

Abordagem Flipped Classroom em Comparação com o Modelo Tradicional de Ensino: uma Investigação Empírica no Âmbito de Teste de Software

Leo Natan Paschoal¹, Laíza Ribeiro Silva¹, Simone R. S. de Souza¹

¹Instituto de Ciências Matemáticas e de Computação (ICMC) –
Universidade de São Paulo (USP)
Caixa Postal 668 – 13.566-970 – São Carlos – SP – Brasil

paschoalln@usp.br, laizaribeiro@usp.br, srocio@icmc.usp.br

Abstract. *This paper presents the results of an experimental evaluation in order to analysis the flipped classroom learning approach against the traditional learning approach. The study was developed in context of Software Testing course offered to the undergraduate students in Computer Science. This study evaluated the impact of flipped classroom approach regarding performance in learning and personal motivation. The results indicate that the flipped classroom approach present different results compared to traditional approach regarding performance in learning. In opposition, the personal motivation of students using flipped classroom approach is not different in relation to students learning with traditional classroom.*

Resumo. *Este artigo apresenta um estudo sobre o uso da abordagem de ensino flipped classroom comparada com a abordagem tradicional no contexto de Teste de Software. Em um experimento controlado, realizado com estudantes de graduação em Ciências de Computação, foi possível verificar o impacto da abordagem no que se refere ao desempenho no aprendizado e motivação pessoal. Os resultados sugerem que os estudantes que participaram do método de estudo baseado na abordagem flipped classroom obtiveram desempenhos diferentes do que estudantes que aprenderam com uma aula expositiva. Em contraste, a motivação pessoal dos estudantes que participaram da flipped classroom não difere dos estudantes que aprenderam com uma aula tradicional.*

1. Introdução

O Teste de Software é uma das áreas da Engenharia de Software que tem o objetivo de garantir a qualidade do software, buscando assegurar que os requisitos do mesmo sejam atendidos. Apesar de sua importância ser algo já confirmado, estudos vêm relatando que o mercado de trabalho está sofrendo com a falta de recursos humanos qualificados e com a disponibilização de mão de obra para atuar na área, uma vez que os egressos de cursos superiores de computação estão adentrando no mercado de trabalho com conhecimento muito limitado sobre o assunto [Astigarraga et al. 2010] e o meio acadêmico não está conseguindo atender satisfatoriamente à demanda existente [Barbosa et al. 2007].

Neste cenário, alguns mecanismos de apoio ao ensino e aprendizagem de Teste de Software vêm sendo estabelecidos. No estudo de [Valle et al. 2015], o autor apresenta um mapeamento sistemático em que são especificados quais os tipos de mecanismos e as abordagens utilizadas. Com base nos resultados obtidos pelos autores, observa-se uma predominância de recursos didáticos digitais construídos para apoiar Teste de Software,

mas são poucos os trabalhos que se detém nos aspectos relativos à abordagem de ensino utilizada. Entretanto, de acordo com [Smith et al. 2012], as abordagens de ensino utilizadas atualmente no âmbito de Teste de Software não são eficientes, sendo um fator que pode causar a desmotivação dos estudantes, uma vez que os professores de Teste de Software estão se deparando com estudantes muito desmotivados, com pouco ou nenhum interesse em se tornarem profissionais desta área [Allison and Joo 2015]. Por outro lado, a capacitação em Teste de Software é uma demanda crescente, dada a necessidade cada vez maior de software em diferentes domínios de aplicação.

No âmbito educacional, relativo à metodologia de ensino, existem alguns resultados sobre o uso de estratégias de ensino diferentes das tradicionais, dado que busca-se encontrar a abordagem mais adequada para que o estudante tenha uma aprendizagem efetiva [Carvalho and Ramos 2015]. Para tanto, a abordagem pedagógica utilizada em sala de aula deve explorar aspectos que proporcionam ao estudante oportunidades para o sucesso profissional. Desse modo, algumas abordagens de ensino são bem vistas pela comunidade científica, como a *flipped classroom* [Gannod et al. 2008].

