

Investigação Empírica sobre os Efeitos da Gamificação de um Juiz Online em uma Disciplina de Introdução à Programação

Title: An Empirical Investigation on the Effects of Gamifying an Online Judge in an Introductory Programming Course

Ralph B. S. Ribeiro
Universidade Federal do Amazonas
rbsr@icomp.ufam.edu.br

Leandro S. G. Carvalho
Universidade Federal do Amazonas
galvao@icomp.ufam.edu.br

David B. F. Oliveira
Universidade Federal do Amazonas
david@icomp.ufam.edu.br

Elaine H. T. de Oliveira
Universidade Federal do Amazonas
elaine@icomp.ufam.edu.br

Marcela Pessoa
Univ. do Estado do Amazonas
msppessoa@uea.edu.br

Resumo

CONTEXTO. Este artigo apresenta um juiz online gamificado desenvolvido com a finalidade de motivar estudantes de graduação iniciantes no aprendizado de programação. O sistema foi aplicado em 11 turmas de diferentes cursos de engenharia e ciências exatas da Universidade Federal do Amazonas no 1º período letivo de 2017. **OBJETIVO.** O objetivo geral deste estudo é avaliar o impacto desse juiz online gamificado em uma disciplina introdutória de programação sobre o desempenho, motivação, experiência e percepção da aprendizagem dos estudantes. **MÉTODO.** Um quase-experimento foi realizado para mensurar a influência do sistema gamificado sobre o desempenho dos estudantes, comparando-se estatisticamente dados das turmas de 2017/1 (com gamificação) com as de 2016/1 (sem gamificação). Além disso, nas turmas de 2017/1, foi aplicado um questionário qualitativo para verificar a percepção dos estudantes em relação a três componentes: motivação, experiência do usuário e percepção da aprendizagem. **RESULTADOS.** Verificou-se um aumento significativo na nota final da disciplina entre os não desistentes, o que não dependeu da experiência prévia reportada pelos alunos, mas que se concentrou mais nos alunos não repetentes. Além disso, houve um aumento no número de logins, número de testes de código e tempo de uso do IDE do juiz online. Não se verificou diferença significativa no número de exercícios corretos, mas houve uma redução na variância, principalmente entre os alunos de desempenho mais baixo. Também nas turmas experimentais, houve uma diminuição da associação entre a nota da prova de Matemática do Enem e o desempenho em programação. Com respeito à percepção dos estudantes em termos dos três componentes da análise qualitativa, a maioria das respostas foi positiva. **CONCLUSÕES.** Embora a taxa de desistência da disciplina tenha aumentado no período letivo de uso do juiz online gamificado, os resultados reforçam as evidências sobre o potencial motivador da gamificação no processo de ensino-aprendizagem de programação.

Palavras-chave: Gamificação; Introdução a Programação; Juiz Online Gamificado.

Abstract

CONTEXT. This paper presents a gamified online judge developed to motivate non-major freshmen in programming classes (CS1). The system was applied in 11 classes of different engineering and science programs of Federal University of Amazonas throughout the first academic term of 2017. **OBJECTIVE.** The main objective of this study is to assess the impact of this gamified online judge in a CS1 course on student performance, motivation, experience and learning perception. **METHOD.** A quasi-experiment was carried out with the aim of measuring the influence of the gamified system on student performance, by statistically comparing the data of 2017/1 (with gamification) to 2016/1 (without gamification). In addition, in 2017/1 classes, a qualitative questionnaire was applied to measure students' perceptions on three components: motivation, user experience and learning perception. **RESULTS.** There was a significant increase in the final grade of CS1 among non-dropout students,

Cite as: Ribeiro, R. B. S., Carvalho, L. S. G., Oliveira, D. B. F., de Oliveira, E. H. T., & Pessoa, M. (2020). An Empirical Investigation on the Effects of Gamifying an Online Judge in an Introductory Programming Course (Investigação Empírica sobre os Efeitos da Gamificação de um Juiz Online em uma Disciplina de Introdução à Programação). *Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE)*, 28, 461-490. DOI: 10.5753/RBIE.2020.28.0.461

which did not depend on the reported previous experience in programming, but focused more on first-time students. In addition, there was an increase in the number of logins, number of code tests and IDE usage time. There was no significant difference in the number of correct exercises, but there was a reduction in variance, especially among students with lower performance. Also, in the experimental classes, there was a decrease in the association between the Enem Mathematics test score and the performance in programming. Regarding students' perceptions of the three components of qualitative analysis, most responses were positive. **CONCLUSIONS.** Although the dropout rate of CSI has increased during the academic term in which the gamified online judge was used, the results reinforce the evidence about the potential of gamification in motivating programming learning.

Keywords: Gamification; Introductory Programming; Gamified Online Judge.

1 Introdução

Programar é uma habilidade necessária para o exercício profissional em diversas áreas de conhecimento. Por conta disso, disciplinas de introdução à programação fazem parte da matriz curricular de vários cursos de graduação que não têm a computação como atividade-fim. Porém, essas disciplinas costumam apresentar altos índices de reprovação e desistência (Bosse & Gerosa, 2015; Chaves, Castro, Lima, Lima, & Ferreira, 2013).

Aprender a programar envolve mais do que apenas entender a sintaxe de uma linguagem de programação e manipular estruturas de dados. Essa habilidade também requer a mobilização de um conjunto de conhecimentos prévios e de processos metacognitivos de modo a desenvolver soluções de problemas por meio de código (Prather et al., 2018). Por isso, aprender a programar é considerada uma tarefa difícil. Os estudantes precisam dedicar tempo para resolver exercícios, a fim de desenvolver essa habilidade. É necessário aplicar estratégias de estudo bem diversas das que usualmente utilizam para o estudo de outros conhecimentos. Além disso, estudantes *non-majors*, ou seja, de cursos de graduação fora da área de computação, não costumam se sentir motivados a cursar essa disciplina, pois não percebem ainda como a programação poderá ser útil na resolução de problemas no seu domínio de atuação.

Além disso, a literatura cita outras dificuldades que são enfrentadas pelos estudantes iniciantes, como: baixa compreensão das noções básicas de programação, dificuldade de pensar algorítmicamente, uma base matemática fraca, uso de linguagens de programação com sintaxes rebuscadas para estudantes iniciantes, o não entendimento dos enunciados, entre outros. Ademais, na maioria das vezes os alunos iniciantes vêm de um sistema educacional fundamentado na memorização/reprodução de informações e mecanização de procedimentos, o que influencia o aproveitamento da disciplina (Silva, Melo, & Tedesco, 2018).

Uma metodologia muito discutida atualmente para aumentar a motivação dos estudantes é o uso de *softwares* e ferramentas gamificadas. Segundo Kapp (2012), a gamificação (do inglês “*gamification*”) pode ser definida como “*a utilização de mecânica, estética e pensamento baseados em jogos para engajar pessoas, motivar a ação, promover a aprendizagem e resolver problemas*”. Para que os alunos estejam motivados na prática diária de resolução de problemas, o processo de aprendizagem precisa de uma experiência interessante sob o ponto de vista do aluno. Nesse sentido, o uso de gamificação pode amenizar esse problema devido à predisposição que os jovens têm pelo ato de jogar (Campos, Gardiman, & Madeira, 2015).

Portanto, este estudo busca avaliar o impacto de um juiz *online* gamificado em uma disciplina introdutória de programação sobre o desempenho, motivação, experiência e percepção da aprendizagem dos estudantes. Para tanto, foram formuladas seis questões de pesquisa:

- **Q1:** Há influência da gamificação na nota do estudante em IPC?
- **Q2:** Há influência dos elementos de gamificação sobre a nota final em IPC?
- **Q3:** Há associação entre a nota no Enem e a nota final do estudante em IPC?
- **Q4:** Os estudantes se sentem motivados ao utilizarem o juiz online gamificado?
- **Q5:** Os estudantes se sentem confortáveis ao utilizarem o juiz online gamificado?

- **Q6:** Os estudantes percebem algum ganho de aprendizagem ao utilizarem o juiz online gamificado?

Com o intuito de investigar essas questões, utilizamos um ambiente de correção automática de códigos (juiz *online*) chamado CodeBench¹, desenvolvido na Universidade Federal do Amazonas (UFAM), e adicionamos técnicas de gamificação para promover o engajamento e motivar os alunos nas atividades de programação. Na Universidade Federal do Amazonas, a disciplina de Introdução à Programação de Computadores (IPC) é ministrada para 17 cursos de graduação em engenharia e de ciências exatas. Em 2016, a taxa de aprovação foi de apenas 43,3% e a taxa de desistência foi de 19,7% entre os 635 estudantes matriculados, o que motivou o presente estudo.

O artigo está organizado da seguinte forma: na Seção 2 apresenta-se a contextualização do trabalho com o referencial teórico a respeito do uso da gamificação na educação e sobre sistemas de juízes *online*. A Seção 3 descreve a ferramenta proposta neste trabalho. O projeto experimental e a metodologia de avaliação do sistema são descritos na Seção 4; e os resultados são discutidos na Seção 5. Por fim, a Seção 6 contém a conclusão e os trabalhos futuros.

2 Contextualização

Nesta seção, abordam-se o uso de gamificação na educação para melhoria da aprendizagem e sistemas de juízes online, com ênfase no sistema CodeBench.

2.1 Gamificação na Educação

A gamificação corresponde ao uso de mecanismos de jogos, tais como: competição, desafio, interação social e recompensa, em contextos não relacionados a jogos (Kapp, 2012). Segundo Peixoto, Silva, Gonçalves, e Vilela (2015), técnicas de gamificação em softwares educacionais estão cada vez mais sendo utilizadas para apoiar a aprendizagem. Mas qual a vantagem de se usar gamificação na educação? Fardo (2013) argumenta que um dos benefícios é disponibilizar aos alunos um sistema no qual eles consigam visualizar o impacto de suas ações e sentir que estão aprendendo, no decorrer da disciplina. Assim, da mesma forma que acontece nos jogos, fica mais fácil entender que atingir pequenos objetivos contribui para alcançar um objetivo maior.

Dessa maneira, para criarmos uma experiência de aprendizagem divertida e eficaz, precisamos conseguir transmitir para o sistema o “pensamento de *game*”, ou seja, necessitamos combinar os elementos de jogos com uma experiência de aprendizagem adequada (Figura 1) (Alves, 2015). Assim, como argumenta Kaap (2012), a gamificação pode ser utilizada para promover a aprendizagem, e cada elemento de jogo deve estar conectado entre si, pois os elementos, sozinhos, não são suficientes para transformar uma aula tradicional em uma experiência motivadora e engajadora, embora cada elemento seja importante e fundamental no processo geral da gamificação.

Werbach e Hunter (2012) mostram que a relação entre a mecânica, a dinâmica e os componentes, expressa na Figura 2, “*é essencial para o sucesso de um projeto com o uso de gamificação*”. Ainda segundo eles, a **dinâmica** é “*constituída por elementos responsáveis por atribuir coerência e padrões regulares à experiência*”. É ela que determina as regras e os elementos mais conceituais, como: restrições, emoções, narrativa, progressão e relacionamento (Alves, 2015). Já na **mecânica** dos jogos, estão os elementos que promovem a “ação”, ou seja, é como se eles fossem a “engrenagem” do sistema que faz a gamificação funcionar no decorrer da execução. Dentre eles, podemos citar: desafios, sorte, cooperação e competição, *feedback*,

¹ <http://codebench.icomp.ufam.edu.br/>

aquisição de recursos, recompensas, transações, turnos e estados de vitória. Por fim, na base da pirâmide, os **componentes** representam a essência da dinâmica e da mecânica do jogo, por exemplo: realizações, avatares, medalhas (*badges*), coleções, combates, desbloqueio de conteúdo, placares (*rankings*), níveis, pontos, exploração, entre outros.



Figura 1: Pensamento de game. Fonte: Adaptado de Alves (2015).



Figura 2: Pirâmide representando a hierarquia de elementos em um projeto de gamificação. Fonte: Adaptado de Werbach e Hunter (2012).

Este estudo busca dotar um juiz *online* com elementos de gamificação, para gerar uma experiência de aprendizagem motivadora em IPC, como será visto a seguir.

2.2 Sistemas de Juiz Online

Juízes *online* “são plataformas web cujo propósito é compilar, executar e avaliar programas submetidos por usuários” (Wasik, Antczak, Badura, Laskowski, & Sternal, 2018). Essas ferramentas, do ponto de vista pedagógico, segundo Francisco, Ambrósio, Pereira, e Fernandes (2018), podem estimular as mudanças na forma de ensinar e aprender programação devido a alguns fatores, como por exemplo, a facilidade do estudante de obter *feedback*.

Na Universidade Federal do Amazonas, utiliza-se o juiz online CodeBench para dar suporte a professores e estudantes em disciplinas de programação. Por meio desse sistema, professores e tutores das disciplinas podem disponibilizar exercícios de programação, listas e até mesmo provas para seus alunos, que por sua vez podem desenvolver soluções em um IDE (*Integrated Development Environment*) acoplado ao próprio sistema. Assim que o estudante submete o código de solução para um dado exercício, o sistema compara a *string* de saída do programa do estudante com a *string* cadastrada como correta, informando instantaneamente o resultado.

Após a adoção desse juiz online como ferramenta de apoio a uma metodologia de ensino híbrido de programação, houve um aumento no índice de aprovações dos alunos de IPC (Ribeiro, Oliveira, Carvalho, & Oliveira, 2018). Porém, esse sistema, como dito anteriormente, apenas dá um suporte ao professor da disciplina automatizando vários processos, como correção manual de exercícios, atribuição de notas, entre outros, mas não foca em motivar ou incentivar o aluno a resolver as questões propostas ou a participar ativamente da disciplina.

2.3 Juízes online gamificados

Sinly, Rusli, e Winarno (2018) desenvolveram um ambiente gamificado para um sistema juiz online chamado UMN OJ, usado para treinamento de um clube de alunos da Universitas

Multimedia Nusantara² interessados em participar de competições de programação. Os autores aplicaram o teste de Bartle (Tondello, Mora, & Nacke, 2017) para identificar o perfil de jogador dos 24 alunos do clube de programação e, a partir do resultado desse teste, foram selecionados os elementos de jogos que seriam implementados na nova gamificação. Os autores avaliaram o sistema de duas formas distintas: i) participação dos alunos do clube de programação no sistema juiz online: antes da gamificação, apenas 37% dos alunos do clube tinham solucionado pelo menos um problema de programação no sistema juiz online UMN OJ. Esse número aumentou para 95% após a implementação do ambiente gamificado. Os autores acreditam que a gamificação teve um papel importante nesse aumento, mas pretendem aumentar o número de usuários para validar os experimentos; e ii) teste de aceitação do usuário: os autores usaram o modelo *Hedonic Motivation System Adoption Model* (HMSAM) (Lowry, Gaskin, Twyman, Hammer, & Roberts, 2012) para avaliar o índice de aceitação do sistema gamificado, e concluíram que a maioria dos entrevistados aprovou a nova gamificação. Embora os sistemas gamificados dos juizes online UMN OJ e CodeBench possuam objetivos similares, o trabalho de Sinly et al. (2018) adota apenas elementos básicos de gamificação, como tabela de classificação, *badges* e perfil. Além disso, o público-alvo do juiz online gamificado UMN OJ é formado por alunos interessados em participar de competições de programação, enquanto o CodeBench pode ser usado como ferramenta de apoio para qualquer disciplina de programação, tendo uma arquitetura ajustável à necessidade do professor da disciplina.

