.ef.costa@gmail.com

Sandro Ronaldo Bezerra Oliveira

Programa de Pós-Graduação em Ciência da Computação
- Universidade Federal do Pará (UFPA)
Belém - Pará - Brasil
srbo@ufpa.br

ABSTRACT

This paper presents the exploratory testing, an approach that has become very relevant in software testing, mainly due to the emerging the use of agile practices in the software development process to satisfy the needs of the market (Time to Market). However, it was possible to notice that it is a subject little discussed in the academic context. For this reason, this work uses gamification as a systematic strategy in teaching and learning of exploratory testing using case studies with two classes, which aims to maintain strong student involvement and an incentive for good performance, obtaining as results students prepared to use this testing approach. As a final result, the students achieved a good overall performance; with reports that gamification facilitated and collaborated greatly to achieve the best performance, converging with the quantitative data obtained. This can be evidenced by the fact that two case studies (classes) reached a percentage superior to 70% of conquest, that is, there was an excellent general performance analyzed by the proportion of medals obtained for the participation.

KEYWORDS

Gamificação, Ensino-Aprendizagem, Teste Exploratório.

1 INTRODUÇÃO

Um dos grandes desafios enfrentados no ensino de Engenharia de Software é suprir a necessidade de utilizar métodos de ensino que tornam esse processo mais efetivo [1]. Nessas circunstâncias, vários estudos vêm sendo realizados sobre o teste de software, principalmente no contexto da aplicabilidade envolvendo métodos ágeis e abordagens sistemáticas de ensino.

O Teste Exploratório (TE) é uma abordagem de teste manual que enfatiza a responsabilidade e a liberdade do testador para explorar o sistema, permitindo que o testador adquira conhecimento do programa ao mesmo tempo em que executa os testes, pois os casos de testes não são pré-estabelecidos em um plano de teste [2, 3, 4, 5].

O TE é flexível e promove o rápido *feedback*, no entanto a escassez na geração de documentação ocasionou o surgimento das técnicas de gerenciamento de teste visando

principalmente a estruturação desta referida abordagem [4, 6]. Observou-se que a técnica *Session-Based Test Management* (SBTM) é a mais difundida dentre as outras técnicas, conforme foi evidenciada na realização de um Mapeamento Sistemático da Literatura (MSL) sobre o grau de importância de TE [7]. Assim, este trabalho buscou integrar o SBTM na aplicação de TE.

No contexto de educação, o uso da gamificação permite que os estudantes consigam visualizar os efeitos de suas ações, o seu desempenho no aprendizado e como isso acontece progressivamente, tornando-se um facilitador no relacionamento entre as partes envolvidas na prática do ensino, imerso como em um jogo [8]. Assim, define-se que a gamificação é a utilização de elementos de jogos fora do seu contexto, onde é utilizado para mobilizar os indivíduos a agir, ajudar, resolver problemas, interagir e promover o aprendizado [9, 10].

Portanto, este trabalho tem como objetivo aplicar uma estratégia sistemática para o ensino de TE, usando gamificação para proporcionar aos alunos um melhor conhecimento sobre a aplicabilidade dessa abordagem. Assim, este estudo tem a seguinte questão de pesquisa: O uso da gamificação auxilia no ensino e aprendizagem dos participantes no TE?

Além desta seção introdutória, O trabalho está estruturado tendo na Seção 2 a metodologia da pesquisa, na Seção 3 a estrutura gamificada usada no estudo de caso, na Seção 4 o estudo de caso realizado, na Seção 5 a discussão dos resultados obtidos e na Seção 6 as ameaças à validade deste estudo. Finalmente, na Seção 7 a conclusão.

2 METODOLOGIA DA PESQUISA

Inicialmente foi conduzido um MSL para identificar o grau de importância da aplicação de TE. Em seguida, definiu-se os elementos de jogos com base no *Octalysis Gamification Framework* [11] e então ocorreu a planejamento do estudo de caso participativo. Em paralelo definiu-se o conteúdo das aulas de treinamento com base no capítulo 4 do SWEBOK (*Software Engineering Body of Knowledge*) Versão 3.0 [5, 12, 13], elaboração das apresentações das aulas, construção dos materiais necessários e a elaboração de questões para discutir no *feedback*. Em seguida, os dados foram coletados durante a execução do estudo de caso, analisados quantitativa e qualitativamente e descritos em um relatório final.

3 O FRAMEWORK GAMIFICADO

No contexto deste estudo, Framework Gamificado constitui-se no conjunto de técnicas, materiais, regras e dinâmicas da abordagem utilizada na perspectiva de promover um ensino-aprendizagem de teste exploratório motivador, proporcionando melhores conhecimentos sobre esta abordagem de teste em questão. Além disso, destaca-se que foi utilizada a técnica de SBTM para prover uma abordagem mais sistemática de teste exploratório [14].

3.1 Visão Geral

Na gamificação foram utilizados vários elementos alusivos ao jogo de caça ao tesouro, bem como os personagens foram baseados no filme “Piratas do Caribe”. Há três perfis, os quais são:

- Especialista, é o ministrador e instrutor no estudo de caso, o qual detém amplo conhecimento da área de teste de software;
- Juiz, que observa as ações dos estudantes e preenche a planilha de *Ranking/Gamificação*;
- Testador, que são os estudantes, os quais se disponibilizaram a participar espontaneamente do estudo de caso [13].

Além disso, cada testador recebeu um nome fictício de um pirata que existiu na história. Assim, alguns elementos básicos foram envolvidos, sendo pontos de atividades, medalhas, brindes, moedas virtuais (bitaskull), cartas, catálogo de defeitos e pontos bônus [13]. A seguir as etapas da gamificação são descritas.

Na etapa **Formar o Pirata**, ocorreu o treinamento dos testadores tendo a aplicação de conceitos às atividades práticas posteriores (vide Quadro 1). Antes de iniciar tais aulas, os estudantes preencheram um formulário contendo perguntas sobre conceitos da abordagem de teste com intuito de coletar dados para identificar o grau de conhecimento inicial sobre o referido assunto. Enfatiza-se que a prática realizada envolveu a aplicação do TE usando o SBTM, conforme as demais etapas.

Quadro 1: Conteúdo das Aulas introdutórias.