Flipped classroom é considerada uma abordagem que possui o propósito de instituir uma mudança sistemática na sala de aula, em que o aluno tem uma postura pró-ativa e estuda os fundamentos teóricos de um conteúdo antes da aula e, em sala de aula, poderá praticar o conteúdo. Desse modo, ocorre uma descentralização da atenção do professor para o estudante, e por consequência, aumenta-se o envolvimento do aluno [Davies et al. 2013]. Nesse sentido, o professor usufruirá de mais espaço na disciplina para potencializar a aprendizagem do estudante [Carvalho and Ramos 2015], direcionando-o para um aprendizado individualizado, uma vez que seu papel se constituirá em orientar o aprendiz, indicando os conteúdos que os alunos deverão estudar, preparando atividades práticas para serem realizadas presencialmente, e norteá-los, trabalhando com as controvérsias e os equívocos que eles tenham cometido.

Em algumas pesquisas pertencentes a área de educação, estudos discorrem sobre a importância dos elementos afetivos nos processos de ensino-aprendizagem, principalmente relacionados à motivação [Tao et al. 2016]. Nesta perspectiva, um modelo que tem direcionado a avaliação de métodos de ensino e recursos digitais, no que se refere ao estímulo motivacional, é o modelo ARCS (Atenção, Relevância, Confiança e Satisfação) proposto por [Keller 1987]. Keller (1987) também desenvolveu um instrumento baseado nesse modelo, denominado IMMS (*Instructional Materials Motivation Survey*), que busca investigar as expectativas e valores que orientam os comportamentos dos estudantes para aprender. Trabalhos apontam o uso do instrumento IMMS em diferentes contextos educacionais relacionados a *flipped classroom* [Tao et al. 2016].

Apesar do uso da abordagem de *flipped classroom* estar se popularizando em diversas áreas, observa-se a falta de estudos que apresentem resultados empíricos sobre a utilização desta abordagem no âmbito de Teste de Software. Assim, este artigo tem o objetivo de apresentar a execução e os resultados de um experimento controlado, no qual o objetivo é comparar o desempenho no aprendizado e o estímulo motivacional de alunos que aprendem Teste de Software com uma abordagem de *flipped classroom* com o desempenho no aprendizado e o estímulo motivacional de alunos que aprendem Teste de Software com uma abordagem tradicional de ensino.

Este artigo está estruturado da seguinte forma. Na Seção 2 é apresentado o protocolo do experimento e os detalhes relacionados à sua execução. Na Seção 3 são descritos os resultados obtidos e sua análise. Na seção 4 são apresentadas as conclusões do trabalho

e também algumas perspectivas para investigações futuras.

2. Protocolo do Experimento

Este experimento foi conduzido seguindo as recomendações de [Wohlin et al. 2012], compreendendo as seguintes etapas: definição do experimento, planejamento, execução e análise. Com base nos procedimentos definidos e resultados, um pacote de laboratório¹ foi construído com o intuito de auxiliar na replicação do experimento e até mesmo a fazer com que o estudo seja melhor compreendido. Esse documento contém os recursos didáticos, formulários, atividades e o termo de consentimento.

2.1. Definição e Hipóteses

As metas do experimento foram especificadas em cinco partes, como um paradigma semelhante ao *Goal/Question/Metric (GQM)*: **analisar** a abordagem *flipped classroom* no ensino de Teste de Software em comparação com uma abordagem tradicional de ensino, **com a finalidade de** verificar sua aplicabilidade, **com respeito a** motivação proporcionada e ao desempenho no aprendizado dos alunos, **do ponto de vista** dos pesquisadores, **no contexto de** estudantes do curso de graduação em Ciências de Computação, matriculados na disciplina de Engenharia de Software, estudando o conteúdo de Teste de Software, mais especificamente a unidade de Teste Funcional. Aqui, entende-se por abordagem tradicional de ensino aquela em que o aluno não se prepara antes da aula e o professor ministra o assunto e, eventualmente, aplica uma atividade de fixação do conteúdo. Essas metas foram estipuladas a partir da definição de duas questões de pesquisa (QP): **(QP1)** Existe uma diferença entre os aspectos motivacionais de estudantes que aprendem Teste de Software com uma abordagem de ensino tradicional daqueles que aprendem com uma abordagem de *flipped classroom*? **(QP2)** Existe uma diferença entre o desempenho no aprendizado de estudantes que aprendem Teste de Software com uma abordagem de ensino tradicional daqueles que aprendem com uma abordagem de *flipped classroom*?