Seguindo um objetivo um pouco diferente, Kasahara, Sakamoto, Washizaki, e Fukazawa (2019) propuseram uma estratégia de gamificação para incentivar os alunos a desenvolverem códigos de alta qualidade em um juiz online chamado Waseda (WOJ). Usando a métrica *cyclomatic complexity* (CC) (McCabe, 1976), os autores implementaram uma tabela de pontuação (*ranking*) que ordena os alunos de acordo com a qualidade dos códigos desenvolvidos durante suas atividades no juiz online. O *ranking* é mostrado para cada problema de programação, e apresenta as classificações dos alunos, nomes pessoais, pontuações de CC e tempo de envio. Para avaliar a eficácia da ferramenta, foi aplicado um experimento em um curso de Programação C com 35 alunos e 6 semanas de duração. Como resultado, os avaliadores destacaram que os alunos recorreram a várias técnicas de qualidade de software para melhorar suas pontuações, como por exemplo, a criação de funções para evitar duplicação de código. Ao contrário do trabalho apresentado em Kasahara et al. (2019), o juiz online CodeBench utiliza outros elementos de gamificação além do *ranking* de alunos, como: animações, avatares, bens virtuais, *badges*, recompensas etc. Além disso, o CodeBench avalia os alunos de acordo com a correteza dos códigos submetidos, e não usa métricas de código para avaliar a qualidade das soluções submetidas.

Petit et al. (2018), apresentam o juiz online denominado Jutge, cujo objetivo é apoiar estudantes e instrutores de programação, oferecendo problemas organizados por curso, conteúdo e dificuldade. Os autores registram que o juiz online costuma ser utilizado na disciplina Estrutura de Dados para gerenciar uma atividade denominada “O Torneio de EAD”, em que alunos competem entre si projetando e implementando estratégias para controlar os movimentos de vários personagens de um jogo de acordo com um conjunto de regras pré-estabelecido. Cada aluno tem três semanas para definir individualmente sua estratégia e criar um jogador que sempre vença o “dummy” (jogador programado pelos professores). O pior jogador de cada rodada é eliminado, e quanto mais um aluno permanecer no jogo, maior será sua nota. Ao final é realizada uma cerimônia onde os melhores programadores são convidados a descrever suas estratégias. Essa estratégia é utilizada há 10 semestres e sempre conta com a participação de mais de 90% das turmas. Segundo os autores, essa estratégia de gamificação melhorou em 20% a taxa de aprovação na disciplina. Diferente do CodeBench, que prioriza a cooperação e utiliza

² <http://www.umn.ac.id/>

um robusto conjunto de elementos de gamificação, esse trabalho modelou uma competição utilizando o conteúdo da disciplina incentivando muito mais a competitividade e utilizando de tabela de *ranking* para apresentar o resultado. Outra diferença entre as duas propostas é que os elementos de jogos do CodeBench não requerem mudanças no conteúdo e nas atividades da disciplina, enquanto os elementos de jogos descritos por Petit et al. (2018) requerem alterações significativas na dinâmica que será aplicada nas turmas.

Bez, Tonin, e Rodegheri (2014) apresentam o juiz online URI, que desde sua primeira versão em 2012 vem recebendo vários novos recursos, incluindo gamificação. Segundo Klock et al. (2014), no URI Online Judge existem quatro elementos de gamificação: os rankings, as medalhas, as missões e a personalização de avatar. As missões permitem que o aluno conquiste diversas medalhas. O Juiz online possui o ranking geral e o ranking por universidade, classificando os alunos de acordo com a quantidade de problemas resolvidos. O CodeBench, além dos elementos de gamificação presentes no URI, apresenta uma narrativa medieval na qual os estudantes, representados por personagens, têm a missão de derrotar um monstro para libertar o seu país.

Wasik, Antczak, Badura, Laskowski, e Sternal (2018) descrevem o estado da arte nos sistemas de juízes online. Eles classificam os juízes online de acordo com seus objetivos: i) sistemas que suportam a organização de concursos competitivos de programação, ii) que aprimoram os processos de educação e iii) recrutamento, iv) que facilitam a solução de desafios de mineração de dados, v) que disponibilizam compiladores online e vi) plataformas de desenvolvimento integradas a componentes de outros sistemas. Especificamente nos juízes online voltados para a educação, eles registram que muitos introduzem elementos de gamificação para fortalecer o engajamento e a motivação dos alunos. Como exemplo, os autores citam o CheckiO, que é uma plataforma preparada para dar suporte ao processo de aprendizagem das linguagens de programação Python e JavaScript e fornece alguns elementos de gamificação e redes sociais. Segundo os autores, o CheckiO também é complementado com um jogo de estratégia online multiplayer chamado *Empire of Code*, em que os jogadores podem resolver problemas da plataforma CheckiO para ganhar bônus no jogo. Para ter sucesso, eles também precisam desenvolver *bots* artificiais, responsáveis pela proteção de suas unidades e por atacar seus inimigos em potencial. Outra plataforma que os autores mencionam é o Codewars, que possui exercícios ordenados por nível de dificuldade e assunto, com o diferencial de possuir interface do usuário inspirada na cultura japonesa. Seguindo nessa linha, os autores citam iniciativas que utilizam jogos para desenvolver habilidades algorítmicas, de programação e de Inteligência Artificial, como o CodeHunt, Leek Wars, CodinGame, entre outros. O CodeBench segue uma das linhas descritas pelos autores ao apresentar um juiz online gamificado para apoiar os professores em disciplinas de programação.

3 O Juiz Online Gamificado

O juiz online CodeBench foi gamificado utilizando-se a metodologia de *design* instrumental ADDIE (*Analysis, Design, Development, Implementation and Evaluation*) (Branch, 2009), em conjunto com o modelo ISD (*Instructional Systematic Design*) de Dick e Carey (2001). Essa abordagem unificada do matriciamento entre as principais etapas desses dois modelos foi proposta por Wangenheim e Wangenheim (2012). Por meio dela, nas etapas iniciais de desenvolvimento, foi concebida a mecânica de gamificação e foram definidas a meta do jogo, suas regras e a forma de interação do sistema, como mostrado na Tabela 1.

Tabela 1: Concepção do juiz *online* gamificado.

Concepção da	Juiz <i>online</i> gamificado
--------------	-------------------------------

gamificação	
Meta	Motivar o aluno a resolver exercícios de programação disponibilizados no juiz <i>online</i>
Tipo de mídia	Digital e <i>online</i>
Contexto	Utilizado ao longo da disciplina
Interação	Individual
Narrativa	A narrativa se passa em um mundo de fantasia medieval onde, ao final da jornada, os personagens (estudantes de uma mesma turma) devem enfrentar um monstro (<i>Quimera</i>) e libertar o país <i>Midgard</i> de sua dominação
Descrição	O estudante escolhe seu avatar entre várias opções disponíveis. À medida que resolve exercícios, anda pelo mapa até <i>Quimera</i> , ganhando pontos de força e armas. Quanto maior a força e melhor a arma do avatar, mais pontos de vida consegue tirar do monstro ao enfrentá-lo
Resultados	Uma porcentagem da turma deve chegar até o final do mapa, onde está <i>Quimera</i> , para matá-la. Esse percentual é definido pelo instrutor no momento de criação da turma e pode ser alterado a qualquer momento, durante o período letivo. Apesar da interação ser individual, esse objetivo coletivo foi inserido para que a dinâmica da gamificação não se tornasse excessivamente competitiva. A morte da <i>Quimera</i> define o “estado vencedor” do jogo
Feedback	Durante a utilização do juiz <i>online</i> , os alunos recebem <i>feedback</i> do seu desempenho, de acordo com a posição, armas e a força do seu personagem

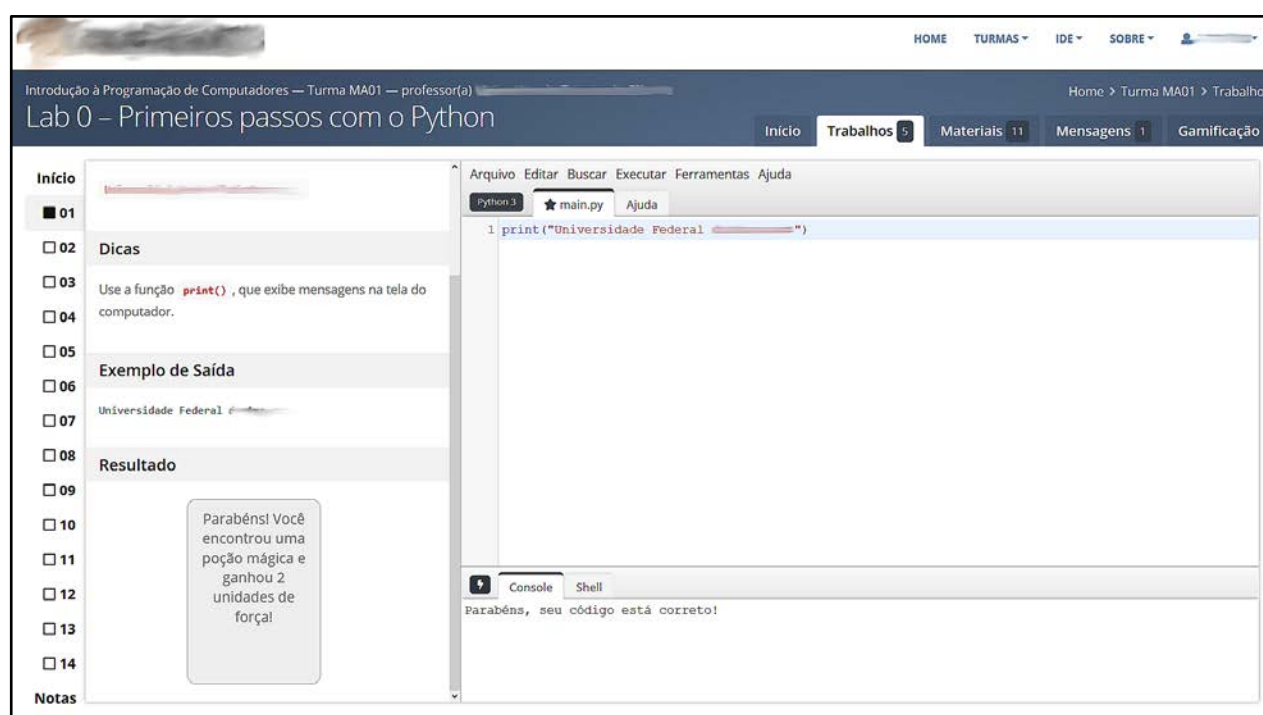


Figura 3: Captura de tela de um exercício submetido com sucesso ao juiz online.

O sistema de pontuação adota este processo: quando o código submetido pelo estudante é verificado como correto, sorteia-se uma carta de baralho (Figura 3), que representa o elemento ‘sorte’ e pode fornecer um dos seguintes itens, configuráveis pelo professor: **força** (1 a n pontos de força) e **posição no mapa** (1 a m casas). Caso o estudante obtenha nota máxima em uma lista de exercícios, seu avatar ganha uma medalha (*badge*), representada por uma arma mais potente. Quanto melhor a arma usada pelo avatar, mais pontos de vida ele consegue tirar da *Quimera*, cuja morte define o estado vencedor do jogo.

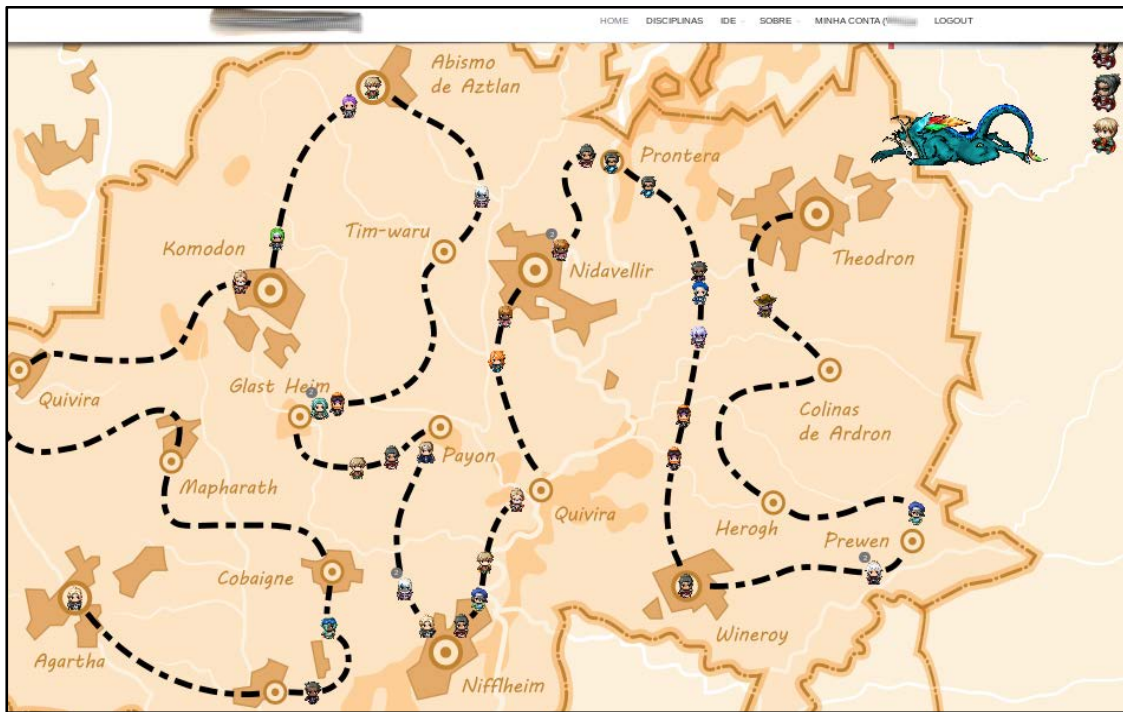


Figura 4: Captura de tela do mapa com a posição dos avatares.