Tópicos	Conteúdo	Nível de Taxonomia*
Fundamentos de Teste de Software	Terminologia relacionada a teste	C
	Relação com outras áreas	C
	Questões chave	AP
Níveis de Teste	Onde aplicar	AP
	Objetivo	AP
Técnicas de Teste	Estrutural	C
	Funcional	AP
	Teste baseado na experiência	AP
	Teste baseado em requisitos	C
	Teste baseado no risco	C
	Teste de Usabilidade e etc.	AP
Métricas Relacionadas a Teste	Avaliação do sistema sob teste	AN
	Avaliação do teste realizado	AN
Processo de Teste	Questões quanto ao gerenciamento	C
	Atividades de Teste	AP

*Taxonomia de Bloom: C: Compreender; AN: Analisar; AP: Aplicar.
Fonte: Adaptado de SWEBOOK, (2014).

Em **Receber Ordens** o especialista explicou os propósitos e as regras de cada etapa e das correspondentes atividades, bem como entregou os materiais a serem utilizados, elucidando quaisquer dúvidas de uso.

Em **Equipar Pirata** cada testador selecionou acessórios para personalizar o seu avatar. Os recursos foram obtidos a partir da escolha de três cartas de forma aleatória. As referidas cartas disponibilizam recursos de ataque, defesa e vestimenta, os quais são relacionados ao mundo de piratas.

Em **Procurar Tesouro**, ocorreu a sessão de teste, onde os testadores permanecem focados na detecção dos defeitos, priorizando-os à medida que registram no relatório de sessão de teste. Inerente a isso, havia a possibilidade de **Solicitar Ajuda**, caso os testadores obtivessem dificuldade de detectar os defeitos. As ajudas poderiam ser dicas de possíveis defeitos.

Na etapa **Discutir Estratégias** os testadores realizaram a revisão ortográfica do registro dos defeitos e a autoanálise das estratégias utilizadas conforme a técnica PROOF (*Past, Results, Outlook, Obstacles, Feelings*)[14]. É uma etapa inerente ao SBTM que foi adaptada semelhante ao *Debriefings*.

Em **Lutar na Batalha** ocorreu a revisão por pares, onde os testadores realizaram uma análise de três aspectos dos defeitos registrados no relatório de sessão pelo adversário: a) priorização, analisa se a prioridade definida foi adequada; b) clareza, analisa se há palavras ambíguas e/ou frases incompletas; c) reprodutibilidade, analisa se é possível reproduzir o defeito somente com o roteiro registrado. Para cada aspecto o participante precisa justificar no relatório de análise o motivo de estar em desacordo. Além disso, há uma atividade lúdica a fim de conter o progresso do adversário, no qual utilizam as cartas de ataque e defesa, para retirar pontos e se defender da perda de pontos, respectivamente.

Em **Validar Resultados** o especialista realizou sua própria análise de cada relatório de sessão observando os três aspectos supracitados, comparando com os relatos no relatório de análise. Também verificou se os defeitos encontrados estavam ou não no catálogo de defeitos, quando estavam presentes foi concedido uma determinada pontuação, caso contrário foi concedido uma pontuação maior, justamente pelo fato de ser um defeito nativo do programa e assim colaborar com a qualidade do sistema. Após isso, os resultados foram apresentados em sala de aula ocorrendo uma discussão em torno dos pontos positivos e negativos.

Em **Premiar Pirata Destaque** o especialista premiou os testadores que obtiveram ótimo desempenho nas atividades relacionadas às etapas **Procurar Tesouro** e **Lutar na Batalha**, sendo que os brindes somente foram concedidos se os enigmas fossem resolvidos. Os enigmas foram perguntas sobre o conteúdo ministrado na etapa **Formar o Pirata** visando a fixação dos conceitos.

Em **Comprar Recursos** ocorreu novamente uma etapa de aquisição de recursos, porém agora os testadores também tiveram a possibilidade de cartas do tipo “desconhecida”, as quais concedem moedas Bitaskull, brindes ou acessórios de personalização do avatar. Esta etapa visa a preparação novamente para uma nova detecção de defeitos e posteriormente “luta”.

Em **Premiar Vencedores**, ao final, são premiados os vencedores que conseguiram ótimo desempenho alcançando o nível 4 (avatar Jack) no estudo de caso geral. Na sala de aula também havia um tesouro escondido (bombom de chocolate) em que esses testadores precisaram encontrar seguindo a resolução de três enigmas que os levou até o local do referido

tesouro. Essa atividade ocorreu para encerrar a dinâmica do estudo de forma ainda mais lúdica.

Diante disso, no trabalho [15] é possível observar como cada elemento de jogo foi definido, materiais utilizados em cada etapa, regras de pontuação, regras de geração de avatar e mais detalhes sobre o *framework* gamificado no trabalho.

3.2 Regras e Prêmios

Basicamente, as regras foram definidas em torno da realização de atividades específicas em cada etapa do fluxo do estudo de caso e do comportamento dos estudantes. Em relação ao primeiro fator, os estudantes poderiam receber de 0 a 10 pontos por atividade, sendo que na etapa **Procurar Tesouro** os pontos concedidos estão condicionados à quantidade de defeitos detectados, podendo ser negativo, caso o defeito registrado seja um falso positivo, isto é, penaliza-se em -1000 pontos caso o estudante registre um fluxo normal do sistema como se fosse um fluxo anormal (*bug*). Em relação ao segundo fator de avaliação, foram observadas as perguntas realizadas, a presença em sala de aula, as sugestões e a participação como aspectos que concediam pontos de bônus, em contrapartida a ausência nas aulas, se atrapalham a realização das atividades e se não realizam as atividades em questão acarretam na perda de pontos bônus. Além disto, há também regras relacionadas ao controle, como cumprir o tempo destinado a uma tarefa e a resolução de enigmas que são pré-requisitos para receberem alguns brindes.

Para cada etapa, o estudante pode alcançar do nível 1 ao nível 4 tanto para o fator de desempenho na realização das atividades, quanto ao seu desempenho em relação aos itens avaliativos sobre o comportamento (ações participativas) em sala de aula. Os quatro níveis (avatars) possíveis, são: Nível 1 - avatar Marty; Nível 2 - avatar Will Turner; Nível 3 - avatar Joshamee Gibbs; e Nível 4 - avatar Jack Sparrow.