A Tabela 1 apresenta as hipóteses que serão investigadas no estudo.

Tabela 1. Formulação de hipóteses

Hipóteses para QP1	Hipóteses para QP2
H_0 : Não existe diferença significativa em termos de motivação entre estudantes que aprendem com uma abordagem tradicional de ensino e uma abordagem de <i>flipped classroom</i> .	H_0 : Não existe diferença significativa em termos de desempenho entre estudantes que aprendem com uma abordagem tradicional de ensino e uma abordagem de <i>flipped classroom</i> .
H_1 : Existe diferença significativa em termos de motivação entre estudantes que aprendem com uma abordagem tradicional de ensino e uma abordagem de <i>flipped classroom</i> .	H_1 : Existe diferença significativa em termos de desempenho entre estudantes que aprendem com uma abordagem tradicional de ensino e uma abordagem de <i>flipped classroom</i> .

2.2. Materiais e Medidas

Neste estudo experimental foram considerados dois diferentes tipos de métricas para avaliar a abordagem de *flipped classroom* no âmbito de Teste de Software. A primeira métrica está relacionada à motivação do estudante ao utilizar a abordagem *flipped classroom* e a abordagem tradicional de ensino. A segunda métrica está relacionada ao desempenho no aprendizado obtido pelos alunos ao estudarem por intermédio da abordagem de *flipped classroom* e abordagem tradicional de ensino.

¹Pacote disponível em: <https://goo.gl/H8m9pW>

A motivação foi mensurada por meio de um instrumento validado, denominado IMMS [Keller 1987]. Esse instrumento é constituído por trinta e seis perguntas e cada uma delas possui cinco opções de resposta, no formato de escala Likert, que varia de um intervalo de 1 até 5, sendo: (1) não é verdadeiro; (2) ligeiramente verdadeiro; (3) moderadamente verdadeiro; (4) na maior parte verdadeiro; (5) muito verdadeiro. Dentre as trinta e seis questões, doze mensuram a atenção, nove medem a relevância, nove determinam a confiança e seis estimam a satisfação do usuário. Uma vez que o questionário IMMS é constituído pelas quatro subescalas do modelo ARCS, foi possível comparar os pontos obtidos pelos participantes em cada uma das subescalas.

O desempenho no aprendizado é mensurado de duas maneira distintas, uma vez que para responder a segunda questão de pesquisa, buscou-se averiguar se há ou não diferença significativa em relação ao conhecimento teórico e o conhecimento prático retido pelos estudantes ao utilizarem as diferentes abordagens de ensino. Os autores desta pesquisa estão admitindo para este experimento as seguintes especificações para diferenciar conhecimento teórico e conhecimento prático: o conhecimento teórico retido é baseado exclusivamente em terminologias e conceitualizações, não abrangendo aplicação prática; o conhecimento prático retido é baseado exclusivamente em aplicação prática.

Para avaliar o conhecimento teórico que os alunos adquiriram a partir da utilização das abordagens, foi construído um questionário composto por treze questões de múltipla escolha que englobam aspectos teóricos do conteúdo de Teste Funcional. Cada questão tem 5 opções de resposta em que apenas uma delas é correta. As questões foram selecionadas e retiradas de concursos e exames de certificação da área de Teste de Software. As respostas são convertidas em uma nota, sendo que o máximo que o aluno pode obter é 10 pontos. Para montagem deste questionário, dois especialistas da área foram consultados.

Com o intuito de estimar o conhecimento prático retido pelos estudantes a partir da utilização das abordagens, foi utilizado um exercício recomendado pelos mesmos especialistas que haviam participado da construção do questionário desenvolvido para coletar os dados relacionados ao conhecimento teórico. Um exercício semelhante foi utilizado na pesquisa de [Barbosa et al. 2007]. Essa atividade consiste na aplicação do critério da técnica de Teste Funcional denominado "Particionamento de Equivalência" em uma determinada especificação de requisitos. Os números médios de casos de teste gerados são considerados no momento de avaliação. Assim, as respostas dos estudantes são convertidas em uma nota, sendo que o máximo que o aluno pode obter é 10 pontos.