HOME DISCIPLINAS IDE SOBRE MINHA CONTA LOGOUT

Participantes
 A listagem abaixo apresenta todos os guerreiros e guerreiras que ajudarão a vencer o monstro Quimera, ordenados de acordo com suas posições no tabuleiro do jogo. Cada guerreiro(a) possui uma quantidade de força e uma arma, que indicam a força dos golpes deste(a) guerreiro(a) sobre o monstro. Os três primeiros guerreiros ou guerreiras estão sempre de óculos escuros, simbolizando a liderança desses(as) guerreiros(as) sobre os demais.

	Força Acumulada: 173 / 200	Posição: 196
Arma:		
	Força Acumulada: 161 / 200	Posição: 196
Arma:		
	Força Acumulada: 178 / 200	Posição: 196
Arma:		
	Força Acumulada: 200 / 200	Posição: 196
Arma:		

Figura 5: Ranking dos avatares no juiz online gamificado.

A Figura 4 mostra a captura de tela do mapa do mundo imaginário e a Figura 5, o ranking dos jogadores. A mecânica e os componentes também utilizam elementos como: desafios, competição, *badges*, *feedback*, recompensas, estado de vitória, combate, *ranking*, pontos e bens virtuais.

4 Projeto Experimental

Foi realizado um quase-experimento para mensurar o desempenho do estudante ao utilizar o CodeBench gamificado. Para tanto, ele foi aplicado na Universidade Federal do Amazonas, durante o primeiro semestre letivo de 2017, na disciplina Introdução à Programação de Computadores (IPC), em onze turmas da área de ciências exatas e engenharias.

As mesmas turmas ofertadas no período letivo anterior correspondente (2016/1) foram utilizadas como “grupo controle” para efeito de comparação, nas quais os alunos usaram o juiz online sem gamificação. Foram coletados dados de 460 alunos do grupo experimental (2017/1) e 493 do grupo controle (2016/1), totalizando 953 alunos, a partir de quatro fontes:

- **Dados de Suporte 1 (DS01)** – Uma planilha de dados, extraída a partir do sistema de registro acadêmico, contendo informações dos estudantes matriculados em IPC em todos os cursos que ofertaram a disciplina em 2016/1 e 2017/1.
- **Dados de Suporte 2 (DS02)** – Um log, extraído do juiz *online*, que contém as ações individuais dos alunos no sistema gamificado durante a disciplina.
- **Dados de Suporte 3 (DS03)** – Uma planilha de dados, disponível publicamente na página da instituição, contendo as notas no Exame Nacional do Ensino Médio (Enem) dos candidatos inscritos no Sistema de Seleção Unificado (SiSU).
- **Dados de Suporte 4 (DS04)** – Uma planilha de dados contendo as respostas dos alunos de 2017/1 ao questionário MEEGA adaptado.

Esses dados estão consolidados em repositório online³, de forma anonimizada, para consulta da comunidade, bem como replicação das análises aqui apresentadas.

Durante a fase de planejamento, foram formuladas 06 hipóteses nulas a partir das questões de pesquisa. Para a Q1 (*Há influência da gamificação na nota do estudante em IPC?*), foi formulada a seguinte hipótese nula:

H01: Não há diferença na nota final em IPC entre a turma de 2016/1 e a de 2017/1.

Essa hipótese foi testada em três análises: turmas sem estratificação, turmas estratificadas por experiência prévia em programação (variável binária) e turmas estratificadas por alunos repetentes (variável binária). Para responder a Q1 de forma apropriada, além de comparar o desempenho em períodos letivos distintos, comparou-se o desempenho de IPC em 2017/1 com a nota dos mesmos estudantes em outras disciplinas comuns às matrizes curriculares. Além de IPC, as duas únicas comuns aos 11 cursos são Álgebra Linear 1 e Cálculo 1. Também se considerou o coeficiente de rendimento (CR), que corresponde à média aritmética de todas as disciplinas cursadas pelo aluno, desde seu ingresso até o período considerado, mas excluindo a nota de IPC no período considerado. Para essas três notas, foi formulada a seguinte hipótese nula:

H02: Não há diferença na nota entre a turma de 2016/1 e a de 2017/1.

Por fim, para complementar o entendimento da Q1, consideraram-se os dados de uso do juiz online (DS03) pelos estudantes, tais como número de logins, tempo de IDE, número de submissões, entre outros. Para essas medições, foi formulada a seguinte hipótese nula:

H03: Não há diferença de medição entre a turma de 2016/1 e a de 2017/1.

Com relação à Q2 (*Há influência dos elementos de gamificação sobre a nota final em IPC?*), foram formuladas as seguintes hipóteses nulas:

H04: Não há correlação entre os pontos de força e a nota final em IPC.

³ <https://github.com/leandrogalvao2/rbie2019>

H05: Não há correlação entre a posição final no mapa e a nota final em IPC.

Com respeito à Q3 (*Há associação entre a nota no Enem e a nota final do estudante em IPC?*), é necessário entender primeiro o conceito de “valor agregado”, utilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) para avaliação de cursos de graduação. Ele representa a contribuição do curso ao desenvolvimento, nos estudantes, de competências, habilidades e conhecimento (INEP, 2017). No presente estudo, consideramos o valor agregado apenas de algumas disciplinas, em vez de todo o curso de graduação.

Em termos quantitativos, o valor agregado pode ser entendido como o poder de estimar o desempenho de um graduando a partir das notas de entrada (notas do Enem). Se essas notas forem um bom estimador, então o curso (neste trabalho, uma disciplina de interesse) agregou pouco valor, pois os estudantes mantiveram seu perfil de desempenho demonstrado na entrada. Caso contrário, isso significa que o curso proporcionou oportunidades de desenvolvimento para que os alunos de menor desempenho pudessem se desenvolver, de modo que a conclusão do curso dependeu em menor grau das condições de entrada.

Dessa forma, seguindo a metodologia do INEP (2017), foi considerado um modelo de regressão múltipla, cuja variável de resposta era, a seu turno, cada uma das notas nas 03 disciplinas comuns (IPC, Álgebra Linear 1 e Cálculo 1) e o CR. Como variáveis explanatórias, foram usadas as notas nas seguintes provas do Enem: Linguagens e Códigos, Ciências Humanas, Ciências da Natureza e Matemática. Para esse modelo, foi testada a seguinte hipótese nula:

H06: O coeficiente de regressão de cada variável explanatória é igual a zero (sem efeito).

Por fim, com respeito às questões de pesquisa Q4 (*Os estudantes se sentem motivados ao utilizarem o juiz online gamificado?*), Q5 (*Os estudantes se sentem confortáveis ao utilizarem o juiz online gamificado?*) e Q6 (*Os estudantes percebem algum ganho de aprendizagem ao utilizarem o juiz online gamificado?*), não foram formuladas hipóteses nulas, pois elas nortearam uma avaliação qualitativa sobre a percepção dos estudantes ao usarem o juiz online gamificado.

Para nortear a avaliação qualitativa, ao fim de 2017/1, os estudantes foram convidados a responder um questionário, adaptado a partir do instrumento MEEGA (*Model for the Evaluation of Educational Games*), desenvolvido para avaliar jogos educacionais (Savi, Wangenheim, & Borgatto, 2011). Os itens foram modificados para o contexto de um sistema gamificado, em vez de jogos. O MEEGA é decomposto em três subcomponentes: **motivação**, **experiência de usuário** e **aprendizagem**. Por sua vez, cada subcomponente é decomposto em um número variável de dimensões. Embora hoje exista um aprimoramento desse modelo, o MEEGA+ (Petri, Wangenheim, & Borgatto, 2018), este ainda não havia sido publicado antes da avaliação, em 2017.

O questionário MEEGA adaptado contém um total de 32 itens, divididos em 11 dimensões de avaliação. As respostas aos itens foram baseadas em uma escala Likert, variando de “discordo fortemente” (-2) até “concordo fortemente” (+2). Além disso, os estudantes também escreveram livremente comentários sobre cada item e citaram pontos fortes e fracos sobre o juiz online gamificado. Esses comentários foram analisados por meio do Método de Explicitação do Discurso Subjacente (MEDS) (Nicolaci-da-Costa, 2007).

4.1 Descrição da disciplina

Na Universidade Federal do Amazonas, a disciplina de IPC corresponde ao que na literatura costuma-se chamar CS1. Seu conteúdo abrange *variáveis*, *comandos de entrada e saída*, *comandos condicionais*, *laços de repetição*, *vetores* e *matrizes*. Ela é ministrada para alunos de

graduação em cursos fora da área de Computação, conhecidos como *non-majors*. A linguagem Python é adotada como base para aplicação dos conceitos.

A carga horária da disciplina é de 60 horas, distribuídas ao longo de um período letivo de cerca de 4 meses de duração. As aulas são ministradas em dois encontros semanais de 2h de duração. O conteúdo foi dividido em 07 módulos, cada um contendo uma avaliação formativa (exercícios práticos) e uma somativa parcial. Todas as atividades (exercícios, avaliações, divulgação do material de aula) são apoiadas pelo juiz online. Cada turma conta com o apoio de um tutor, um aluno de pós-graduação, matriculado em disciplina de estágio em docência.

Os **exercícios práticos** podem ser realizados em qualquer lugar com acesso à internet, com peso de cerca de 1,2% sobre a nota final. Os alunos podem consultar qualquer pessoa ou fonte para resolver os exercícios. As **avaliações parciais** só podem ser realizadas no laboratório de computadores, sob a supervisão do professor e do tutor, sem possibilidade de consulta. O juiz online utiliza o endereço físico da interface de rede como condição de acesso às avaliações parciais. Estas têm um peso variável entre 4,2% e 12,5% sobre a nota final, conforme a dificuldade do assunto. Pelas regras da instituição de ensino, a nota mínima de aprovação é 5,0, em uma escala de 0 a 10. Para mais detalhes, consultar (Carvalho, Gadelha, Nakamura & Oliveira, 2016).

As 11 turmas de 2016/1 foram alocadas para 07 diferentes professores. As 11 de 2017/1 foram alocadas para 08 professores distintos, entre os quais 04 foram comuns a ambos os períodos letivos. Entre os outros 04 docentes, 02 já haviam ministrado a disciplina em períodos anteriores, com a mesma metodologia, e os outros 02 eram novos na metodologia, embora tivessem experiência no ensino de programação introdutória. Todos os docentes, em ambos os períodos, eram oriundos do quadro efetivo da instituição, em regime de dedicação exclusiva. Os tutores foram todos renovados de um ano para o outro.

Um questionário demográfico é aplicado no momento do cadastro dos alunos no juiz online. Entre seus vários itens, pergunta-se a **experiência prévia** em alguma linguagem de programação. Em 2016/1, o preenchimento do questionário era opcional, mas passou a ser obrigatório a partir de 2017/1, devido à expressiva quantidade de não-respostas (23,5%).

A mesma metodologia de ensino utilizada em 2016/1 também foi aplicada em 2017/1. A diferença essencial foi a gamificação do juiz online e a disponibilização de 12 exercícios práticos a mais em 2017/1 (total de 87) em relação a 2016/1 (total de 75). As questões das avaliações parciais de 2017/1 foram diferentes daquelas aplicadas em 2016/1. Cada questão era sorteada entre pelo menos 10 variações, a fim de minimizar problemas de desonestidade acadêmica.

4.2 Caracterização das amostras

Como dito anteriormente, a versão gamificada do juiz online foi aplicada em 11 turmas ofertadas para cursos de graduação da área de ciências exatas e engenharias na Universidade Federal do Amazonas, durante o primeiro semestre letivo de 2017 (2017/1). Os dados das turmas de 2017/1 foram comparados com os dados gerados pelos alunos das turmas ofertadas em 2016/1 para os mesmos cursos, mas com o apoio da versão original do juiz online, sem gamificação.

Em 2016/1, houve um total de 493 alunos matriculados em IPC; já em 2017/1, houve 460. A interseção entre estes dois grupos foi de 51 alunos (11,1% dos alunos matriculados em 2017/1). Nas análises estatísticas, foram considerados somente os alunos **não desistentes** em cada período letivo, pois os desistentes deixam de gerar dados necessários para algumas delas. Assim, havia 351 alunos não desistentes (71,2% dos matriculados) em 2016/1 e 294 alunos não

desistentes (63,9% dos matriculados) em 2017/1. A interseção entre estes dois grupos foi de apenas 9 alunos (3,1% dos alunos não desistentes em 2017/1).

A Figura 6 apresenta a distribuição de frequência das principais características demográficas dos alunos não desistentes em 2016/1 (controle) e em 2017/1 (experimental). À direita de cada par de gráficos, é indicado o valor- p do teste qui-quadrado (χ^2) que avaliou a semelhança das distribuições de frequência da característica evidenciada. Em todos os casos, não foi possível rejeitar a hipótese nula de que não há diferença entre as amostras. Ou seja, as características demográficas das turmas de 2016/1 e 2017/1 são muito semelhantes e são explicitadas a seguir:

- **Sexo:** indica o sexo biológico do estudante (feminino ou masculino).
- **Idade:** corresponde ao número de anos transcorridos entre a data de nascimento do estudante e o dia 1º de março do ano considerado. Essa característica foi discretizada em 06 faixas mais frequentes.
- **Tempo de curso:** quantidade de anos transcorridos entre o ano de entrada do estudante e o ano considerado. Foi estratificada em três faixas: calouros (zero ano), 01 ano de curso e mais que 01 ano de curso.
- **Primeira vez:** indica se o estudante está cursando IPC pela primeira vez ou não (repetente).
- **Experiência prévia:** indica se o estudante reportou ter conhecimento ou não de alguma linguagem de programação antes de iniciar a disciplina.
- **Processo Seletivo:** indica se o estudante ingressou em seu curso pelo SiSU (50% das vagas), ou pelo processo próprio da instituição (50% das vagas), ou por outro meio (remanejamento, mandado judicial, transferência ex-offício).
- **Cota:** indica por qual política afirmativa o estudante ingressou em seu curso. Foram consideradas 03 categorias: **ampla concorrência** (AC); cota **dependente de renda** (renda familiar bruta *per capita* igual ou inferior a 1,5 salário mínimo e que tenha cursado integralmente o ensino médio em escola pública, tendo ou não declarado cor da pele); cota **independente de renda** (sem declaração de renda e que tenha cursado integralmente o ensino médio em escola pública, tendo ou não declarado cor da pele).

Por meio da Figura 6, nota-se a semelhança na distribuição das categorias de cada característica demográfica entre 2016/1 e 2017/1. A única diferença notável refere-se à experiência prévia em programação. Entre os alunos de 2016/1, 35% declararam conhecer alguma linguagem de programação, contra 28% em 2017/1. No entanto, o teste do qui-quadrado não encontrou diferença significativa entre as distribuições. Como nenhum teste de conhecimento prévio foi aplicado, usamos esse indicador como referência na análise que se segue.