A partir do desempenho nos dois fatores avaliativos gera-se um avatar final de cada etapa, posteriormente é gerado o avatar geral, o qual está correlacionado ao desempenho geral do estudante no estudo de caso, a partir do cálculo da média aritmética considerando todas as etapas executadas. Nesse caso, considera-se a nota de peso dos avatares proporcionais aos níveis, sendo peso 1 para o nível 1, peso 2 para nível 2 e assim, por conseguinte. Dessa maneira, o avatar geral com base no resultado do cálculo da média aritmética (MA) pode ser: a) Nível 1, se $MA \leq 1$; b) Nível 2, se $1 < MA \leq 2$; c) Nível 3, se $2 < MA \leq 3$; d) Nível 4, se $3 < MA \leq 4$.

De maneira geral, os prêmios são concedidos em medalhas e moedas, que dependem do desempenho nas ações participativas, conforme descrito detalhadamente em [15].

4 O ESTUDO DE CASO

O estudo foi aplicado duas vezes com o objetivo de avaliar se o arcabouço gamificado poderia auxiliar no ensino e aprendizado sobre TE, observando sua viabilidade no nível técnico e de graduação. Além disso, com base no *feedback* dos participantes, a estrutura foi refinada (minimizando as lacunas da primeira execução) para melhorar sua aplicabilidade.

4.1 Ambiente de Aplicação

Na primeira execução o estudo de caso utilizou-se de um laboratório de informática, onde os estudantes envolvidos no estudo de caso foram três pessoas que faziam parte de uma equipe de teste, os quais utilizaram essa abordagem de teste exploratório em um projeto de desenvolvimento de um sistema para o Instituto de Ciências Exatas e Naturais (ICEN), da Universidade Federal do Pará.

Na segunda execução o estudo de caso também ocorreu em um laboratório de informática, com um grupo de seis alunos que haviam se formado recentemente no curso de Técnico em Informática do Instituto Federal do Pará, no caso eram alunos com formação técnica.

Os dois grupos foram selecionados para observar a viabilidade no contexto acadêmico e técnico de ensino (Instituto Federal de Educação, Ciência e Tecnologia - IFET). Nesse contexto, houve a perspectiva de alcançar resultados semelhantes, pois ocorre o treinamento com o objetivo de equilibrar o conhecimento dos participantes.

4.2 Execução do Estudo de Caso

No Quadro 2 é possível observar as etapas ocorridas em cada dia de execução do estudo.

Quadro 2: Etapas executadas por dia.

Dia	Etapa
1º, 2º, 3º, 4º	Formar o Pirata; (*4º dia houve a prova tradicional para avaliar o aprendizado do 1º ao 3º dia)
5º	Receber Ordens; Equipar Avatar; Procurar Tesouro; Solicitar Ajuda; Discutir Estratégia; Lutar na Batalha;
6º	Validar Resultados; Premiar Pirata Destaque; Comprar Recursos; Procurar Tesouro; Solicitar Ajuda; Discutir Estratégia; Lutar na Batalha;
7º	Validar Resultados; Premiar Pirata Destaque; Comprar Recursos; Procurar Tesouro; Solicitar Ajuda; Discutir Estratégia; Lutar na Batalha;
8º	Validar Resultados; Premiar Pirata Destaque; Premiar Vencedores; Feedback;

Fonte: Elaborado pelo autor (2019).

Basicamente, o estudo de caso é iniciado na etapa **Formar o Pirata** e finaliza na etapa **Premiar Vencedores**, contudo houve algumas iterações onde se iniciava na etapa **Procurar Tesouro** até a etapa **Premiar Pirata Destaque**.

Assim, o estudo decorreu do uso de procedimentos específicos que envolveram a aplicação de variáveis relacionadas à fixação dos conceitos, aplicação correta e aprendizado da abordagem de teste exploratório:

- para a **fixação dos conceitos** foram aplicados exercícios práticos durante a etapa **Treinar o Pirata**, além disso, houve a tarefa de resolução de enigmas na etapa **Premiar Pirata Destaque**, onde tais enigmas representavam questionamentos sobre os conceitos ministrados no treinamento e os benefícios (brindes) eram concedidos somente após a resposta correta do enigma em questão;
- para a **aplicação correta** houve o momento de validação dos resultados, onde o especialista analisava os relatórios de sessão de teste e de análise crítica de cada participante para verificar a correteza dos procedimentos realizados, além de expor os resultados desta validação em um momento de discussão em sala de aula para que os participantes pudessem observar quais os procedimentos poderiam ser melhorados, quais foram bem executados e aperfeiçoar as suas estratégias, seja de exploração do

sistema ou de otimização do processo de registro dos defeitos;

- para o **aprendizado da abordagem** foram usados questionário pré-estudo de caso, prova tradicional ao final do treinamento e questionário pós-estudo de caso, a fim de observar o nível de aprendizado antes do estudo de caso, após o treinamento e após o fim do estudo de caso por completo, bem como observou-se o *feedback* dos participantes para realizar ajustes e/ou melhorias na abordagem.

De maneira geral, foi definido que seriam observados o desempenho participativo e a realização das atividades.

Além disso, aplicou-se a estratégia de semear propositalmente alguns defeitos relacionados a interface gráfica do usuário, aos requisitos funcionais e a usabilidade sendo descritos em um documento (catálogo de defeitos). Tal estratégia visou obter dados para analisar qual a quantidade de defeitos encontrados entre semeados e nativos do sistema (presente desde o desenvolvimento). Destaca-se que os estudantes desconheciam quais defeitos haviam sido semeados ou não, apenas os observavam após a validação dos resultados.

A validade dos defeitos e classificação em nativos ou não, ocorreu em um momento extraclasse (etapa **Validar Resultados**) em que especialista analisou se era de fato um defeito ou não. Caso se confirmasse como um defeito, observava-se se estava presente no catálogo de defeitos ou era nativo do sistema, caso fosse nativo era concedido uma pontuação maior ao testador. Antes de toda essa validação, os participantes realizavam também uma análise crítica do relatório de sessão de teste de um “adversário” (etapa **Lutar na Batalha**) descrevendo em um relatório de análise o resultado dessa análise. O especialista validou também este relatório para confirmar se estavam sendo coerentes nas análises e fornecer dados para que os estudantes pudessem melhorar suas estratégias de exploração e otimização de tempo no registro de defeitos.