Em adição aos instrumentos destinados à mensuração de dados relacionados ao desempenho e motivação, um questionário com o objetivo de coletar alguns dados sociodemográficos dos participantes foi construído, para caracterizar os sujeitos da pesquisa. Além disso, um termo de consentimento foi desenvolvido, com o propósito de informar e esclarecer os sujeitos da pesquisa para que possam tomar sua decisão de forma justa e sem constrangimentos sobre a sua participação no experimento.

Além dos instrumentos mencionados, materiais didáticos foram preparados para a execução do experimento. Para apoiar a execução da abordagem de *flipped classroom*, foram elaborados slides e selecionados dois vídeos com licença livre, além de um artigo científico já publicado. O conteúdo existente no artigo e nos vídeos serve para complementar os conteúdos dos slides, uma vez que nos momentos de estudos, os estudantes de *flipped classroom* não terão contato com o professor. Assim, admite-se para esta pesquisa que os detalhes que seriam explanados em uma aula tradicional estão contidos no artigo, nos vídeos e slides. Para apoiar a execução da abordagem tradicional de ensino, foram

utilizados apenas os slides.

2.3. Seleção dos Sujeitos

Foram convidados para participar do experimento cento e seis estudantes de graduação, todos devidamente matriculados na disciplina de Engenharia de Software, do curso de Ciências de Computação de uma instituição de ensino superior. No entanto, apenas cinquenta e dois conduziram a atividade de forma completa para analisar a amostra. Nesse sentido, a amostragem foi do tipo não-probabilística, selecionada por conveniência. Salienta-se que apesar do experimento estar no contexto da disciplina de Engenharia de Software e envolver os conteúdos de uma unidade da disciplina, a participação dos estudantes foi voluntária e não contabilizou como atividade avaliativa da disciplina, isto é, os estudantes que participarem do experimento não foram beneficiados.

2.4. Execução do experimento

O experimento foi composto por três fases, distribuídas ao longo de uma semana, compreendendo duas aulas. A primeira aula ministrada em uma terça-feira e a segunda aula em uma sexta-feira, passando-se exatamente três dias entre as aulas. Foram utilizadas apenas duas aulas, pois a execução do experimento precisava simular um cenário real de aula, não comprometendo as atividades da disciplina. A seguir são descritas as atividades executadas pelos sujeitos em cada uma das fases.

Fase 1 (em sala de aula): Os pesquisadores apresentaram a proposta do experimento para o grupo de alunos matriculados. Os que aceitaram participar da pesquisa preencheram um termo de consentimento, autorizando a liberação de dados. Os que não aceitaram participar, foram liberados da aula e foram informados que deveriam procurar o monitor da disciplina para inteirar-se sobre o conteúdo.

Todos os participantes que assinaram o termo realizaram uma atividade (nesta fase denominada como pré-teste conceitual). Essa atividade contém questões sobre o conteúdo de Teste Funcional. O objetivo da aplicação foi mensurar o conhecimento teórico que os alunos possuíam sobre o conteúdo, antes de ter contato com o conteúdo na disciplina. Após os estudantes realizarem a atividade, foi feito um sorteio randômico dos participantes que iriam fazer parte do grupo controle e do grupo experimental. Assim, os estudantes que foram selecionados para compor o grupo experimental são aqueles que utilizaram a abordagem de *flipped classroom* e os selecionados para compor o grupo controle são aqueles que aprenderam por meio da abordagem tradicional de ensino.

Após a distribuição dos participantes, os integrantes do grupo controle foram liberados da aula e não receberam nenhuma orientação sobre o experimento, já os integrantes do grupo experimental receberam orientações sobre as ações que deveriam realizar. Os pesquisadores disponibilizaram no Ambiente Virtual de Aprendizagem Moodle todo o material didático preparado para o grupo experimental. Para tanto, os pesquisadores criaram um grupo dentro da disciplina, neste ambiente, a fim de disponibilizar o material didático somente aos integrantes que utilizavam *flipped classroom* para estudar.

Fase 2 (em ambiente externo à sala de aula): Os alunos que compuseram o grupo experimental estudaram o conteúdo, em ambiente externo à sala de aula. Eles poderiam escolher qualquer ambiente como laboratório da universidade, biblioteca, entre outros. Um fórum de discussões foi disponibilizado no Moodle, com a intenção de permitir que os estudantes pudessem interagir com colegas e com o professor, disponibilizando suas dúvidas, a fim de saná-las.