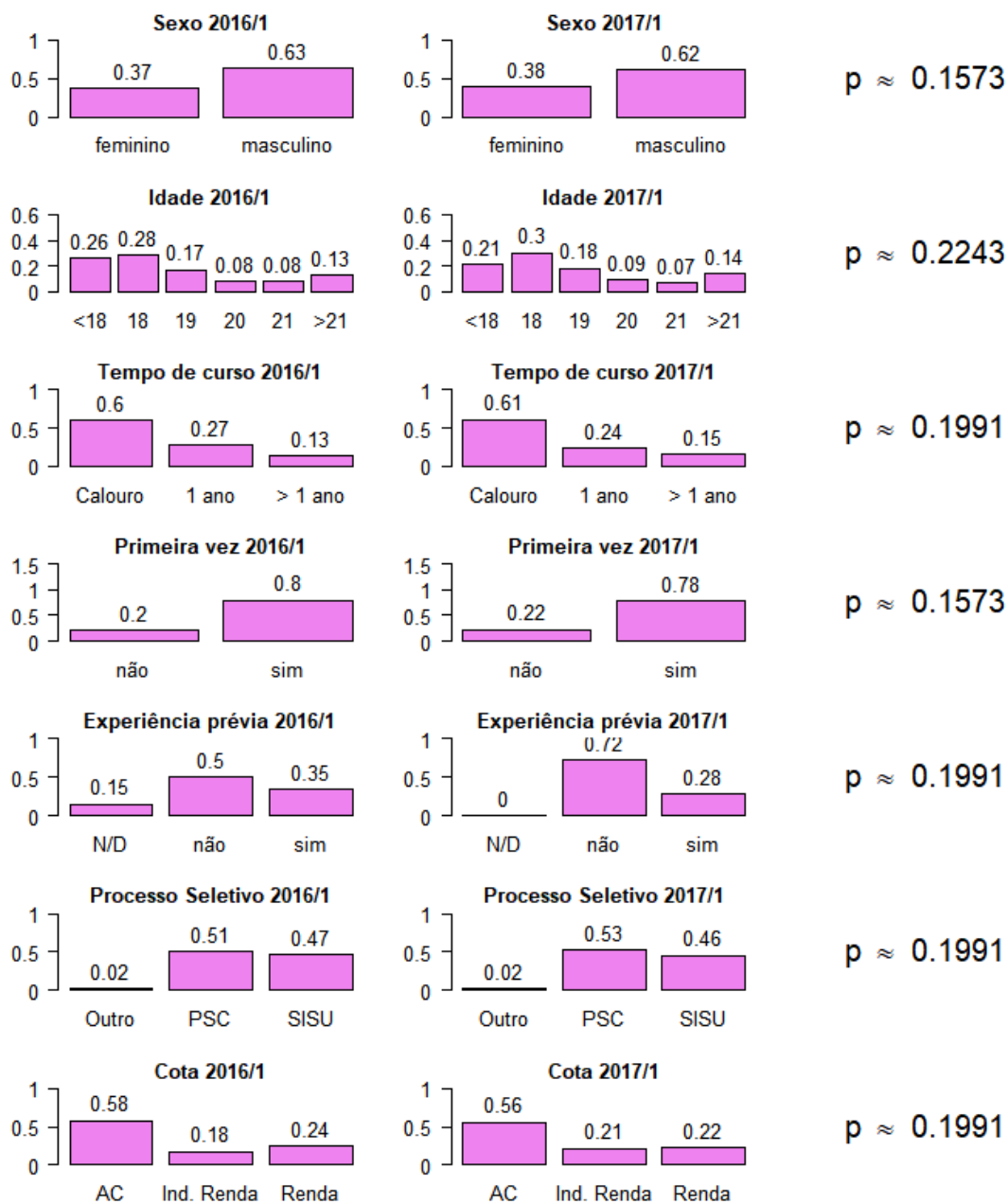


Figura 6: Distribuição de frequência das características demográficas dos alunos não desistentes nas turmas de IPC ministradas em 2016/1 (sem gamificação) e em 2017/1 (com gamificação).

5 Resultados e Discussão

Nesta seção são apresentados os resultados obtidos a partir de testes estatísticos realizados sobre as hipóteses levantadas. Mais precisamente são analisados: i) o desempenho em IPC, ii) a experiência prévia em programação; iii) o desempenho entre os alunos repetentes; iv) o desempenho acadêmico no Enem e em outras disciplinas; v) o uso do juiz online; vi) os elementos de Gamificação × Desempenho e vii) o valor agregado em IPC. Os testes estatísticos consideraram um nível de confiança de 95%. Além disso, esta seção apresenta a percepção do estudante através das dimensões motivação, experiência do usuário e aprendizagem, adaptadas do questionário MEEGA para o contexto gamificado, uma análise qualitativa que teve como fomento questões abertas para que o aluno indicasse os pontos positivos e negativos e possíveis ameaças à validade dos resultados. A discussão dos resultados acompanha a respectiva apresentação.

5.1 Desempenho em IPC

Primeiramente, testou-se a H_0 , ou seja, se as notas finais em IPC dos estudantes não desistentes em 2017/1 eram significativamente diferentes das obtidas pelos não desistentes em 2016/1. Uma vez verificado que nenhuma das amostras seguia a distribuição normal, aplicou-se o teste não-paramétrico de *Mann-Whitney* (amostras não pareadas), cujo valor p foi $6,61e-09$, rejeitando-se assim a H_0 . Isto é, podemos afirmar que há diferença nas notas finais em IPC dos estudantes não desistentes, entre as turmas do grupo experimental (2017/1) e controle (2016/1).

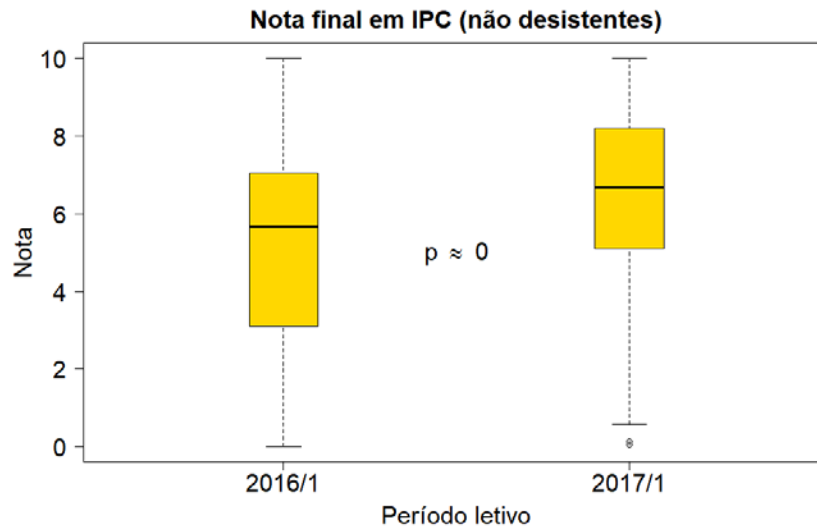


Figura 7: Comparação da distribuição de notas entre os estudantes não desistentes das turmas de 2016/1 e 2017/1.

Na Figura 7, são exibidos os *boxplots* das notas finais em IPC. Na caixa à esquerda, tem-se distribuição das notas de 351 alunos não desistentes entre os 493 matriculados em 2016/1 e, na caixa à direita, tem-se a distribuição das notas de 294 não desistentes entre os 460 matriculados de 2017/1.

A Figura 7 também indica uma melhora da mediana das notas, de 5,67 em 2016/1 para 6,67 em 2017/1. Outra melhoria evidente ocorreu na variância das notas, indicando maior homogeneidade no desempenho na disciplina. Além disso, a taxa de aprovação na disciplina subiu de 66,1% em 2016/1 para 79,6% em 2017/1. Esses resultados são indícios quantitativos de que o uso do juiz *online* gamificado contribuiu para a melhoria do desempenho dos estudantes não desistentes, isto é, aqueles que não trancaram a disciplina ou não reprovaram por falta.

Nas seções a seguir, avalia-se se essa melhoria possa ter dependido ou não dos seguintes fatores: ter experiência prévia em programação (Seção 5.2), estar cursando a disciplina pela primeira vez (Seção 5.3), haver essa mesma tendência de melhoria em disciplinas comuns do currículo e no coeficiente de rendimento (Seção 5.4).

5.2 Experiência prévia em programação

Embora o resultado anterior seja favorável, o desempenho acadêmico de uma turma em uma disciplina pode ser explicado por uma diversidade de fatores, como por exemplo: conhecimento prévio dos alunos, pedagogia e empatia dos docentes que ministram a disciplina, metodologia de ensino-aprendizagem empregada, ferramentas utilizadas, alteração curricular, entre outros. Nesta subseção, demonstramos que a experiência prévia em programação não foi um fator que impactou o desempenho acadêmico dos estudantes em IPC.

Nesta análise, os estudantes não desistentes em 2016 e em 2017 foram divididos em dois estratos: aqueles que reportaram conhecer alguma linguagem de programação antes do início da disciplina e aqueles que reportaram não conhecer. O primeiro grupo, que aqui chamamos de

“experientes”, correspondia a 34,8% dos alunos não desistentes em 2016/1 e a 27,9% dos de 2017/1. Os não-experientes correspondiam a 50,4% em 2016/1 e a 71,8% em 2017/1. Em 2016/1, quando o questionário demográfico era de preenchimento opcional, 14,8% dos alunos não desistentes não responderam. Em 2017/1, um único aluno (0,3%) não respondeu. Os alunos que não responderam esse item do questionário foram descartados desta análise.

Assim, a hipótese nula H_0 foi novamente testada, estratificando-se cada amostra (2016/1 e 2017/1) em dois níveis (com e sem experiência prévia). Aplicando-se o teste não paramétrico de Mann-Whitney, essa hipótese foi rejeitada em ambos os estratos. Dessa forma, observou-se que a diferença de desempenho entre os tratamentos (com e sem gamificação do juiz online) não pode ser atribuída à aleatoriedade das amostras, tanto para os estudantes com experiência prévia ($p \approx 0,0004$), quanto para os sem experiência prévia ($p \approx 0,0018$).

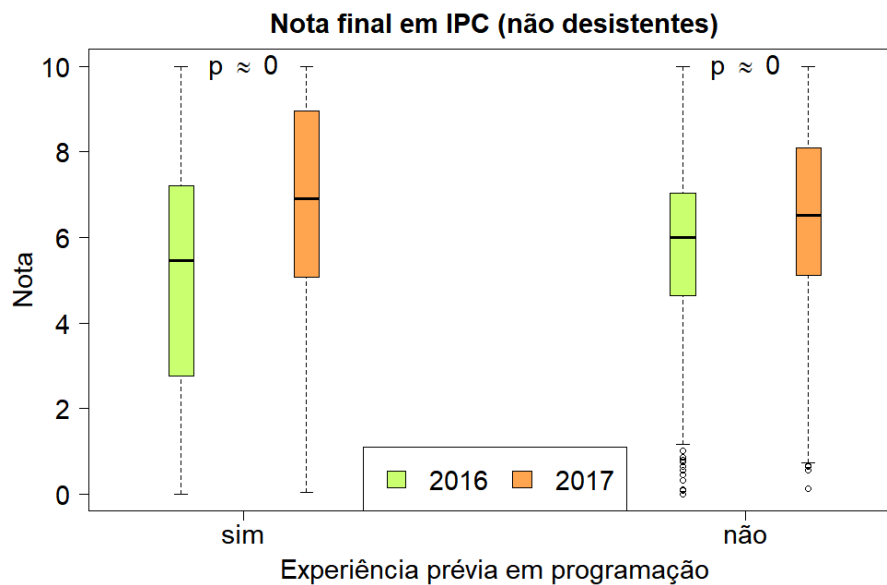


Figura 8: Comparação do desempenho em IPC (nota final) entre as turmas de 2016/1 e 2017/1, estratificadas em grupos de alunos que autorrelataram ter ou não ter experiência em programação previamente ao início da disciplina.

A Figura 8 ilustra a distribuição das notas finais para esses 04 grupos. É interessante notar como a variância da nota final dos alunos sem experiência prévia é menor que a dos alunos autorrelatados experientes. Isso se explica pelo fato de haver uma única maneira de não haver experiência: nunca ter tido contato com alguma linguagem de programação. Por outro lado, o contato prévio com programação pode ter acontecido de diversas maneiras, paradigmas e graus de aprofundamento, de modo que o grupo de alunos experientes é mais heterogêneo do que o de não experientes, o que se reflete na variância das notas finais.

5.3 Desempenho entre os alunos repetentes

Procurou-se investigar também se o desempenho entre alunos repetentes era diferente daquele entre os não repetentes. Por “repetente”, entende-se o aluno que não está cursando pela primeira vez a disciplina de IPC. Isso não quer dizer que esse estudante se autoconsidere “experiente em programação”, como na análise anterior. Por exemplo, o estudante pode ter desistido da disciplina no início de um período anterior, sem ter visto a maior parte do conteúdo.

Em 2016/1, entre os 351 alunos não desistentes em IPC, 59 (16,8%) eram repetentes e os demais (83,2%) eram não repetentes, ou seja, estavam cursando a disciplina pela primeira vez. Essa informação foi extraída do histórico de cada aluno (fonte de dados DS01). Já em 2017/1, entre os 294 alunos não desistentes em IPC, 46 (15,6%) eram repetentes e os demais (84,4%) eram não repetentes. Portanto, a distribuição de alunos repetentes era semelhante em cada aplicação da disciplina, como comprovado pelo teste do qui-quadrado (ver Seção 4.1).

Assim, a hipótese nula H_{01} foi novamente testada, estratificando-se cada amostra em dois níveis (repetente e não repetente). Aplicando-se o teste de Mann-Whitney, verificou-se que, entre os estudantes não repetentes, o desempenho em 2017/1 foi significativamente maior que em 2016/1 ($p \approx 1,03e-08$). Porém, não se verificou diferença significativa entre os tratamentos para os repetentes ($p \approx 0,0863$). No entanto, o *boxplot* da Figura 9 mostra que houve um ligeiro aumento na mediana e, principalmente, uma redução da variância.