Por fim, para observar se houve aprendizado de TE com a gamificação proposta, foram utilizadas várias estratégias de coleta de dados: (i) nas aulas introdutórias, referente à etapa **Treinar o Pirata**, foram coletados dados antes de iniciar as aulas, a partir de um formulário contendo questões para identificar o grau de conhecimento inicial sobre conceitos de teste de software e da abordagem de teste exploratório, e no fim, a partir da aplicação de uma prova tradicional visando comparar o desempenho de cada estudante; (ii) quanto ao desempenho nas atividades, foi dada uma nota entre 0 a 10 pela realização de cada atividade e, a partir disto, gerou-se uma média de desempenho relacionando aos avatares; e (iii) no *feedback* e questionário pós-estudo de caso foram coletados dados sobre a opinião dos estudantes se conseguiram assimilar a abordagem de TE.

5 RESULTADOS E DISCUSSÕES

Quanto às características dos participantes, a média de idade foi de 21 anos. O sexo (masculino - M ou feminino - F) e a idade dos respectivos alunos são: no primeiro estudo participaram Cofresí (F, 23), Henry (M, 22) e Anne (M, 21), que estavam no último semestre do curso de graduação em

Ciência da Computação; no segundo estudo participaram Kidd (F, 19), Bart (M, 20), Edward (M, 19), Mohamed (F, 27), Felix (M, 19) e Kotaro (F, 23), que haviam concluído recentemente o curso Técnico de Informática. Todos participaram livremente e voluntária, sem oferecer nenhum benefício para participar. O único requisito era que ambos os grupos estivessem cursando pelo menos o penúltimo semestre do seu curso, a fim de garantir que obtivessem o conhecimento necessário.

5.1 Resultados Quantitativos

Para compreensão do desempenho dos estudantes nas atividades de treinamento (etapa **Formar o Pirata**) foi elaborada uma relação entre nota e conceito. Nesse caso, uma nota menor que 5 é conceito Insuficiente, de 5 a 6.9 é conceito Regular, de 7 a 8.9 é conceito Bom e de 9 a 10 é conceito Excelente. Houve também a aplicação do formulário inicial onde foi identificado que nenhum participante nas duas execuções do estudo de caso havia conhecimento prévio sobre teste de software e muito menos sobre a abordagem de teste exploratório [16].

Na primeira execução os três estudantes obtiveram um bom desempenho nos exercícios, alcançando a nota máxima. Na prova tradicional, Anne, Cofresí e Henry obtiveram notas 8, 6 e 5, respectivamente, no entanto a obtenção de conceito Regular por Cofresí e Henry pode ser justificada em razão de estarem ausentes em algumas aulas. Os três estudantes supracitados não obtiveram um conceito melhor em virtude de confundirem o que seriam as etapas do processo de teste com as etapas do processo V&V (Verificação e Validação) [16].

Na segunda execução todos os seis estudantes obtiveram também um bom desempenho nos exercícios, haja vista que alcançaram nota entre 8 e 10 nestas tarefas. Na prova tradicional, os estudantes Kidd e Edward foram mais interativos em sala de aula, principalmente com perguntas, com isso obtiveram conceito Bom. Os estudantes Felix e Mohamed obtiveram um aproveitamento Regular e os estudantes Bart e Kotaro obtiveram um aproveitamento abaixo do esperado, com conceito Ruim. Esses alunos que obtiveram um desempenho abaixo foram menos participativos. No caso, Kotaro foi o estudante que menos interagiu em sala de aula com questionamentos para esclarecer suas dúvidas em virtude de sua timidez, Mohamed não esteve presente no último dia da aula introdutória e o Bart não esteve presente nos dois primeiros dias das aulas introdutórias.

Em ambas as execuções, indagou-se no *feedback* sobre o motivo de tal desempenho baixo, então os estudantes relataram que não estudaram para a prova, isto é, todos responderam a prova somente com o aprendizado em sala de aula. Destaca-se também que após a realização do exercício teórico e da prova tradicional houve o momento de discussão das respostas para esclarecer as dúvidas e favorecer que os estudantes conseguissem aplicar tais conceitos nas atividades práticas.

Anne e Cofresí obtiveram um bom desempenho por terem alcançado em sua maioria o avatar Gibbs. O fato de Henry faltar dois dias do estudo de caso justifica a grande parte da obtenção do avatar Gibbs, tendo somente um bom desempenho justamente quando esteve presente e

participativo na dinâmica. Apesar de Henry realizar uma boa análise do relatório de sessão, o seu desempenho baixo justifica-se pelo fato de encontrar poucos defeitos e não registrar com clareza alguns desses defeitos. Nesse contexto, a média aritmética de Cofresí, Henry e Anne foram 3.07, 1.71 e 3.14, respectivamente. Com isso, somente Cofresí e Anne foram os que puderam procurar o tesouro em sala de aula por terem alcançado no geral o nível 4 (Jack Sparrow). Importante mencionar que existe mais uma participante que se inscreveu no estudo de caso, o Drake, porém ele faltou a todas as aulas da primeira execução, levando assim o avatar de nível 1 em todas as etapas do fluxo.

Dentre os 31 defeitos, 4 foram considerados duplicados, em virtude de dois estudantes registrá-los, além de terem 2 defeitos considerados como falsos positivos. Resume-se que foram detectados 27 defeitos diferentes, sendo 5 contidos no catálogo, tendo assim 22 defeitos pertinentes ao programa desde o seu desenvolvimento, isto é, em torno de 81,5% dos defeitos eram nativos do programa. Para este caso, foram executadas 5 sessões de testes com 30 minutos cada, assim sendo 27 defeitos detectados em 150 minutos. Então, a média de tempo para cada detecção e registro de um defeito foi de 5,5 minutos, evidentemente considerando três testadores e defeitos não duplicados.

Na segunda execução, a maioria dos estudantes conseguiu obter um bom desempenho ao analisar a quantidade de avatares de nível Jack alcançada. Esses estudantes obtiveram um ótimo desempenho nas etapas cruciais da dinâmica, as quais tinham como atividade a detecção de defeitos e de análise do relatório de sessão do adversário. Nesse contexto, Kidd, Bart e Eduard alcançaram o avatar Jack por terem as respectivas médias aritméticas de 3.05, 3.15, 3.05, e, consequentemente, puderam encontrar o tesouro em sala de aula. Enquanto Mohamed, Felix e Kotaro obtiveram avatar Gibbs, mas muito próximo de alcançarem nível 4 com as respectivas médias aritméticas de 2.95, 3.0 e 2.65. Da mesma forma, importante mencionar que existe mais três participantes que se inscreveram no estudo de caso, o Bonny, o Read e o Rackham, porém eles faltaram a todas as aulas da segunda execução, levando assim o avatar de nível 1 em todas as etapas do fluxo.