Fase 3 (em sala de aula): Esta fase foi dividida em duas etapas:

Primeira etapa: Os alunos do grupo controle tiveram uma aula tradicional com o professor da disciplina sobre o mesmo conteúdo (Teste Funcional) disponibilizado para os estudantes do grupo experimental. Nesta etapa, os alunos que faziam parte do grupo experimental não estavam em sala de aula. Nessa ocasião, o professor da disciplina ministrou o conteúdo em um período de aproximadamente 50 minutos.

Segunda etapa: Após o professor explicar os conteúdos, os integrantes do grupo experimental que não estavam presentes em sala na primeira etapa desta fase, foram convidados a adentrar na sala de aula. Em um momento posterior, o professor da disciplina disponibilizou um espaço de tempo de aproximadamente 20 minutos para os estudantes apresentarem suas dúvidas. Posteriormente, dois exercícios foram disponibilizados para todos os estudantes que estavam participando do experimento. Esses exercícios foram realizados individualmente por cada participante.

A primeira atividade disponibilizada para os estudantes foi a mesma que havia sido realizada na primeira aula (nesta fase denominada como pós-teste conceitual), sendo utilizada nesta fase com o intuito de mensurar o conhecimento teórico que os alunos possuíam sobre o conteúdo, após ter contato com o conteúdo na disciplina. A segunda atividade foi disponibilizada com o objetivo de mensurar o conhecimento prático dos estudantes (denominada como pós-teste prático). Por fim, ao finalizarem as atividades, os participantes responderam o instrumento IMMS e o questionário demográfico.

3. Análise dos Resultados

Após a execução de todas as fases do experimento, os dados foram tabulados e analisados com auxílio do software BioEstat (versão 5.3)². Para realização dos testes estatísticos, foi adotado um intervalo de confiança para $\alpha = 0,05$. Após α ser especificado, foi definido que durante a comparação de médias entre dois grupos, utilizar-se-ia o teste t de student no caso de amostras dependentes e distribuição gaussiana nos dois grupos. No caso de amostras dependentes e distribuição não gaussiana foi utilizado o teste de Wilcoxon; caso contrário, foi usado o teste de Mann-Whitney.

3.1. Caracterização dos sujeitos

O grupo controle foi constituído por um total de 24 participantes e o grupo experimental foi composto por 28 participantes. Nenhum dos participantes possui formação em nível superior e nenhum deles trabalha ou já trabalhou no âmbito de Teste de Software.

3.2. Questão de Pesquisa 1

Buscando averiguar se existe diferença significativa em termos de motivação entre estudantes que aprendem com abordagem tradicional de uma abordagem de *flipped classroom*, foram conduzidas análises referentes ao grau geral de motivação e a cada um dos componentes de motivação (Atenção, Relevância, Confiança e Satisfação). Assim, para acompanhar a pontuação obtida pelos participantes, foi calculado a pontuação média para cada subescala, seguindo as especificações descritas por [Keller 1987].

O grupo que utilizou a abordagem *flipped classroom* apresentou a maior pontuação no questionário de motivação (μ Flipped Classroom= 3,14, μ Tradicional= 2,97). Levando em consideração as quatro escalas do modelo ARCS: O grupo que utilizou a abordagem

²Disponível em: <https://goo.gl/5BcFTV>

flipped classroom apresentou a maior pontuação na média de resposta na escala de atenção (μ Flipped Classroom= 3,27, μ Tradicional= 2,93), na escala de confiança (μ Flipped Classroom= 3,41, μ Tradicional= 3,27) e na escala de satisfação (μ Flipped Classroom= 2,59, μ Tradicional= 2,41). O grupo que utilizou a abordagem tradicional apresentou a maior pontuação na média de resposta na escala de relevância (μ Flipped Classroom= 3,05, μ Tradicional= 3,10).