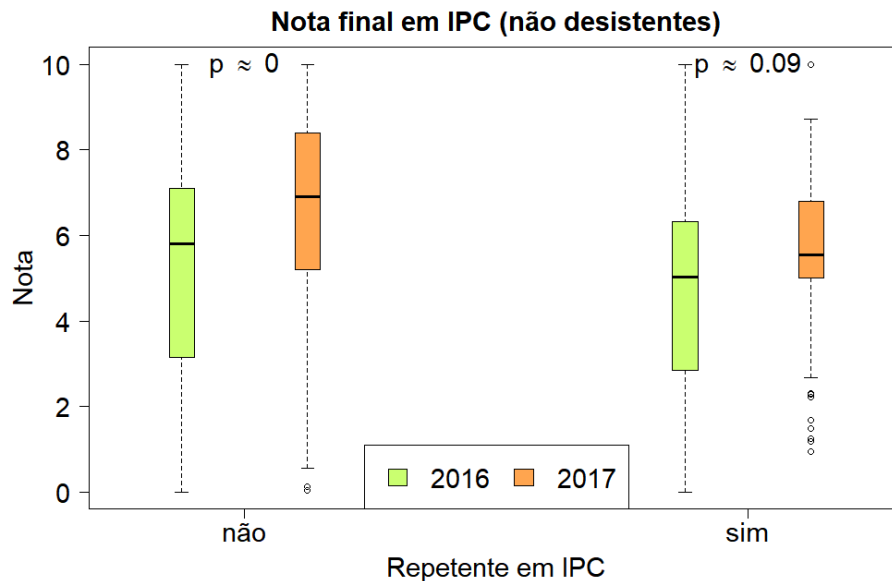


Figura 9: Comparação do desempenho em IPC (nota final) entre as turmas de 2016/1 e 2017/1, estratificadas em grupos de alunos repetentes e não repetentes.

Dessa forma, notam-se indícios de que a gamificação do juiz online, da maneira como foi conduzida, trouxe um impacto positivo para os alunos que tiveram o primeiro contato com a disciplina. Já entre os alunos repetentes, não se verificou esse impacto, porém os resultados foram mais homogêneos, principalmente entre os alunos de baixo rendimento (1º e 2º quartis).

5.4 Desempenho acadêmico no Enem e em outras disciplinas

A diferença de desempenho acadêmico em IPC entre as turmas de 2016/1 e de 2017/1 poderia ser explicada não apenas pela gamificação no juiz online, mas também pela sazonalidade do processo de seleção. Na prática docente, nota-se que algumas turmas costumam ter um melhor desempenho que anteriores, considerando uma mesma disciplina, ministrada por um mesmo professor, seguindo uma mesma metodologia. Assim, nesta subseção, investigamos se os alunos de 2017/1 tiveram um desempenho acadêmico geral melhor que os de 2016/1, o que poderia explicar o bom desempenho em IPC, em vez do uso do juiz online gamificado.

Para isso, foram consideradas quatro notas dos alunos não desistentes em IPC: nota média do Enem, coeficiente de rendimento (CR) e notas finais nas disciplinas Cálculo 1 e Álgebra Linear 1. A nota média do Enem corresponde à média aritmética das provas das cinco áreas de conhecimento avaliadas: Linguagens e Códigos, Ciências Humanas, Ciências da Natureza, Matemática e Redação.

Assim, a hipótese nula H_{02} foi testada para cada nota (média do Enem, Cálculo 1, Álgebra Linear 1 e CR). Aplicando-se o teste de Mann-Whitney, verificou-se não ser possível rejeitá-la em nenhum dos casos. Ou seja, não houve diferença de notas entre as turmas de 2016/1 e 2017/1 na média do Enem ($p \approx 0,0942$), em Cálculo 1 ($p \approx 0,8402$), em Álgebra Linear 1 ($p \approx 0,7533$) e no CR ($p \approx 0,1232$). A Figura 10 ilustra a distribuição dessas quatro notas.

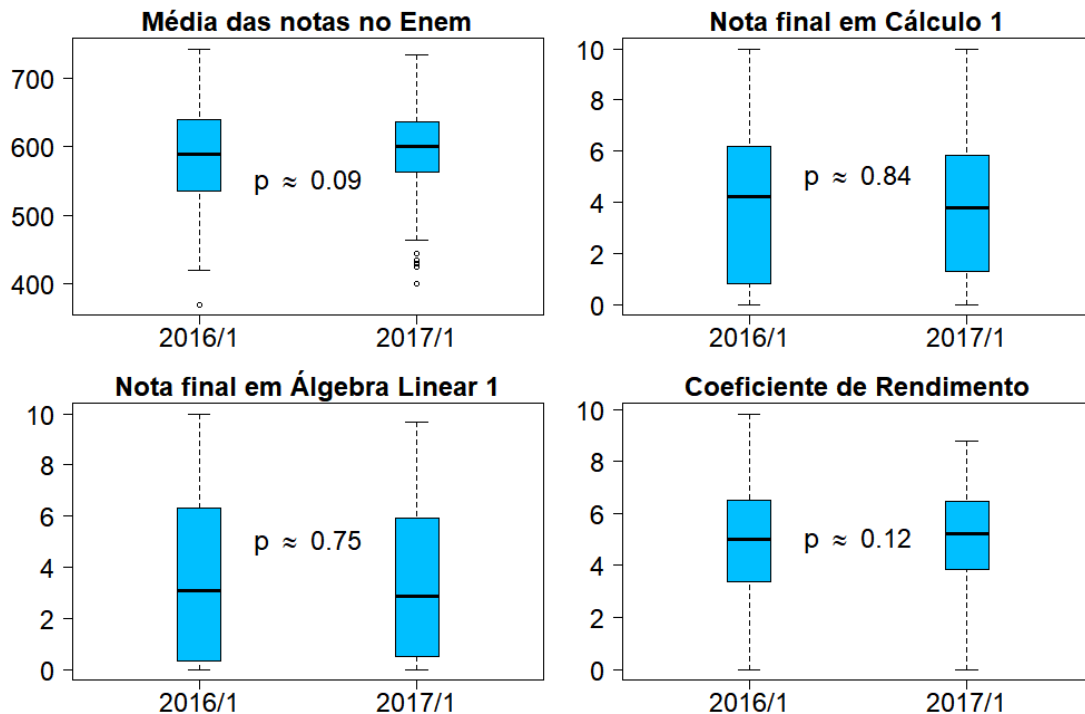


Figura 10: Comparação do desempenho em IPC (nota final) entre as turmas de 2016/1 e 2017/1, com respeito às notas do Enem, Cálculo 1, Álgebra Linear 1 e Coeficiente de Rendimento.

Para as notas de Cálculo 1, Álgebra Linear 1 e CR, havia informação disponível para todos os estudantes considerados. Já para a nota no Enem, obteve-se acesso a 77,8% dos estudantes não desistentes em 2016/1 e a 68,4% em 2017/1, pois a instituição adota dois processos seletivos: um próprio e o Sistema de Seleção Unificada (SiSU), que utiliza as notas do Enem. Além disso, também não se encontrou diferença significativa entre as duas turmas nas notas de cada uma das cinco provas do Enem: Linguagens e Códigos ($p \approx 0,1007$), Ciências Humanas ($p \approx 0,0973$), Ciências da Natureza ($p \approx 0,0794$), Matemática ($p \approx 0,0796$) e Redação ($p \approx 0,1470$).

Esse resultado indica que as turmas 2016/1 e 2017/1 de IPC eram compostas por estudantes com desempenhos semelhantes, tanto no processo seletivo, quanto nas disciplinas em geral de seus cursos de graduação. Portanto, é possível que a mudança da ferramenta empregada (gamificação do juiz online) tenha contribuído para a diferença de desempenho observado em IPC ($p = 6,61e-9$), uma vez que as turmas não têm diferença de desempenho acadêmico na seleção, nem nas duas disciplinas comuns, nem no coeficiente de rendimento em geral.

5.5 Uso do juiz online

Além da correção automatizada de códigos, o CodeBench oferece um ambiente de desenvolvimento (IDE) próprio. Nele, o estudante pode codificar e testar sua solução antes de submetê-la para correção, evitando ter que instalar e configurar um IDE, o que é particularmente útil para quem não possui um computador próprio. Além disso, o CodeBench permite coletar informações sobre os hábitos de uso e de estudo dos estudantes.

Dessa forma, averiguou-se se as turmas tiveram diferentes comportamentos de uso do juiz online após a sua gamificação. Os atributos observados foram os seguintes:

- **Número de logins:** quantidade de acessos ao juiz online.
- **Número de testes:** quantidade de testes de código realizados no IDE do juiz online. Considera apenas os testes feitos durante os exercícios práticos (avaliação formativa).

Ao desenvolver um código para um determinado exercício, o aluno pode executar seu código e testá-lo com suas próprias entradas. Ele pode fazer tantos testes quanto desejar, e esse procedimento é exatamente igual à execução de seu programa em um *shell* Linux.

- **Número de submissões:** quantidade de submissões feitas pelos alunos para correção do juiz online. Considera apenas os exercícios práticos (avaliação formativa). Ao concluir seu código, o aluno pode submetê-lo para avaliação de corretude, que verifica se o código-fonte atende aos casos de testes (entrada e saída) definidos pelo instrutor ao criar o exercício. Os alunos podem fazer tantas submissões quanto desejarem, sem penalidade.
- **Tempo de uso do IDE:** tempo (em segundos) de uso efetivo do IDE do juiz online, ao longo do período letivo considerado. Considera-se apenas o tempo em que o aluno digita algo no IDE. Não é computado o tempo que o aluno despense usando outras abas do navegador, ou usando outras janelas.
- **Número de exercícios corretos:** quantidade de exercícios resolvidos pelo aluno e que foram considerados como corretos pelo juiz online.
- **Número de exercícios incorretos:** quantidade de exercícios resolvidos pelo aluno e que foram considerados como incorretos pelo juiz online.

Para a presente análise, esses atributos foram coletados apenas para os exercícios práticos, pois, em uma avaliação, a tendência é que o aluno permaneça logado até o final do horário de prova. Além disso, esses dados foram normalizados pela quantidade de exercícios disponibilizada em cada ano (75 em 2016/1 e 87 em 2017/1).

Por meio do teste de normalidade de Shapiro-Wilk, verificou-se que nenhum dos seis atributos possuía uma distribuição normal, em nenhuma das turmas. Por isso, utilizou-se o teste de Mann-Whitney para verificar a hipótese nula H_0 , para cada atributo. Os resultados, mostrados na Figura 11, indicam que houve diferença significativa entre as turmas em quatro dos atributos. A mediana do número de logins por exercício prático aumentou de 1,2 por aluno em 2016/1 para 1,3 em 2017/1. Do mesmo modo, aumentaram as medianas do número de testes (de 3,6 para 6,4) e do tempo de uso da IDE (de 429s para 485s, cerca de um minuto a mais). Por outro lado, o número de submissões para correção por exercício e por aluno caiu de 3,7 em 2016/1 para 2,8 em 2017/1. Não se verificou diferença significativa entre as turmas na proporção de exercícios resolvidos corretamente e dos incorretos.

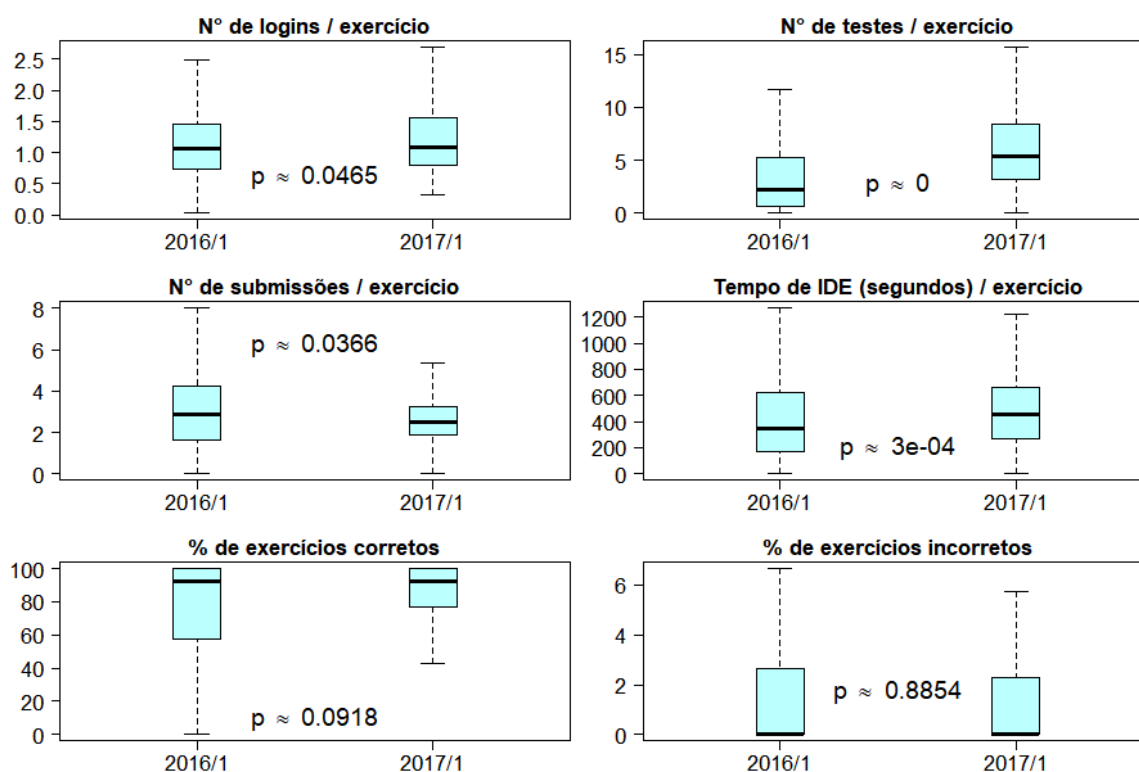


Figura 11: Comparação entre as turmas de 2016/1 e 2017/1 com respeito a atributos de uso do juiz online, durante a resolução de exercícios práticos (avaliação formativa).

Provavelmente, a gamificação do juiz online estimulou um maior número de acessos à plataforma, bem como um maior tempo de uso dela, na tentativa de acertar mais exercícios e, conseqüentemente, conquistar mais itens. Embora o percentual de exercícios corretos não tenha aumentado significativamente, o *boxplot* da Figura 11 mostra uma redução expressiva dos 1º e 2º quartis em 2017/1, isto é, houve um maior empenho dos alunos de menor rendimento.

5.6 Elementos de Gamificação × Desempenho

Ao final da disciplina, todas as 11 turmas conseguiram derrotar a *Quimera*. Além disso, dos 234 alunos aprovados, 80 (34,2%) conseguiram chegar à última posição do mapa. Mesmo depois de o monstro ter sido derrotado, os demais alunos que não alcançaram a *Quimera* ainda podiam resolver exercícios e avançar casas no mapa.

As hipóteses H04 e H05 foram verificadas por meio da correlação de Spearman. Os resultados na Tabela 2 indicam que houve uma correlação forte ($r_s > 0,7$) entre os pontos de força conquistados pelos estudantes no juiz online gamificado e seu desempenho na disciplina. O mesmo ocorreu na correlação entre a posição dos jogadores no mapa e seu desempenho em IPC.

Então, conclui-se que os elementos da gamificação contidos no juiz *online* têm forte associação com as notas dos estudantes. Ou seja, o elemento “sorte” inserido na mecânica da gamificação não distorce a equidade que se espera de que os estudantes com maior pontuação no jogo apresentem também maior desempenho em termos de notas.