Dentre os 66 registros de defeitos, 6 foram considerados falsos positivos e 14 foram duplicados (dois ou mais estudantes registraram o mesmo defeito). Resume-se que foram detectados 46 defeitos diferentes e sendo 6 contidos no catálogo, tendo assim 40 defeitos pertinentes ao programa desde o seu desenvolvimento, isto é, em torno de 80,5% dos defeitos eram nativos do programa. Nesse caso, foram executadas 18 sessões de testes com 30 minutos cada, assim sendo 46 defeitos detectados em 540 minutos. Então, a média de tempo para cada detecção e registro de um defeito foi de 11,7 minutos, evidentemente considerando seis testadores e defeitos não duplicados.

Diante dos dados, percebe-se que em ambas as execuções houve uma efetividade bastante semelhante na detecção de defeitos nativos, apesar da diferença de nível de escolaridade. Além disso, destaca-se que todos os estudantes em ambas as execuções obtiveram desempenho geral bem semelhante, exceto Henry por não estar presente em três dias do estudo.

5.2 Resultados Qualitativos

Conforme já tem sido supracitado, os dados coletados no momento de *feedback* foram classificados com base na Matriz de SWOT (*Strengths, Weaknesses, Opportunities and Threats*) [17]. No *feedback* os participantes expressaram-se realizando uma análise crítica informando os pontos positivos, negativos, oportunidades e ameaças, inclusive as sugestões de melhorias, as quais foram registradas por meio de gravador de áudio. De maneira geral, esses dados analisados foram sobre os materiais, a didática, os perfis, a dinâmica (gamificação), as regras e cada etapa existente.

Nas duas execuções do estudo de caso os alunos consideraram fundamental a etapa **Procurar Tesouro**, principalmente quando houve uma discussão dos resultados dos exercícios e da prova tradicional. Durante essa discussão, eles foram capazes de perceber a correção de suas respostas, fixar os sujeitos melhorando o aprendizado sobre como criar missões e como usar estratégias de exploração, consequentemente, melhorando a aplicabilidade do TE [16].

De acordo com o relato dos estudantes, o momento de *feedback* foi importante para transmitir as suas opiniões, sugestões e a visão crítica do estudo de caso. Dessa maneira, os estudantes relataram como fatores fortes:

- *Exemplos demonstrados*, bem como a forma de apresentá-los, contribuíram para melhorar o aprendizado da abordagem de teste, e que os exemplos de casos reais foram de grande valia para compreenderem sobre tal abordagem;
- *Tópicos e Conteúdos das aulas introdutórias* contribuíram para expandir o conhecimento sobre teste;
- *O cenário de caça ao tesouro*, utilizando alguns personagens do filme Piratas do Caribe, essa relação foi bem adequada, pois os estudantes sentiram-se imersos no mundo de pirata tendo que explorar a procura de tesouros (defeitos), além disso, ele gostaram bastante dos personagens relacionados ao nível que atingiam e também da estratégia lúdica de modo geral;
- *A dinâmica* contribuiu para uma melhor compreensão da aplicabilidade do teste exploratório, sendo o momento que os estudantes mais gostaram foi na etapa **Lutar na Batalha**, enfatizando que queriam ter mais prática por ter sido bem divertido e empolgante à medida que encontravam mais defeitos e interagiam uns com os outros;
- *Cartas e Avatares*, os estudantes gostaram bastante das cartas e até mesmo sugeriram o uso de um boneco de avatar real para colocarem seus acessórios à medida que obtinham acessórios.

Em relação a estes pontos, os quatro primeiros foram citados nas duas execuções, e o item “v” foi mencionado apenas na segunda execução.

Os estudantes mencionaram como fatores fracos:

- *Didática*, na primeira execução o especialista prendeu-se em vários exemplos sobre um mesmo assunto, então isso foi relatado como um fator negativo em alguns casos justamente quando houve uma compreensão rápida no primeiro exemplo, já na segunda execução o especialista explicou em alguns momentos com rapidez, então isso foi

observado também como um fator negativo para compreensão de alguns exemplos;

- *Prática*, na primeira execução foi enfatizada que os participantes esperavam mais prática, pois se sentiriam mais seguros e eficientes na exploração;
- *Personalização do avatar*, percebeu-se um gasto de tempo maior do que o esperado nas etapas **Equipar Pirata** e **Comprar Recursos**, pelo fato dos estudantes necessitarem personalizar o avatar, isto é, na primeira execução do estudo de caso um estudante por vez dirigiu-se aos juízes para selecionar as cartas em virtude de haver uma quantidade mínima de cada tipo de carta, dessa forma ocasionou a perda de tempo na personalização do avatar.

Em relação a estes pontos, os dois primeiros itens foram citados nas duas execuções, e o item “iii” foi mencionado somente na primeira execução.

Os estudantes mencionaram como oportunidades:

- *Cartas de Ataque e Defesa*, diversificação nos recursos destinados à ataque e defesa;
- *Ferramenta para gerir os recursos*, seria importante obter uma ferramenta que permitisse gerar de maneira aleatória as cartas, principalmente, para gerir todos os recursos obtidos e os possíveis;
- *Aplicação de Teste*, obtenção de mais sessões de teste para aprimorar a descoberta de defeito;
- *Avatar*, fosse a partir de um boneco para personalizá-lo e também ficar visível a todos os recursos obtidos;
- *Solicitar Cartas*, selecionar mais cartas, por ser bastante divertido o momento de “luta”;
- *Exemplos*, exemplificação realizando analogias com algo do cotidiano para facilitar a fixação dos exemplos.

Em relação às oportunidades, os três primeiros itens foram citados nas duas execuções, e os três últimos itens foram mencionados apenas na segunda execução.

Os estudantes mencionaram como ameaças:

- *Cartas*, possui muitas cartas que podem comprometer o acompanhamento e andamento das etapas **Equipar Pirata** e **Comprar Recursos**, pelo fato de ser necessário reorganizá-las toda vez que uma pessoa retire as cartas;
- *Participantes*, possui muitas pessoas (perfil testador) que podem comprometer o andamento das etapas **Lutar na Batalha** e **Validar Resultados**, por haver bastante interação entre os participantes e uma quantidade relativamente grande de relatórios de sessão e análise para validar, respectivamente;
- *Exploração Pareada*, os estudantes relataram que se a etapa de **Procurar Tesouro** fosse realizada em duplas não seria tão vantajosa, pois muito provavelmente eles não conseguiriam manter o foco na exploração;
- *Especialista*, o instrutor deve ser de fato especialista na área de teste, caso contrário pode prejudicar o andamento e o processo de ensino-aprendizagem do estudo de caso.