Ao contrastar as pontuações obtidas pelo grupo controle com o grupo experimental no contexto das escalas ARCS, considerando as pontuações da mediana, obtidos em cada escala por grupo, observa-se que a sequência de respostas dos participantes do grupo experimental demonstram supremacia às que foram indicadas pelos alunos do grupo controle em todas as escalas do modelo ARCS, uma vez que atenção (μ Flipped Classroom= 3,25, μ Tradicional= 3,08), relevância (μ Flipped Classroom= 3,33, μ Tradicional= 3,28), confiança (μ Flipped Classroom= 3,17, μ Tradicional= 3,11) e satisfação μ Flipped Classroom= 2,5, μ Tradicional= 2,09).

Ao analisar os dados coletados com as algumas técnicas de estatística descritiva, percebe-se que o grupo experimental (μ Flipped Classroom) obteve na maioria das vezes resultados superiores aos apresentados pelo grupo controle (μ Tradicional). Uma possível explicação está relacionada ao fato de que a abordagem *flipped classroom* permite que o estudante tenha mais liberdade para manipular o conteúdo, aprendendo de acordo com o seu ritmo e isso pode afetar a motivação intrínseca do estudante.

Com a finalidade de verificar as tendências existentes entre os diferentes tipos de abordagens e os quatro componentes de motivação, um gráfico do tipo radar (Figura 1) foi desenvolvido a partir das médias de cada componente. Conforme pode ser observado, enquanto que o grupo experimental teve uma disposição maior para o componente relevância, o grupo controle demonstrou uma disposição maior para o componente confiança. Acredita-se que os estudantes que aprenderam com a abordagem *flipped classroom* reconhecem quão relevante é a abordagem, mas se sentiram pouco confiantes ao utilizá-la, porque é a primeira vez que estudam com a abordagem *flipped classroom*. Já os estudantes que aprenderam com a abordagem tradicional de ensino se sentiram mais seguros que os estudantes de *flipped classroom*, uma vez que estavam acostumados em aprender com esta abordagem. Por fim, ao considerar os valores relacionados a um ponto central, percebe-se que existe uma supremacia no que se refere a abordagem de *flipped classroom*, sugerindo que estudantes do grupo experimental sentiram-se mais motivados que estudantes do grupo controle nas quatro subescalas do modelo ARCS.

Por fim, ao realizar o teste de hipótese, foi verificado que não existe diferença significativa entre as médias de pontos obtidos pelos integrantes do grupo controle e do grupo experimental, no que se refere a motivação, uma vez que o p-value (0,3047) do teste *t* de *student* é maior que alfa, sendo que: graus de liberdade = 50 e *t* observado = -1,0370. Ao considerarmos o estado da arte relacionado a abordagem de *flipped classroom* e o uso do modelo ARCS para mensurar motivação, [Tao et al. 2016] também não conseguiram observar discrepâncias significativas nos resultados entre os grupos a partir da análise estatística. Assim, os resultados obtidos corroboram com o que existe na literatura, porém, é importante salientar que não é possível generalizar para outros domínios, porque a aplicação foi instanciada para Teste Funcional de Software.

3.3. Questão de Pesquisa 2

Com o objetivo de analisar se existe diferença significativa de desempenho entre estudantes que aprendem com abordagem de *flipped classroom*, foram analisados os dados

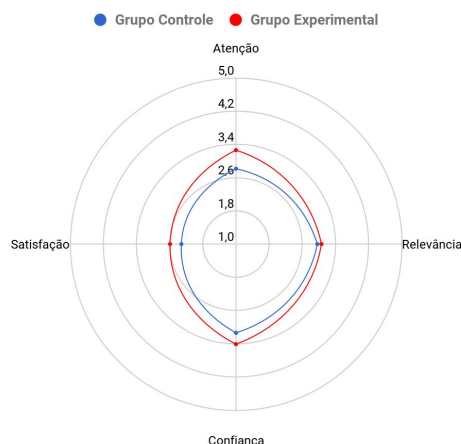


Figura 1. Relação entre os componentes de motivação e abordagens de ensino

coletados no pré-teste, pós-teste conceitual e pós-teste prático, levando em conta técnicas de estatística descritiva e testes de hipóteses. De acordo com a aplicação da técnica de estatística descritiva nos dados, o grupo que utilizou a abordagem *flipped classroom* apresentou a maior média no pós-teste conceitual (μ Flipped Classroom= 6,47, μ Tradicional= 5,93), em contraposição a abordagem tradicional de ensino apresentou a maior média no pós-teste prático (μ Flipped Classroom= 1,90, μ Tradicional= 1,97).