Tabela 2: Coeficiente de correlação de Spearman (r_s) entre os pontos e a posição no mapa × notas dos estudantes em IPC.

	Pontos de Força × Notas	Posição no Mapa × Notas
valor p	2,2e-16	2,2e-16
Coeficiente de correlação (r_s)	0,74	0,74

5.7 Valor agregado em IPC

Como exposto na Seção 4, o valor agregado corresponde ao desenvolvimento de competências, habilidades e conhecimento nos estudantes, proporcionado pelo curso. O INEP calcula seu valor, ao avaliar cursos de graduação, com base em uma regressão linear múltipla a partir da nota das provas do Enem (INEP, 2017). Neste trabalho, adaptamos essa metodologia de modo que a variável dependente fosse, de cada vez, as notas das 03 disciplinas comuns e o CR, em vez da nota no Enade (Exame Nacional de Desempenho dos Estudantes). Assim, para cada variável dependente, e cada período letivo, foi testada a hipótese nula H_0 .

A Tabela 3 apresenta os valores p de cada um dos modelos de regressão testados, destacando-se os valores significativos ($p > 0,05$) em negrito. Não são apresentados os valores dos coeficientes de regressão porque estamos interessados essencialmente em saber se o efeito deles é ou não significativo sobre o modelo, ou seja, se alterações no valor das variáveis preditoras (notas do Enem) estão relacionadas a alterações em cada variável de resposta considerada (nota final das 03 disciplinas comuns e CR).

Na Tabela 3, observa-se que somente a nota da prova de Matemática no Enem foi uma variável significativa para explicar a maior parte das notas testadas, o que confirma o senso comum, por se tratar de cursos da área de Ciências Exatas. Em 2017/1, nota-se um efeito significativo da nota de Ciências Humanas sobre o CR e a nota de Álgebra Linear 1. Nas outras duas provas, os valores p foram pouco acima do nível de significância adotado.

O principal resultado é que em IPC, em 2017/1, nenhuma das notas do Enem tiveram efeito significativo sobre o modelo de regressão, diferentemente do que ocorreu em 2016/1, quando a prova de Matemática teve efeito significativo para explicar a nota em IPC. Essa evidência se reforça quando se observa que, nas outras duas disciplinas comuns e no CR em geral, a prova de Matemática do Enem manteve-se também em 2017/1 como fator de efeito significativo sobre o modelo de regressão.

Tabela 3: Valores p para hipótese nula de que o termo (nota em prova do Enem) não contribui para a saída do modelo de regressão (coeficiente de rendimento ou nota final em disciplina).

Área de conhecimento do Enem	IPC		Cálculo 1		Álgebra Linear 1		CR	
	2016/1	2017/1	2016/1	2017/1	2016/1	2017/1	2016/1	2017/1
Ciências Humanas	0,541	0,051	0,852	0,054	0,916	0,014	0,185	0,003
Ciências da Natureza	0,238	0,272	0,089	0,495	0,714	0,498	0,471	0,772
Linguagens e Códigos	0,525	0,734	0,766	0,029	0,449	0,976	0,743	0,852
Matemática	0,000	0,242	0,000	0,003	0,000	0,020	0,000	0,042

Assim, verifica-se que há indícios de que a gamificação do juiz online contribuiu para que o aprendizado de programação rompesse o mimetismo das condições de entrada. Isto é, a metodologia adotada em 2017/1 fez com que alunos de diferentes desempenhos no Enem tivessem iguais oportunidades de alcançar uma boa nota em IPC, o que não foi verificado nas duas disciplinas comuns ou no desempenho geral expresso pelo CR.

5.8 Questionário MEEGA adaptado

Para complementar a análise quantitativa do juiz online gamificado em termos de desempenho acadêmico, investigou-se a percepção do estudante em relação à gamificação. Assim, no final de 2017/1, foi aplicado o questionário MEEGA (Savi, Wangenheim, & Borgatto, 2011), que adaptamos para o contexto de um sistema gamificado de aprendizagem. Entre os 460 estudantes matriculados nas 11 turmas, 373 responderam o questionário. Os itens do questionário MEEGA são divididos em três componentes: motivação, experiência do usuário e aprendizagem, os quais correspondem às questões de pesquisa Q4, Q5 e Q6 formuladas na Seção 1.

5.8.1 Percepção sobre motivação

O primeiro subcomponente analisado foi o de **motivação**, cujos itens são decompostos com base no modelo ARCS (Keller, 1983), que define quatro categorias para representar a motivação no design instrucional: atenção, relevância, confiança e satisfação. A Figura 12 apresenta os itens correspondentes a cada categoria, bem como os percentuais de respostas marcadas pelos estudantes aos cinco níveis da escala Likert utilizada, que variava de “discordo fortemente” (-2) a “concordo fortemente” (+2). Embora o objeto da avaliação tenha sido um sistema gamificado, manteve-se a palavra “jogo” nos itens do questionário porque os elementos de jogos causavam a impressão, na percepção leiga dos estudantes, de que estavam usando um “jogo”. Portanto, procurou-se utilizar uma palavra com a qual eles tinham maior familiaridade, em vez de introduzir um novo jargão técnico.

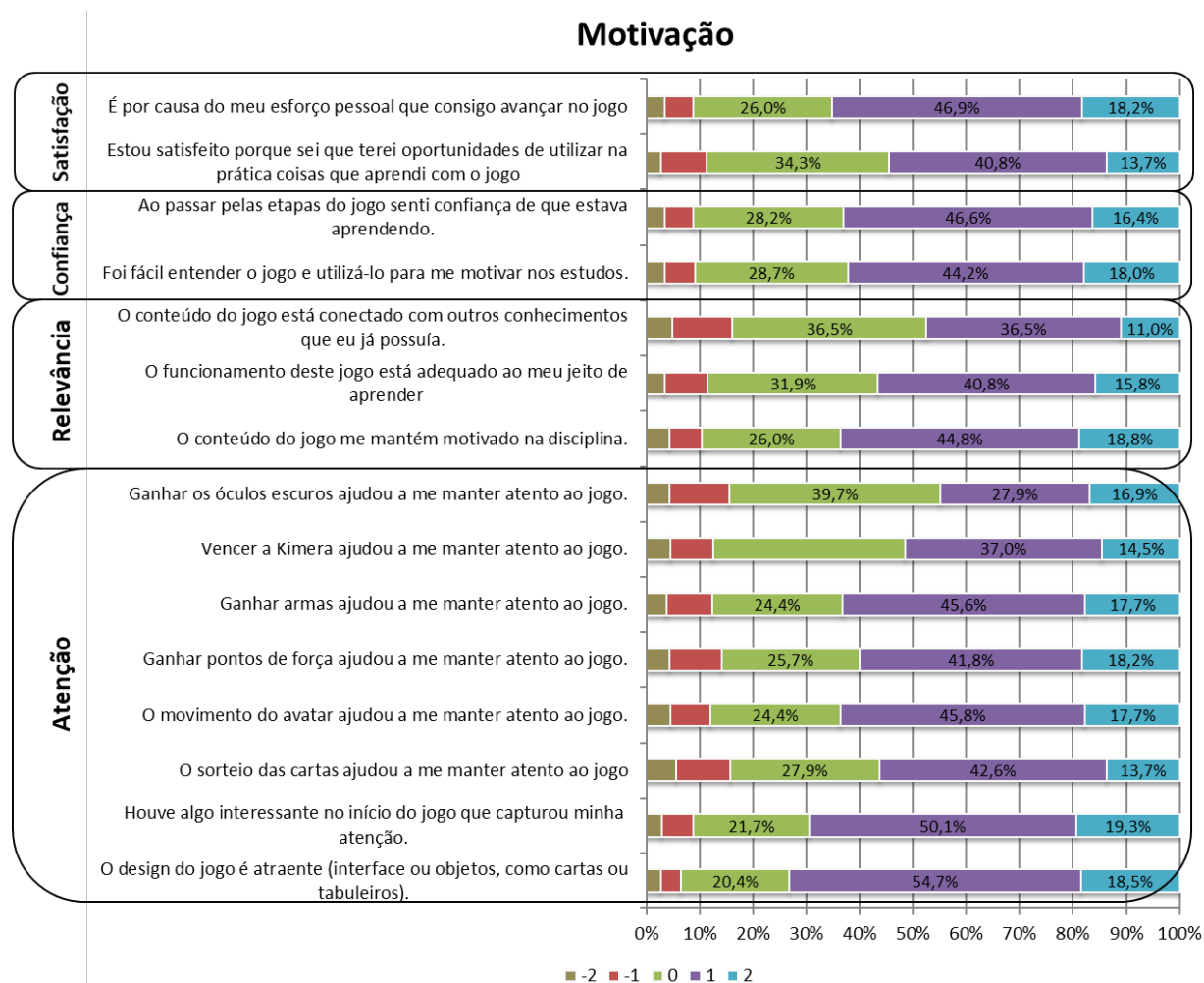


Figura 12: Itens para avaliação do subcomponente “motivação” no juiz online gamificado.

Com relação à categoria **satisfação**, a maioria dos estudantes concordou com as duas afirmações propostas (65,1% e 54,5%). Por exemplo, na afirmação “*É por causa do meu esforço pessoal que consigo avançar no jogo*”, 46,9% dos estudantes apenas concordaram e 18,2% concordaram fortemente. De fato, os alunos perceberam que, quanto mais resolviam os exercícios e se esforçavam na disciplina, mais avançavam no mapa do juiz online.

A categoria **confiança** continha dois itens, os quais foram bem avaliados pelos alunos, que entenderam que o juiz online gamificado os ajudou durante a disciplina. Isso fica patente na nota atribuída à afirmação “*Foi fácil entender o jogo e utilizá-lo para me motivar nos estudos*”, avaliada positivamente por 62,0% dos estudantes (notas 1 e 2). Além disso, a outra afirmação,

“Ao passar pelas etapas do jogo senti confiança de que estava aprendendo”, também foi avaliada com maioria de notas 1 e 2. Alguns alunos ainda deixaram comentários, tais como:

- “Dá uma sensação que estamos mesmo aprendendo e se (sic) envolvendo.”
- “Foi fácil porque quanto mais acertávamos, mais o avatar ganhava pontos de força e de deslocamento [...]”

A categoria **relevância** diz respeito à sensação do estudante sobre a importância do conteúdo da gamificação. A afirmação “o conteúdo do jogo me mantém motivado na disciplina”, recebeu uma nota muito expressiva, em que 44,8% dos estudantes avaliaram com nota 1 e 18,8% dos estudantes avaliaram com a nota 2. Esses resultados são um forte indício de que os elementos de gamificação contidos no CodeBench influenciam na motivação do estudante. Além desses resultados, um dos alunos teceu o seguinte comentário a respeito desse item: “Sim, o jogo realmente me deixou mais motivado para disciplina. Fez quebrar a tensão de uma matéria relativamente complicada”.

A categoria de **atenção** também foi avaliada positivamente pelos alunos. Ela inclui itens sobre a capacidade do design (interface ou objetos, como cartas ou tabuleiros), das animações (como o movimento do personagem no mapa), ou até mesmo do fator sorte (como o sorteio de cartas) de manter a atenção dos estudantes. Os dois itens que receberam maiores notas nessa categoria foram: “houve algo interessante no início do jogo que capturou minha atenção” e “o design do jogo é atraente (interface ou objetos, como cartas ou tabuleiros)”.

5.8.2 Percepção sobre experiência do usuário

O segundo subcomponente analisado pelo questionário é o da **experiência do usuário** (Figura 13). Ele é dividido em seis categorias: **competência, divertimento, desafio, controle e interação social, imersão**.

A categoria **competência**, que avalia se o jogador conseguiu vencer os desafios do jogo através de suas habilidades, foi bem classificada pelos estudantes em geral. Por exemplo, mais de 60% dos alunos respondeu de forma positiva (notas 1 ou 2) ao item “consegui atingir os objetivos do jogo por meio das minhas habilidades”, ou seja, eles sentiram que, para conseguir ir bem no jogo, dependiam exclusivamente das habilidades adquiridas ao longo da disciplina.

O mesmo se deu na categoria **diversão**, que recebeu maioria de notas 1 e 2. O objetivo dela é identificar se os estudantes acharam o jogo divertido e prazeroso e se, dessa forma, eles indicariam o jogo para seus amigos. O item “me diverti com o jogo” recebeu 63,8% de notas positivas, indicando assim que o jogo realmente foi divertido. Perguntados sobre a possibilidade de jogarem esse jogo em outra disciplina, a maioria dos alunos respondeu que gostaria de utilizar, que gostaria de ter jogado mais e que recomendaria o jogo para os seus colegas.

A categoria **desafio** recebeu uma grande quantidade de notas neutras (0). Por exemplo, o item “o jogo evolui num ritmo adequado e não fica monótono – oferece novos obstáculos, situações ou variações de atividades” recebeu 31,9% como nota 0, ou seja, indiferente. Essa nota reflete-se nos seguintes comentários dos estudantes:

- “Ao passar do tempo, o jogo fica meio monótono, pois não apresenta muitas variedades de atividades.”
- “Dava pra ter umas questões extras para praticar mais, com uns desafios mais difíceis e interessantes (mas não necessariamente valendo nota)”
- “Poderia ser um pouco mais extenso, apenas um pouco.”