Em relação às ameaças, os dois primeiros itens foram citados nas duas execuções, o item “iii” foi citado somente na primeira execução, já o item “iv” apenas na segunda execução.

Em relação às melhorias, ao fim da primeira execução do estudo de caso houve:

- a adição do 8º dia voltado à atividade prática, como visto na Seção 4.2;
- a inclusão de mais cartas, uso do álbum de cartas ao invés da planilha de recursos para minimizar os problemas com gasto excessivo de tempo; e
- os ajustes no conteúdo teórico para a concessão de mais tempo destinado à realização dos exercícios.

5.3 Discussão dos Resultados

A análise sobre todos os resultados foi realizada observando dois fatores: a) medalhas obtidas em cada execução do estudo de caso, comparando em termos de porcentagem a quantidade de medalhas esperadas em relação à quantidade obtida; b) avatar de atividades, ação participativa e avatar final, também comparando com a quantidade de avatares possíveis de serem obtidos em relação à quantidade obtida em cada nível. A quantidade de medalhas recebidas dependeu inteiramente do desempenho nas ações participativas e também se os participantes conseguissem atingir nota máxima nas atividades específicas a cada etapa, por esse motivo a análise foi realizada no contexto de obtenção das medalhas.

Na Figura 1 evidencia-se que em ambas as execuções do estudo de caso houve um aproveitamento ótimo, pois alcançou mais de 70% das medalhas possíveis, assim sendo considerado satisfatório. Detalhadamente, na primeira execução do estudo de caso era possível obter um total de 294 medalhas (100%), porém foram obtidas 210 medalhas (71,4%), sendo que 34 (aproximadamente 16,2%) dessas medalhas foram obtidas por alcance da nota máxima em algumas atividades. Na segunda execução do estudo de caso era possível obter 840 medalhas (100%), porém foram obtidas 644 medalhas (76,6%), sendo que 110 (aproximadamente 17%) dessas medalhas foram por alcance da nota máxima em algumas atividades. Tudo isso demonstra que os estudantes foram bastante participativos e interagiram tentando alcançar o máximo das pontuações.

Para gerar os gráficos expressos nas Figuras 2, 3 e 4 foram realizados os cálculos descritos a seguir. Na primeira execução do estudo de caso, têm-se três participantes, sendo executadas 14 etapas em virtude das duas iterações, assim são 48 avatares a serem considerados como total (100%). Já na segunda execução do estudo de caso, foram seis participantes, sendo executadas 20 etapas em virtude das três iterações, assim são 120 avatares a serem considerados como total (100%). Em ambas as execuções foram desconsideradas as etapas **Solicitar Ajuda** e **Premiar Vencedores**.

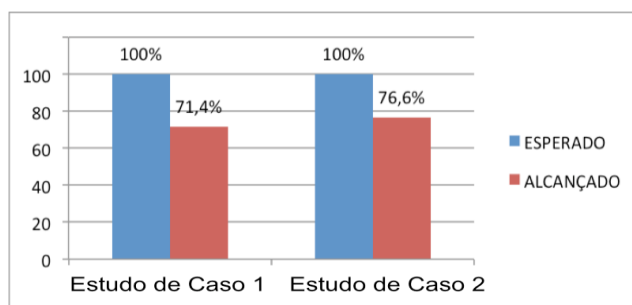


Figura 1: Desempenho Geral relacionado às Medalhas Obtidas

Na Figura 2 observa-se uma significativa quantidade de avatares de nível 1 ocasionada pela ausência de determinados estudantes em alguns dias. Por outro lado, a expressiva quantidade obtida de avatares de nível 4 não sofreu tanta influência pelo fato de ser concedido automaticamente o referido nível nos momentos em que não havia como serem analisadas as atividades, pois isso ocorreu em apenas duas etapas, no caso **Receber Ordens** e **Comprar Recursos**. Diante disso, é possível afirmar que o desempenho geral na realização das atividades foi bastante satisfatório, haja vista que a maioria dos avatares obtidos foram de nível 4.

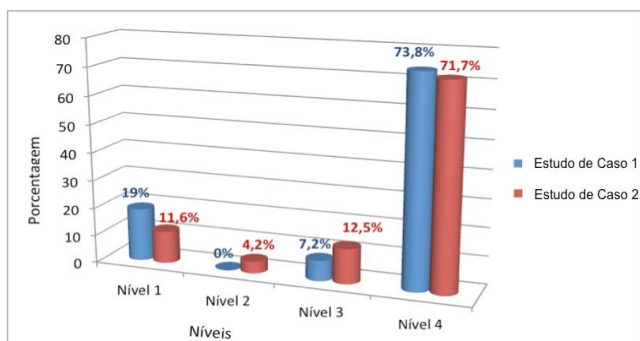


Figura 2: Desempenho Geral nas Atividades

Na Figura 3 o percentual de nível 4 demonstra que na segunda execução os estudantes interagiram um pouco mais, algo que foi congruente ao observado nas aulas. Isso pode ser justificado pelo fato dos estudantes da primeira execução consultarem bastante o website para fins de esclarecimento de dúvidas ou para se lembrarem das regras. Na segunda execução os estudantes praticamente não consultaram o website, em virtude da conexão de Internet ser instável e assim impulsioná-los a realizarem mais questionamentos.

Observa-se na Figura 4 que o desempenho geral nas etapas foi melhor dos alunos na primeira execução do estudo de caso por obterem 31% de avatares de nível 4. Nesse contexto, o fato de Cofresí e Anne conseguirem encontrar mais defeitos dentre todos os estudantes pode ter corroborado para isso. Ao analisar o desempenho obtido somando as porcentagens de nível 3 e 4, os estudantes da segunda execução foram melhores, contudo, um desempenho bastante semelhante e considerado satisfatório ao esperado. Evidentemente que esse desempenho geral poderia ainda ser bem melhor, caso todos estivessem sempre presentes e se mantivessem participativos.