Observando a mediana das pontuações no pré-teste, pós-teste conceitual e pós-teste prático dos grupos, nota-se que para todos os casos, os estudantes do grupo experimental obtiveram pontuações maiores do que os estudantes do grupo controle, uma vez que pré-teste (μ Flipped Classroom= 4,62, μ Tradicional= 3,85), pós-teste conceitual (μ Flipped Classroom= 6,47, μ Tradicional= 5,93) e pós-teste prático (μ Flipped Classroom= 1, μ Tradicional= 0,88). Percebe-se que os resultados obtidos no pós-teste prático foram muito baixos, revelando que os estudantes necessitam de mais tempo para entender o conteúdo e realizar aplicações práticas em ambas as abordagens.

O conhecimento adquirido a partir das intervenções empregadas nos grupos foi analisado. Para tanto, comparou-se o pré-teste com o pós-teste conceitual. Observa-se na Figura 2 que a partir da utilização do teste de Wilcoxon, percebe-se que existe uma diferença significativa entre os estudantes que aprenderam com a abordagem de *flipped classroom*. Com tais dados, percebe-se que a intervenção realizada fez com que ocorresse uma variação nas notas obtidas pelos estudantes do grupo controle e pelo grupo experimental. Nesse sentido, acredita-se que a intervenção fez com que os estudantes de ambos os grupos obtivessem notas maiores no pós-teste conceitual do que no pré-teste, revelando que os alunos aprenderam com ambas as abordagens.

Apesar dos participantes desse experimento terem sido distribuídos randomicamente para os grupos controle e experimental, não houve uma verificação se algum grupo foi beneficiado ou não antes da intervenção ser realizada. Assim, durante o teste de hipóteses, com o objetivo de confirmar que nenhum grupo foi beneficiado, foi verificado se existe diferença significativa entre as notas obtidas no pré-teste pelos participantes do experimento. O teste de Mann-Whitney revela que não existe diferença estatística significativa, confirmando que nenhum grupo foi beneficiado.

Por fim, para responder a segunda questão de pesquisa por intermédio de teste de hipóteses, foi realizada uma comparação entre os pontos do pós-teste conceitual obtidos

Pré-teste $\bowtie$ Pós-teste conceitual	
	<i>p-value</i> (Wilcoxon)
Tradicional	0,0003
Flipped Classroom	0,0092

Desempenho no aprendizado	
	<i>p-value</i> (Mann-Whitney)
Pré-teste $\mu_{\text{Tradicional}}$ $\bowtie$ Pré-teste $\mu_{\text{Flipped Classroom}}$	0,6073
Pós-teste conceitual $\mu_{\text{Tradicional}}$ $\bowtie$ Pós-teste conceitual $\mu_{\text{Flipped Classroom}}$	0,0495
Pós-teste prático $\mu_{\text{Tradicional}}$ $\bowtie$ Pós-teste prático $\mu_{\text{Flipped Classroom}}$	0,2870

Figura 2. Testes de hipóteses

pelos integrantes do grupo controle com os pontos dos integrantes do grupo experimental. Percebe-se que existe uma diferença significativa entre as pontuações, indicando que uma das abordagens destaca-se em relação a outra. Observando a análise de estatística descritiva, acredita-se que os estudantes que aprenderam com *flipped classroom* obtiveram as melhores pontuações. Também foram confrontados os pontos do pós-teste prático de ambos os grupos, usando o teste de Mann-Whitney, e observou-se que não existe diferença estatística significante. Portanto, existe diferença significativa em termos de desempenho entre estudantes que aprendem com uma abordagem tradicional de ensino e uma abordagem de *flipped classroom*. Porém, não é possível afirmar tal condição ao contrastar os dados do pós-teste prático do grupo controle com o do grupo experimental.

4. Considerações Finais

Esta pesquisa teve o intuito de apresentar um experimento conduzido no contexto de Teste de Software para estudantes da disciplina de graduação de Engenharia de Software. O objetivo foi averiguar se existe diferença significativa no que se refere ao desempenho obtido e motivação pessoal entre alunos que aprendem com uma abordagem de ensino de *flipped classroom* em oposição a abordagem de tradicional de ensino.