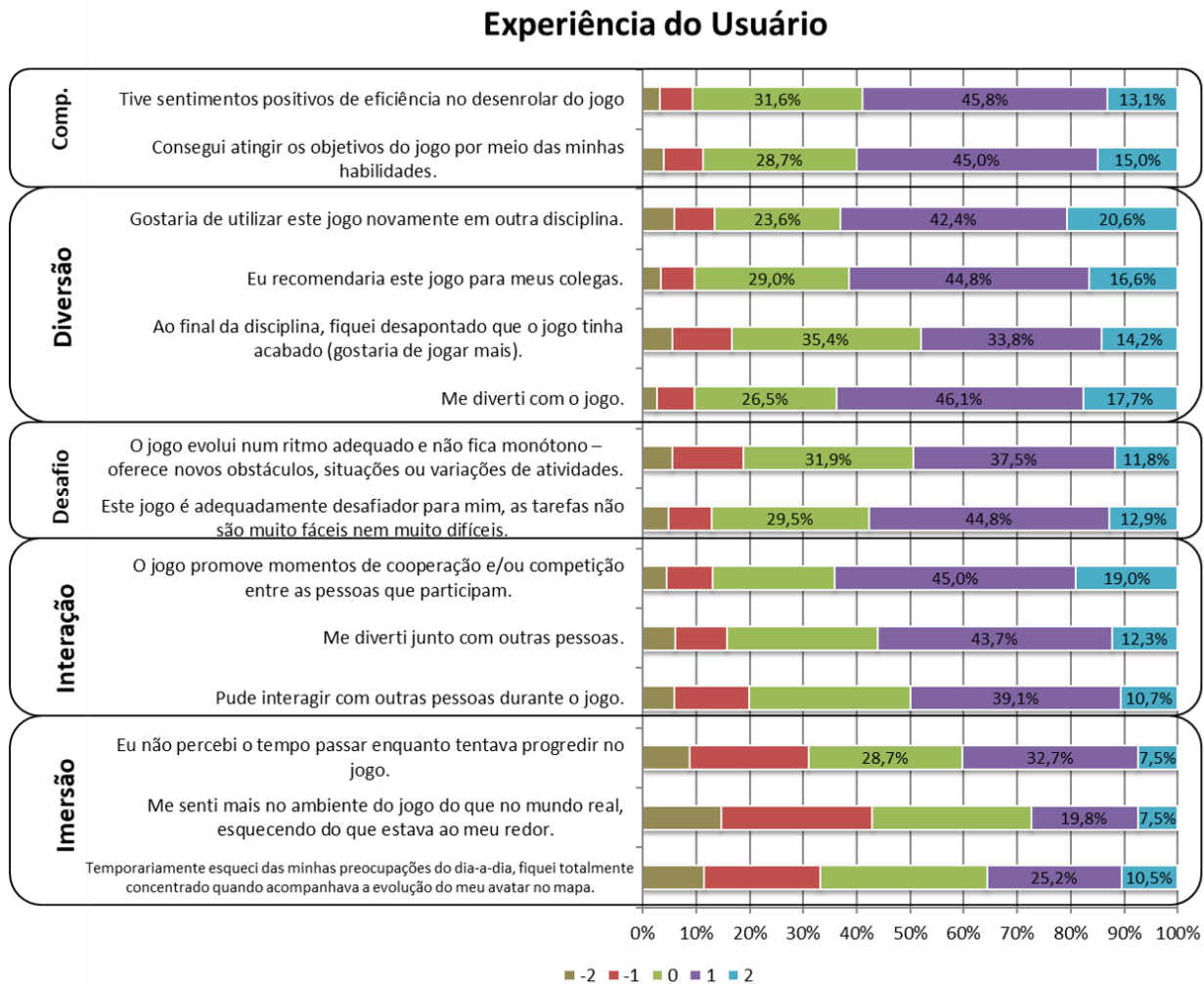


Figura 13: Itens para avaliação do subcomponente “experiência do usuário” ao utilizar o juiz online gamificado.

Como será relatado na Seção 5.9, as sugestões e comentários dos estudantes foram agrupadas sistematicamente por meio do Método de Explicitação do Discurso Subjacente (MEDS) (Nicolaci-da-Costa, 2007), a fim de, em um trabalho futuro, aprimorar esses itens que foram avaliados de forma indiferente ou negativa.

A categoria **interação social** inclui itens relacionados ao sentimento de conexão, empatia, cooperação e competição entre os estudantes. Ela foi muito bem avaliada na maioria dos itens, entre os quais destaca-se a competição promovida pelo juiz online gamificado (“o jogo promove momentos de cooperação e/ou competição entre as pessoas que participam”). Nesse item, 64% dos alunos responderam de forma positiva, sendo que 19,0% responderam com a nota máxima 2. Porém, apesar da afirmação “pude interagir com outras pessoas durante o jogo” ter sido avaliada de forma positiva, ela foi a que teve mais votos indiferentes (0) na categoria. De fato, a gamificação promovia poucas oportunidades de colaboração entre os estudantes. No entanto, novas ferramentas de cooperação serão desenvolvidas em versões futuras do jogo.

Por fim, a categoria **imersão**, recebeu a maior porcentagem de respostas negativas no subcomponente **experiência do usuário**. Provavelmente, isso se deve ao fato de o juiz online gamificado não se tratar de um jogo em si, pois as conquistas dependem dos resultados alcançados pelos alunos nos exercícios e avaliações. Isso pode ser percebido por meio da

afirmação “*Me senti mais no ambiente do jogo do que no mundo real, esquecendo do que estava ao meu redor*” que recebeu, dentre todos os itens, a maior porcentagem (28,2%) de nota -1.

5.8.3 Percepção sobre ganho de aprendizagem

Por fim, o questionário também continha itens relacionados à percepção dos estudantes com respeito à **aprendizagem**, cujos resultados são reportados na Figura 14. Como destaque, a afirmação “*o jogo contribuiu para a minha aprendizagem na disciplina*” teve um resultado muito positivo, em que 46,1% dos alunos responderam com a nota 1 e 17,7% responderam com a nota 2. Dessa forma, podemos afirmar que, segundo a percepção dos alunos, o sistema gamificado realmente contribuiu para um melhor aproveitamento da disciplina.

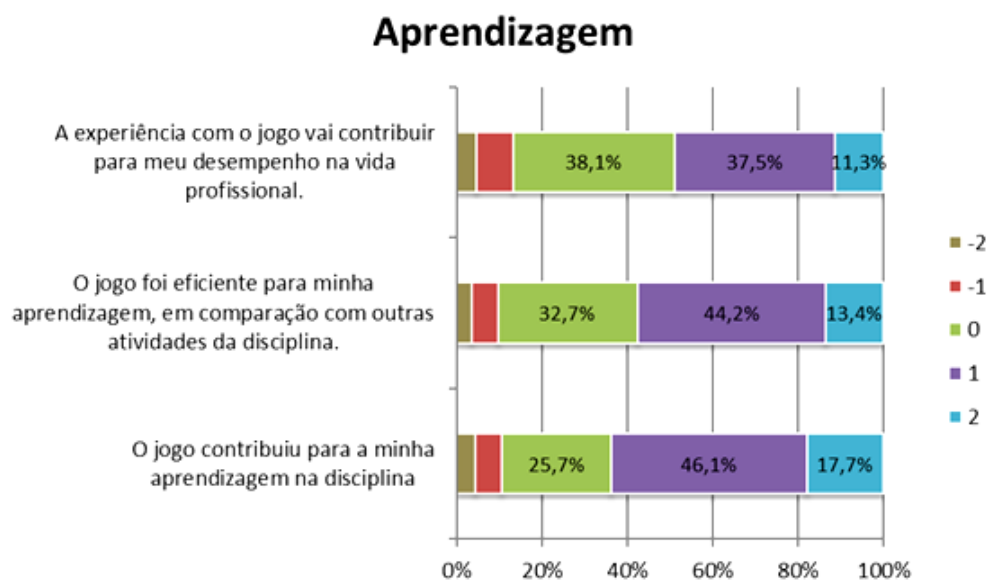


Figura 14: Itens para avaliação do subcomponente “aprendizagem” no juiz online gamificado.

5.9 Análise Qualitativa

O questionário aplicado ao final da disciplina também continha alguns campos livres para que os estudantes escrevessem três pontos fortes e sugestões de melhoria do sistema gamificado. A seguir, são listados alguns exemplos de pontos fortes comentados pelos alunos:

- “*O mapa interativo fornece uma noção mais dinâmica do quanto o aluno está aprendendo*”.
- “*O sentimento adicional de ‘competitividade’ do jogo motiva o aluno*”.
- “*Sentimento de progressão e evolução contínua no aprendizado da disciplina*”.
- “*O sorteio das cartas*”.
- “*O jogo estimula o aluno a fazer os exercícios*”.
- “*... o fato de ir ganhando armas e acessórios motivava*”.
- “*O designer (sic) do jogo, as armas e os avatares*”.
- “*O jogo ajudou a manter a atenção no conteúdo já que só assim se poderia avançar*”.
- “*Criou uma aliança entre os jogadores para juntos poderem derrotar Quimera*”.
- “*... é um diferencial no que se refere à vida acadêmica, onde tudo é sério e complexo*”.

A seguir, são listadas algumas das sugestões recomendadas pelos estudantes, que podem ser incorporadas em futuras versões do sistema gamificado:

- “*Deveria ter mais animações*”.
- “*Ganhar Defesa também seria uma boa melhoria*”.
- “*Menos ‘aleatoriedade’ e mais merecimento*”.

- “Ter uns exercícios "extras", como desafio, para andarmos uma quantidade maior de casas, e ganharmos uma quantidade maior de energia”.
- “Colocar interação entre várias turmas (poder ver o progresso de outras turmas)”.
- “Inserir inimigos/desafios antes da Quimera”.
- “Aumentar a quantidade de questões nos desafios e nas provas”.
- “Mais cartas, mais poderes e novas armas”.
- “Ganhar pontos ao matar o Quimera”.

Essas sugestões dos estudantes foram agrupadas em categorias segundo o MEDS (Nicolaci-da-Costa, 2007). Essas categorias, apresentadas na Tabela 4, resultam da comparação sistemática das respostas em busca de recorrências.

Tabela 4: Categorias que emergiram a partir da análise dos comentários dos alunos.

#	Categoria	Ocorrências
01	Design do sistema	92
02	Desafios	90
03	Recompensas	80
04	Aleatoriedade na gamificação	66
05	Jogabilidade	31
06	Opções para relatar bugs no sistema	27
07	Sem sugestão de melhoria	22
08	Interação com outros participantes da mesma turma	20
09	Ajuda com as regras	19
10	Tempo do jogo	15
11	Motivação a vencer	8
12	Interação com outras turmas	7
13	Facilidade no uso da ferramenta	5
14	Diversão	3
15	Feedback	2
16	Imersão	1

A categoria 01, que se refere ao design do sistema gamificado, foi a que mais recebeu sugestões dos alunos (92 sugestões). Os alunos tiveram bons sentimentos quanto ao design, porém, no questionário, a maioria das sugestões referem-se à adição de novos avatares, opções de roupas e mais animações.

Três categorias que estão interligadas, a 02 (desafios), a 03 (recompensas) e a 04 (aleatoriedade na gamificação), tiveram uma grande quantidade de sugestões (90, 80 e 66 respectivamente). Os alunos criticaram bastante a mecânica de aleatoriedade das cartas, que davam recompensas (andar no mapa ou pontos de força). O mecanismo de sorteio de cartas inicialmente implementado era puramente estocástico, ou seja, “sem memória”. Contudo, os seres humanos têm memória. Por isso, quando os alunos ganhavam a mesma recompensa repetidas vezes, tinham a percepção de que estavam sendo injustiçados. Para contornar isso, o mecanismo de sorteio foi alterado para harmonizar a ordem do sorteio das recompensas e dessa forma coincidir com a ideia distorcida que o cérebro humano possui de “aleatoriedade”.

Os estudantes também sugeriram a criação de mais desafios durante a jornada do seu avatar, tais como introduzir monstros intermediários, incluir mais questões para resolver, estender o mapa para conter mais obstáculos, entre outros. Já na categoria 05 (jogabilidade), com 31 sugestões, os estudantes abordaram a adição de novas dinâmicas e até mesmo novas regras, como: “sistema flexível de escolhas de rotas”.

Outras sugestões foram relatadas devido à falta de suporte sentida pelos alunos: na categoria 06 (opções para relatar bugs no sistema), trata-se da possibilidade do sistema ter uma

opção para relatar bugs e, na categoria 09 (ajuda com as regras), alguns estudantes relataram que precisavam de mais explicações das regras antes de começarem a usar a gamificação.

Em relação à categoria 08, os alunos sugeriram competição entre os jogadores e até mesmo entre outras turmas (categoria 12), como batalhas, lutas ou desafios com exercícios. Por fim, o tema imersão, apesar de ter recebido notas baixas no questionário MEEGA adaptado, como indica a Figura 12, recebeu apenas uma sugestão de melhoria.

5.10 Ameaças à Validade

Acreditamos que este estudo contribui valiosamente para a comunidade que pesquisa o uso de sistemas gamificados em disciplinas introdutórias de programação. Contudo, temos consciência de que há algumas limitações quanto a possíveis generalizações. Nesta seção, discutem-se as ameaças que podem afetar a validade dos resultados, concentrando-se em quatro categorias, segundo Whölin, Höst, e Henningsson (2003): interna, externa, de conclusão e de construto.

Validade interna: a validade interna diz respeito a fatores que podem afetar as variáveis dependentes sem o conhecimento do pesquisador. Como não foi possível aleatorizar os sujeitos para dois grupos distintos (experimental e controle), esse foi um estudo quase-experimental. O estudo foi conduzido com alunos de períodos diferentes (2016/1 e 2017/1), com conhecimentos prévios distintos, tornando-se uma possível ameaça à validade. Outro risco é que a experiência prévia com alguma linguagem de programação foi por autodeclaração. Em 2016/1, um percentual maior declarou conhecer alguma linguagem de programação e também um percentual maior não respondeu essa questão. No entanto, os dados dos alunos com e sem experiência prévia foram analisados separadamente, como pode ser visto na Figura 8. Além desse fator, em nossa experiência como docentes, normalmente não notamos diferença significativa de base de formação do Ensino Médio em dois anos consecutivos, a não ser quando há mudança no sistema de entrada à universidade, que não foi o caso desses dois anos.

Validade externa: a validade externa está relacionada com a possibilidade de generalizar os resultados dos experimentos para outros contextos. Sabemos que através do nosso quase-experimento não é possível generalizar para todos os contextos, particularmente porque há participação de seres humanos. No entanto, acreditamos que os resultados podem ser válidos para contextos similares, tal como ensino de programação em Python para *non-majors*. Quanto ao material e atividades utilizados, apesar de terem sido levemente diferentes de um período para outro, ou seja, revisados e com 12 questões a mais em 2017/1, o mesmo professor coordenador da disciplina (um dos autores deste artigo) participou do processo de revisão deles. Assim, acreditamos que o nível de qualidade e objetividade se manteve de um período para o outro. A generalização para outros contextos (outras universidades, alunos *majors*, outras linguagens etc.) é desejável e está em nossos planos para trabalhos futuros.

Validade de conclusão: a validade de conclusão está relacionada com a possibilidade de chegar a conclusões corretas a partir da relação entre o tratamento e os resultados de experimentos. Para evitar possíveis ameaças à validade relacionadas ‘a coleta de dados, estes foram coletados a partir do log do próprio juiz online e do sistema de registro acadêmico da universidade. Foram realizados testes estatísticos para rejeitar as seis hipóteses nulas e, com isso, a validade dos resultados obtidos são aceitáveis. Como a desistência de uma disciplina pode ocorrer por diversos fatores não controláveis, optou-se por excluir da análise os alunos que trancaram ou reprovaram por falta em IPC. Dessa forma, tentamos amenizar o viés causado por esses fatores, conseguindo mensurar com mais exatidão o desempenho real do aluno na disciplina.