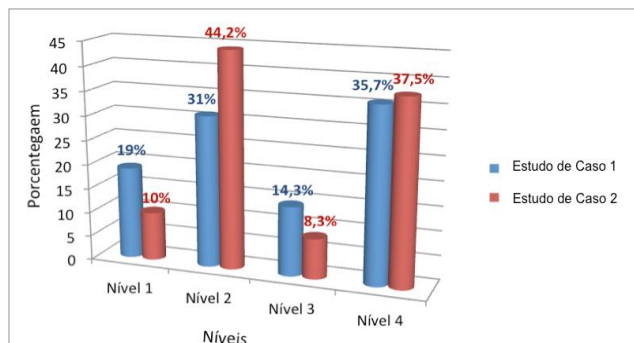


Figura 3: Desempenho Geral nas Ações Participativas

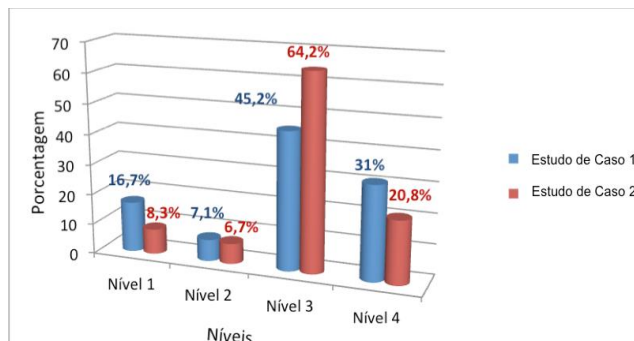


Figura 4: Desempenho Geral nas Etapas

Diante dos dados apresentados, observa-se que a gamificação contribuiu no provimento de um melhor ensino-aprendizado, haja vista a obtenção de um ótimo desempenho geral nas duas execuções do estudo de caso. Neste contexto, destacam-se alguns relatos que corroboram com todos esses dados apresentados:

- “A gamificação ajudou bastante no aprendizado da abordagem de TE, pois não ficou uma forma tradicional. Obtivemos um objetivo e assim conseguíamos melhorar a cada aula a sessão de teste sem ter medo de encontrar defeitos no sistema, que provocou uma disputa entre nós”;
- “A gamificação estimulou mais a explorar o sistema para encontrar defeitos e sempre agindo de maneira correta e organizada, conforme a abordagem de TE com SBTM”;
- “A gamificação fez que reforçasse o aprendizado sobre a abordagem de TE”;
- “Quando eu vi que não ganhei brinde por não ser destaque, logo quis encontrar mais defeitos para me destacar”;
- “A todo momento eu queria estar explorando no meu celular e vê se encontrava um defeito”;
- “Fiquei explorando tudo até que encontrei um defeito no excel, basta colocar vários números em uma única célula que o excel não permitia salvar o arquivo”.

Os estudantes afirmaram que todos os materiais estavam simples e bem compreensíveis. Os estudantes enfatizaram também que a apresentação e a discussão dos resultados da análise dos relatórios foram essenciais para adquirir uma melhor visão crítica, bem como propiciou despertar um pensamento mais criativo na exploração do programa sob teste.

Os estudantes relataram que a aplicabilidade do TE durante a dinâmica proporcionou a observação de vários fluxos e detalhes de interface e usabilidade do programa sob teste. Afirmaram também que com a liberdade obtida para explorar sentiram-se mais confortáveis para conseguirem detectar defeitos dos mais diversos níveis de criticidade. Além disso, o uso de SBTM proporcionou de fato uma melhor estruturação e organização da aplicabilidade do TE, sendo evidenciado, principalmente, na análise do relatório de sessão, pois ficaria bastante complicado de analisar e reproduzir os defeitos detectados pelo adversário sem o auxílio de um documento com o registro de tais defeitos.

Em especial na segunda execução do estudo de caso, os estudantes relataram que se sentiram bastante motivados e intuitivamente exploravam os seus dispositivos móveis e programas de computador quando estavam utilizando. Fato que ocasionou na detecção de dois defeitos no programa Microsoft Excel. Esse fato ocorreu por livre e espontânea ação desses estudantes que relataram estar entusiasmados para testar qualquer software que utilizassem.

Ainda com base nos relatos desses estudantes, eles se sentiram impressionados em como realmente a abordagem de TE permite com que várias pessoas possam detectar defeitos totalmente diferentes em uma mesma funcionalidade. Além de tudo isso, destaca-se também que os estudantes responderam um formulário para identificar o grau de satisfação. De maneira geral, todos os estudantes consideraram excelente o estudo de caso, uma vez que sete estudantes consideraram excelentes e dois relataram terem sido bom a abordagem de TE, a abordagem gamificada e a didática.

Nesse mesmo formulário sobre a satisfação foi identificada que a maior dificuldade dos estudantes na primeira execução foi o entendimento sobre alguns conceitos, mas essa dificuldade foi suprida com a prática. Na segunda execução os estudantes sentiram-se com medo de errar no registro dos defeitos para ocasionar a penalização quando fosse considerado um defeito falso positivo, bem como houve insegurança em encontrar defeitos, pois desconheciam o programa a ser testado, mesmo estando ciente da opção de ajuda. Além disso, tais estudantes não tinham bastante habilidade com o editor de texto. Contudo, a detecção de defeitos foi tornando-se mais tranquila à medida que ocorria a exploração, evidenciado por ninguém solicitar ajuda.

6 AMEAÇAS À VALIDADE

Quanto à ameaça interna, argumenta-se que o *Framework* Gamificado foi aplicado adequadamente, pois os estudantes obtiveram um desempenho proporcional a sua participação no estudo de caso. Além disso, ficou evidente que a gamificação proporcionou aos estudantes a imersão no mundo do jogo e engajamento para que obtivessem o melhor desempenho possível, tudo isso pôde ser confirmado pelos relatos no momento de *feedback*.

Quanto à ameaça externa, apesar de serem nove estudantes (testadores) o número de participantes do estudo de caso, considera-se que é possível ser reproduzido em outros ambientes ao analisar alguns fatores:

- os testadores não obtinham conhecimento prévio sobre teste de software;

- o programa utilizado já havia sido submetido a alguns testes iniciais, todavia ainda foi possível descobrir uma grande quantidade de defeitos nativos ao programa;
- as regras e os materiais estão disponíveis em outros estudos iniciais [15], podendo ser consultados.