A partir da análise dos resultados, percebe-se que não existe diferença significativa no que se refere à motivação pessoal, mas há uma diferença significativa no que se refere ao desempenho no aprendizado. Todavia, salienta-se que só existe diferença significativa no que se refere ao desempenho no aprendizado quando considerado a aplicação de uma atividade que não demanda do estudante a realização de uma atividade prática. Pressupõe-se que os estudantes não obtiveram uma pontuação maior no pós-teste prático, pois essa atividade requer um processo de reflexão maior que o exigido no pós-teste conceitual, exigindo habilidades multidisciplinares como a interpretação de requisitos, familiaridade com desvios condicionais e lógica de programação.

Como se tem ciência que os resultados obtidos podem ser consequência de alguma ameaça à validade (*e.g.* influências nos exercícios escolhidos, didática do professor, tempo de dedicação dos estudantes), este estudo acabou oportunizando aos estudantes a familiarizarem-se com o fenômeno em questão, a fim de obter uma percepção e novos *insights* em relação ao trabalho. Assim, pretende-se realizar replicações desse experimento com novas turmas de alunos.

Com o objetivo de compreender os motivos pelos quais os aspectos motivacionais dos estudantes que aprendem com *flipped classroom* e com a abordagem tradicional não apresentaram diferença, pretende-se contactar esses alunos e identificar os seus estilos

cognitivos. A partir desta identificação, é possível verificar a preferência por materiais didáticos, já que estes podem ter influenciado nos resultados. Adicionalmente, novos estudos precisam ser realizados, considerando outros conteúdos de Teste de Software, pois o conteúdo explorado também pode ter influenciado nos resultados, e uso de *flipped classroom* em outros conteúdos pode gerar resultados mais favoráveis à abordagem.

5. Agradecimentos

Os autores agradecem a CAPES pelo apoio financeiro e os estudantes da disciplina de Engenharia de Software (ICMC-USP – 1º semestre de 2017) por aceitarem participar do estudo experimental.

Referências

- Allison, M. and Joo, S. F. (2015). An adaptive delivery strategy for teaching software testing and maintenance. In *10th International Conference on Computer Science Education (ICCSE)*, pages 237–242.
- Astigarraga, T., Dow, E. M., Lara, C., Prewitt, R., and Ward, M. R. (2010). The emerging role of software testing in curricula. In *IEEE Transforming Engineering Education: Creating Interdisciplinary Skills for Complex Global Environments*, pages 1–26.
- Barbosa, E. F., Souza, S. R. S., Domingues, A. L. S., Chan, A., Höhn, E. N., and Maldonado, J. C. (2007). Uma experiência no ensino de inspeção e teste de software. In *VI Simpósio Brasileiro de Qualidade de Software*, pages 309–324.
- Carvalho, R. J. O. and Ramos, M. A. S. (2015). Flipped classroom - centrar a aprendizagem no aluno recorrendo a ferramentas cognitivas. In *International Conference on ICT in Education*, pages 369–381.
- Davies, R. S., Dean, D. L., and Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Educational Technology Research and Development*, 61(4):563–580.
- Gannod, G., Burge, J., and Helmick, M. (2008). Using the inverted classroom to teach software engineering. In *30th ICSE*, pages 777–786.
- Keller, J. M. (1987). Development and use of the arcs model of instructional design. *Journal of Instructional Development*, 10(3):02–10.
- Smith, J., Tessler, J., Kramer, E., and Lin, C. (2012). Using peer review to teach software testing. In *Ninth Annual International Conference on International Computing Education Research*, pages 93–98.
- Tao, S.-Y., Huang, Y.-H., and Tsai, M.-J. (2016). Applying the flipped classroom with game-based learning in elementary school students' english learning. In *International Conference on Educational Innovation through Technology (EITT)*, pages 59–63.
- Valle, P. H. D., Barbosa, E. F., and Maldonado, J. C. (2015). Um mapeamento sistemático sobre ensino de teste de software. In *XXVI Simpósio Brasileiro de Informática na Educação*, pages 71–80.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer Berlin Heidelberg.