Validade de construto: a validade do construto está associada à relação entre os conceitos e teorias que embasam os experimentos e o que é medido e afetado. Uma delas é que, em cada ano, as turmas foram ministradas por professores distintos. Para amenizar esse possível viés, todos os professores utilizaram os mesmos materiais didáticos e atividades, tanto no grupo

experimental quanto no controle. Além disso, pode ter havido diferença nas dificuldades das questões das avaliações entre os anos de 2016 e 2017. Contudo, o professor coordenador da disciplina (um dos autores deste artigo) revisou todas as questões antes de serem aplicadas, amenizando os efeitos desse risco de ameaça à validade. Em relação ao questionário qualitativo, existe a possibilidade de alguns alunos terem se sentido pressionados a fornecer respostas favoráveis, ainda que respondido de forma anônima e esclarecido que isso não interferiria nas notas escolares. Nesse caso, em vez de termos opiniões sinceras, as verdadeiras percepções e experiências pessoais podem ter sido mascaradas. Para tentar contornar esse risco, as sugestões e comentários dos alunos foram agrupadas sistematicamente por meio do MEDS.

6 Conclusão e Trabalhos Futuros

O ato de programar e resolver problemas exige a mobilização de diversas habilidades cognitivas do aprendiz. Grande parte dos estudantes que ingressam em cursos superiores vem do ensino médio com uma base lógico-matemática deficiente, o que influencia na aprendizagem de algoritmos. Além disso, muitos estudantes que não têm familiaridade com programação tendem a não se interessar por esse tipo de disciplina. Essa falta de interesse gera desmotivação e faz o aluno desistir da disciplina ante a pequenas dificuldades. A gamificação pode ajudá-los nesse sentido, pois torna a aprendizagem divertida e desafiadora. Além disso, os estudantes sentem que estão evoluindo na disciplina quando avançam no sistema gamificado. Esse *feedback*, junto com a colaboração e a competição, são poderosas ferramentas de engajamento estudantil.

Neste artigo, foram apresentadas as características de um juiz *online* gamificado com a finalidade de motivar alunos na aprendizagem de programação. Por meio de análises quantitativas, foi possível verificar aumento significativo nota final na disciplina entre os alunos não desistentes, fato que não dependeu da experiência prévia reportada pelos alunos, mas que se concentrou mais nos alunos que cursaram IPC pela primeira vez (não repetentes). Além disso, verificou-se um aumento significativo em alguns atributos de uso do juiz online, por exercício prático: número de logins, número de testes de código e tempo de uso do IDE. Quanto ao número de exercícios corretos, não se verificou diferença significativa em relação à turma de controle, mas houve uma redução na variância, principalmente entre os alunos de desempenho mais baixo. Por fim, observou-se que o uso do juiz online gamificado aponta para uma diminuição da associação entre a nota da prova de Matemática do Enem e o desempenho em programação, o que não se verificou em outras disciplinas comuns aos cursos analisados.

Também foi realizada uma análise qualitativa da percepção dos estudantes sobre o impacto percebido do juiz online gamificado sobre o próprio aprendizado, por meio de um questionário de avaliação de jogos educacionais adaptado. A maioria dos alunos respondeu de forma positiva às perguntas referentes a motivação, experiência do usuário e percepção da aprendizagem.

Portanto, este estudo traz algumas contribuições para a pesquisa sobre o uso da gamificação com o fim de engajar estudantes em disciplinas de programação introdutória, entre as quais destacam-se as seguintes:

- Gamificação de um juiz *online* para promover o engajamento estudantil;
- Experimento empírico realizado com mais de 600 estudantes de graduação, de diversos cursos de engenharia e ciências exatas;
- Resultados qualitativos sobre a motivação e a experiência do estudante durante o uso de um juiz online gamificado;

- Resultados quantitativos que demonstraram o aumento das competências dos estudantes na disciplina de IPC em relação aos indicadores de desempenho em outras atividades curriculares.

Como trabalhos futuros, pretende-se aplicar as melhorias sugeridas pelos alunos, utilizando um modelo para desenvolvimento sistemático de sistemas gamificados, como o proposto por Klock et al. (2015). Também é necessário estudar o perfil de jogadores, para entender quais elementos de jogos são mais utilizados por cada perfil e proporcionar uma experiência imersiva e personalizada. Também é preciso aplicar um instrumento pré e pós intervenção, como um inventário de conceitos, a fim de medir-se de forma mais objetiva a experiência prévia em programação. Por fim, planeja-se implementar estratégias colaborativas de gamificação, para aumentar a interação entre os estudantes.

Referências

- Alves, F. (2015). *Gamification: como criar experiências de aprendizagem engajadoras* (2nd ed). DVS Editora. [GS Search]
- Bez, J. L., Tonin, N. A., & Rodegheri, P. R. (2014). URI Online Judge Academic: A tool for algorithms and programming classes. In *Proceedings of the International Conference on Computer Science & Education*, 149-152. Vancouver, BC. doi: [10.1109/ICCSE.2014.6926445](https://doi.org/10.1109/ICCSE.2014.6926445) [GS Search]
- Bosse, Y., & Gerosa, M. A. (2015). Reprovações e Trancamentos nas Disciplinas de Introdução à Programação da Universidade de São Paulo: Um Estudo Preliminar. In *XXIII Workshop sobre Educação em Computação*. Recife.
- Branch, R. M. (2009). *Instructional design: The ADDIE approach* (Vol. 722). Springer Science & Business Media. doi: [10.1007/978-0-387-09506-6](https://doi.org/10.1007/978-0-387-09506-6) [GS Search]
- Campos, A., Gardiman, R., & Madeira, C. (2015). Aplicação da Gamificação na Disciplina de Empreendedorismo. In *XXIII Workshop sobre Educação em Computação*. Recife. [GS Search]
- Carvalho, Leandro S. G.; Gadelha, Bruno; Nakamura, Fabíola G.; Oliveira, David B. F.; Oliveira, Elaine H. T. (2016). Ensino de Programação para Futuros Não-Programadores: Contextualizando os Exercícios com as Demais Disciplinas de mesmo Período Letivo. In *Anais do XXIV Workshop sobre Educação em Computação*. p. 2116-2125. [GS Search]
- Chaves, J. O. M., Castro, A. F., Lima, R. W., Lima, M. V. A., & Ferreira, K. H. (2013). Integrando Moodle e Juizes Online no Apoio a Atividades de Programação. In *Anais do Simpósio Brasileiro de Informática na Educação* (Vol. 24, No. 1, p. 244). doi: [10.5753/cbie.sbie.2013.244](https://doi.org/10.5753/cbie.sbie.2013.244) [GS Search]
- Dick, W., Carey, L., & Carey, J. O. (2001). *The Systematic Design of Instruction* (7th ed). Allyn & Bacon.
- Fardo, M. L. (2013). *A gamificação como estratégia pedagógica: estudos de elementos dos "games" aplicados em processos de ensino e aprendizagem*. Dissertação de mestrado em educação, Universidade de Caxias do Sul, Caxias do Sul, RS. [GS Search]
- Francisco, R. E., Ambrósio, A. P. L., Pereira Júnior, C. X., & Fernandes, M. A. (2018). Juiz online no ensino de CS1- lições aprendidas e proposta de uma ferramenta. *Revista Brasileira de Informática na Educação*; v. 26, n. 03. doi: [10.5753/rbie.2018.26.03.163](https://doi.org/10.5753/rbie.2018.26.03.163) [GS Search]

- Galvão, L., Fernandes, D., & Gadelha, B. (2016). *Juiz online como ferramenta de apoio a uma metodologia de ensino híbrido em programação*. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE) (Vol. 27, No. 1, p. 140). doi:[10.5753/cbie.sbie.2016.140](https://doi.org/10.5753/cbie.sbie.2016.140). [GS Search]
- INEP. (2017) *Nota Técnica N° 33/2017/CGCQES/DAES - Cálculo do IDD*. Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2016/nota_tecnica_n33_2017_cgcqes_daes_calculo_idd.pdf. Acesso em 30 de setembro de 2019.
- Kapp, K. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons. [GS Search]
- Kasahara, R., Sakamoto, K., Washizaki, H., & Fukazawa, Y. (2019). Applying Gamification to Motivate Students to Write High-Quality Code in Programming Assignments. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '19)*, 92–98. ACM, New York. doi:[10.1145/3304221.3319792](https://doi.org/10.1145/3304221.3319792) [GS Search]
- Keller, J. M. (1983). Development and Use of the ARCS Model of Motivational Design. *Classic writings on instructional technology*, 10(3), 2–10. doi:[10.1007/BF02905780](https://doi.org/10.1007/BF02905780)
- Klock, A. C. T., da Cunha, L. F., de Carvalho, M. F., Rosa, B. E., Anton, A. J., & Gasparini, I. (2015). Gamification in e-learning systems: A conceptual model to engage students and its application in an adaptive e-learning system. *Learning and collaboration technologies*, 595–607. Springer, Cham. doi:[10.1007/978-3-319-20609-7_56](https://doi.org/10.1007/978-3-319-20609-7_56) [GS Search]
- Lowry, P. B., Gaskin, J., Twyman, N., Hammer B., & Roberts, T. (2012). Taking ‘Fun and Games’ Seriously: Proposing the Hedonic Motivation System Adoption Model (HMSAM). *Journal of the Association for Information Systems*, vol. 14, no. 11, pp. 617–671. doi:[10.17705/1jais.00347](https://doi.org/10.17705/1jais.00347) [GS Search]
- McCabe, T. J. (1976). A Complexity Measure. *IEEE Transactions on Software Engineering SE-2*, 4, 308–320. doi:[10.1109/TSE.1976.233837](https://doi.org/10.1109/TSE.1976.233837)
- Nicolaci-da-Costa, A. M. (2007). O campo da pesquisa qualitativa e o Método de Explicitação do Discurso Subjacente (MEDS). *Psicologia: reflexão e crítica*, 20(1), 65–73. doi:[10.1590/S0102-79722007000100009](https://doi.org/10.1590/S0102-79722007000100009) [GS Search]
- Oliveira Júnior, J. G., Noronha, R. V., & Kaestner, C. A. A. (2017). Método de Seleção de Atributos Aplicados na Previsão da Evasão de Cursos de Graduação. *Revista de Informática Aplicada*, 13(2). [GS Search]
- Peixoto, M. M., Silva, C., Gonçalves, E., & Vilena, J. (2015). Um Mapeamento Sistemático de Gamificação em Software Educativo no Contexto da Comunidade Brasileira de Informática na Educação. In *Anais do XXI Workshop de Informática na Escola*. doi:[10.5753/cbie.wie.2015.584](https://doi.org/10.5753/cbie.wie.2015.584) [GS Search]
- Petit, J., Roura, S., Carmona, J., Cortadella, J., Duch, A., Gimenez, O., Mani, A., Mas, J., Rodriguez-Carbonella, E., Rubio, A., Pedro, J., & Venkataramani, D. (2018). Jutge.org: Characteristics and Experiences. In *IEEE Transactions on Learning Technologies*, vol. 11, no. 3, 321–333. doi:[10.1109/TLT.2017.2723389](https://doi.org/10.1109/TLT.2017.2723389) [GS Search]
- Petri, G., Wangenheim, C. G., & Borgatto, A. F. (2018). MEEGA+, Systematic Model to Evaluate Educational Games. In N. Lee (Ed.) *Encyclopedia of Computer Graphics and Games*. Springer. doi:[10.1007/978-3-319-08234-9_214-1](https://doi.org/10.1007/978-3-319-08234-9_214-1)
- Prather, J., Pettit, R., McMurry, K., Peters, A., Homer, J., & Cohen, M. (2018). Metacognitive Difficulties Faced by Novice Programmers in Automated Assessment Tools. In *Proceedings*

- of the 2018 ACM Conference on International Computing Education Research, 41–50. doi:[10.1145/3230977.3230981](https://doi.org/10.1145/3230977.3230981) [GS Search]
- Ribeiro, Ralph B. S.; Oliveira, David B. F.; Carvalho, Leandro S. G.; Oliveira, Elaine H. T. de. (2018). Gamificação de um Sistema de Juiz Online para Motivar Alunos em Disciplina de Programação Introdutória. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE). pp. 805–814. doi:[10.5753/cbie.sbie.2018.805](https://doi.org/10.5753/cbie.sbie.2018.805) [GS Search].
- Savi, R., Wangenheim, C., & Borgatto, A. (2011). Um modelo de avaliação de jogos educacionais na engenharia de software. In *Anais do XXV Simpósio Brasileiro de Engenharia de Software (SBES 2011)*, São Paulo. doi:[10.1109/SBES.2011.27](https://doi.org/10.1109/SBES.2011.27) [GS Search]
- Silva, A. C., Correa, C. S., Coelho, D. A., Silva Neto, D. T., Ferraz, L., Xavier, M. M., Reis, R. S., Rocha, F. A., & Santos, P. A. (2016). Análise dos Índices de Reprovação nas Disciplinas de Cálculo I e AVGA do Curso de Engenharia Elétrica do Instituto Federal da Bahia de Vitória da Conquista. In *Proceedings of the XIV International Conference on Engineering and Technology Education*. doi:[10.13140/2.1.3348.0487](https://doi.org/10.13140/2.1.3348.0487) [GS Search]
- Silva, T. S. C., Melo, J. C. B., & Tedesco, P. C. D. A. R. (2018). A Model to Promote Student Engagement in Programming Learning Using Gamification. *Brazilian Journal of Computers in Education*, 26(03), 120. doi:[10.5753/rbie.2018.26.03.120](https://doi.org/10.5753/rbie.2018.26.03.120) [GS Search]
- Sinly, C., Rusli, A., & Winarno, P. M. (2018). Utilizing Gamification to Improve User Participation in Online Judge. In *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*(pp. 543-547). Toyama, Japan. doi: [10.1109/SCIS-ISIS.2018.00096](https://doi.org/10.1109/SCIS-ISIS.2018.00096) [GS Search]
- Tondello, G. F., Mora, A., & Nacke, L. E. (2017). Elements of Gameful Design Emerging from User Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, New York. doi:[10.1145/3116595.3116627](https://doi.org/10.1145/3116595.3116627) [GS Search]
- Wangenheim, C. G., & Wangenheim, A. (2012). *Ensinando computação com jogos*. Bookess Editora, Florianópolis, SC, Brasil.
- Wasik, S., Antczak, M., Badura, J., Laskowski, A., & Sternal, T. (2018). A Survey on Online Judge Systems and Their Applications. *ACM Computing Surveys*, 51, 1, Article 3. doi: 10.1145/3143560. doi:[10.1145/3143560](https://doi.org/10.1145/3143560) [GS Search]
- Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press. [GS Search]
- Wöhlin, C., Höst, M., & Henningsson, K. (2003). Empirical research methods in software engineering. *Empirical methods and studies in software engineering*, 7–23. Springer, Berlin, Heidelberg. doi:[10.1007/978-3-540-45143-3_2](https://doi.org/10.1007/978-3-540-45143-3_2) [GS Search]