No que diz respeito à ameaça de construção, de início adotou-se como estratégia a condução de um MSL [7] para identificar vários fatores positivos e negativos da aplicabilidade de TE, a partir disto houve a identificação dos elementos de jogos, a descrição das regras, os propósitos e os materiais necessários. Todo esse processo ocorreu sempre com o acompanhamento e a validação de um especialista na área de Engenharia de Software com inúmeros trabalhos publicados sobre gamificação na educação. Assim, construiu-se e executou-se o estudo de caso obtendo resultados com aderência ao que se almejou investigar.

Por fim, quanto à ameaça de conclusão, é perceptível que os resultados qualitativos convergiram com os resultados quantitativos obtidos, e que isso permite afirmar que a aderência dos resultados representa um fator bastante relevante para a aplicabilidade do *Framework* Gamificado sobre ensino-aprendizagem de TE. A validade conclusiva baseia-se nos resultados obtidos, evidenciando que a gamificação realmente auxiliou no processo de ensino-aprendizagem dos estudantes.

7 CONCLUSÃO

Percebe-se que os alunos estavam satisfeitos com o estudo de caso, porém gostariam de mais sessões de teste, uma vez que era uma prática mais envolvente com a gamificação, causando uma diminuição de qualquer resistência à interação. Diante disso, percebe-se que a questão lúdica causou maior interesse nos alunos, quando se evidenciou que eles se sentiam estimulados a obter um bom desempenho.

O *Framework* Gamificado foi planejado e projetado para treinar pessoas para aplicar autonomamente a abordagem de TE de maneira estruturada e sistemática. Assim, era importante envolver etapas de treinamento (**Formar o Pirata**) precisamente na perspectiva de obter resultados semelhantes aos ocorridos.

Com relação à análise de habilidades por tipo de erro, não foram coletados dados para a mensuração específica, pois será um estudo no futuro, devido as estratégias de exploração utilizadas para influenciar os tipos e as severidades dos defeitos. No presente estudo 7 estratégias de exploração [18] foram exemplificadas para auxiliar na detecção de defeitos, pois os participantes não possuíam conhecimento sobre o SUT (*System Under Test* – Sistema Sob Teste) e, de acordo com relatos dos participantes, essas estratégias facilitaram a detecção de defeitos, principalmente sobre usabilidade, funcionalidades e GUI (*Graphic User Interface*). Além disso, pretende-se aplicar o *Framework* Gamificado em um contexto industrial, para uma análise focada na observação da sua eficácia e eficiência.

AGRADECIMENTOS

Os autores gostariam de agradecer a CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela concessão

de Bolsa de Pesquisa de Mestrado para o aluno-autor deste trabalho, a partir do PPGCC/UFPA. Além disso, agradecemos aos participantes dos estudos de casos relatados neste trabalho pelo comprometimento em realizar as atividades definidas no *framework* gamificado.

REFERÊNCIAS

- [1] R. Santos, P. Santos, C. Werner, G. Travassos, 2008. Using Experimentation to Support Research in Software Engineering Education in Brazil. In: *1 Fórum de Educação em Engenharia de Software*. DOI: 10.13140/2.1.2946.4002.
- [2] C. Kaner, 2008. *A Tutorial in Exploratory Testing*. QUEST.
- [3] D. Pfahl, H. Yin, M. Mantyla, J. Munch, 2014. How is Exploratory Testing Used?: A state of the Practice Survey. *EPPEM'14*, September 18-19, Torino, Italy. Copyright 2014 ACM. ISBN: 978-1-4503-2774-9/14/09.
- [4] J. Bach, 2004. *Exploratory Testing*. In: *The Testing Software engineer*, 2nd ed., E. van Veenendaal (Ed.) Den Bosch: UTN Publishers, pp. 253-265.
- [5] IEEE, 2014. *SWEBOOK V3.0: Guide to the Software Engineering Body of Knowledge*. Computer Society.
- [6] J. Itkonen e M. Mantyla, 2013. Are test cases needed? Replicated comparison between exploratory and test-case-based software testing. *Empirical Software Engineering*, pp. 1-40.
- [7] I. Costa e S. Oliveira, 2019. Um Estudo Baseado em Evidências sobre a Eficácia e Eficiência de Teste Exploratório. *16th International Conference on Information Systems & Technology Management – CONTECSI*. São Paulo, Brasil.
- [8] M. Fardo, 2013. Gamification Applied in Learning Environments. *Renote - Novas Tecnologias na Educação*. 11. ISSN 1679-1916.
- [9] K. Kapp, 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. North Carolina: Pfeiffer, 366 p.
- [10] K. Werbach e F. Hunter, 2012. *For The Win: How Game Thinking Can Revolutionize Your Business*. Filadélfia, Pensilvânia: Wharton Digital Press.
- [11] Y. Chou, 2016. *Actionable Gamification - Beyond Points, Badges, and Leaderboards*. Octalysis Media.
- [12] F. Benitti, 2015. Evaluating Learning Objects for Teaching Software Testing. *Nuevas Ideas en Informática Educativa TISE*. ISBN: 978-956-19-0929-8.
- [13] I. Costa e S. Oliveira, 2019. A Systematic Strategy to Teaching of Exploratory Testing Using Gamification. *14th International Conference on Evaluation of Novel Approachs to Software Engineering - ENASE*. Greece.
- [14] J. Bach, 2000. Session-Based Test Management, *STQE*, vol. 2, no. 6.
- [15] I. Costa e S. Oliveira, 2019. A Gamified and Systematic Approach to Teaching and Learning of Exploratory Testing. *16th International Conference on Information Systems & Technology Management – CONTECSI*. São Paulo, Brasil.
- [16] I. Costa, S. Oliveira, L. Cardoso, A. Ramos, R. Sousa, 2019. Uma Gamificação Para Ensino e Aprendizagem de Teste Exploratório de Software: Aplicação em um Estudo Experimental. *XVIII Simpósio Brasileiro de Jogos e Entretenimento Digital (Education Track – Short Papers)*. Rio de Janeiro, Brasil.
- [17] M. Santos, J. Grechi, P. Bermejo, 2010. Assessing the Impact of SCRUM on Software Development Using SWOT Analysis. In: *XXX ENEGEP*. São Paulo.
- [18] M. Micallef, C. Porter, A. Borg, 2016. Do Exploratory Testers Need Formal Training? An Investigation Using HCI Techniques. *IEEE Ninth International Conference on Software Testing, Verification and Validation Workshops*. DOI 10.1109/ICSTW.2016.31.