

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM INFORMÁTICA

THYAGO TENÓRIO MARTINS DE OLIVEIRA

**Um Modelo de Avaliação por Pares
Gamificado para Ambientes Educacionais
Online: Um Experimento com o MeuTutor**

Maceió
2015

Thyago Tenório Martins de Oliveira

**Um Modelo de Avaliação por Pares Gamificado para
Ambientes Educacionais Online: Um Experimento
com o MeuTutor**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Prof. Dr. Ig Ibert Bittencourt
Santana Pinto

Maceió
2015

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

O48m Oliveira, Thyago Tenório Martins de.
Um modelo de avaliação por pares gamificado para ambientes educacionais online: um experimento com o meu tutor / Thyago Tenório Martins de Oliveira. – 2015.
190 f. : il.

Orientador: Ig Ibert Bittencourt Santana Pinto.
Dissertação (Mestrado em Informática) – Universidade Federal de Alagoas. Instituto de Computação, Maceió, 2015.
Bibliografias: f. 166-172.
Apêndices: f. 173-190.

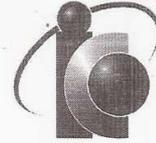
1. Ambientes online – Modelagem – Educação. 2. Avaliação por pares.
3. Gamificação – Técnicas. I. Título.

CDU: 004.42:37



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Informática – PpgI
Instituto de Computação

Campus A. C. Simões BR 104-Norte Km 14 BL 12 Tabuleiro do Martins
Maceió/AL - Brasil CEP: 57.072-970 | Telefone: (082) 3214-1401



Membros da Comissão Julgadora da Dissertação de Mestrado de Thyago Tenório Martins de Oliveira, intitulada: “*Um Modelo de Avaliação por Pares Gamificado para Ambientes Educacionais Online: Um Experimento com o MeuTutor*”, apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas em 15 de junho de 2015, às 09h00min, no Miniauditório do Instituto de Computação da UFAL.

COMISSÃO JULGADORA

Prof. Dr. Ig Ibert Bittencourt Santana Pinto
UFAL – Instituto de Computação
Orientador

Prof. Dr. Seiji Isotani
USP – Universidade de São Paulo
Examinador

Prof. Dr. Balduino Fonseca dos Santos Neto
UFAL – Instituto de Computação
Examinador

Prof. Dr. Alan Pedro da Silva
UFAL – Instituto de Computação
Examinador

À Deus, pois sem ele eu não teria forças para essa longa jornada. À minha família, pelo cuidado e dedicação que me deram, e, em alguns momentos, a esperança necessária para seguir. À minha noiva, por sua capacidade de acreditar e investir em mim.

AGRADECIMENTOS

Em primeiro lugar, acima de tudo, agradeço à Deus pois sem ele esta jornada não teria sido cumprida.

Agradeço à minha família que sempre me apoiou ao longo deste processo, especialmente aos meus pais que sempre estiveram comigo, principalmente nos momentos difíceis ao longo do curso e da vida, por terem me propiciado a oportunidade de estudar e sempre acreditar em meu potencial.

À minha noiva Regina Stein, por todo carinho e dedicação, pela preocupação e zelo, presente em todos os momentos, inclusive nos momentos mais difíceis. Agradeço imensamente pelos seus ensinamentos. E a sua mãe, minha amiga Célia Regina, por toda a ajuda necessária em momentos que eu necessitei, disponibilizando seu precioso tempo para me ajudar.

Ao meu orientador Ig Ibert, cujos ensinamentos foram, são e serão de grande valia para o meu futuro acadêmico e profissional. Obrigado por aguçar a minha "inquietação filosófica" e por me mostrar outras maneiras de tentar compreender aquilo que aparentemente não pode ser explicado. Agradeço por acreditar em meu potencial e pelos puxões de orelha quando foi necessário, me trazendo de volta ao caminho nos momentos em que foi preciso.

Aos professores do curso de computação, no qual tive a oportunidade de aprender com eles, contribuindo na minha vida acadêmica e por tanta influência na minha futura vida profissional.

Agradecimento em especial a todos os integrantes do grupo NEES, ao qual pertencço desde o início do curso de graduação, e da empresa MeuTutor (especialmente Olavo Holanda, Endhe Elias e Marcelo Nunes). Pudemos compartilhar nossas experiências, horas de trabalho e dedicação com foco em nosso objetivo. Agradeço pela colaboração, apoio, conversas, amizade, enfim, por tudo que pudemos compartilhar, desde a convivência, as alegrias, as frustrações e as descobertas.

É muito difícil agradecer a todas as pessoas que de algum modo, nos momentos serenos e/ou apreensivos, fizeram ou fazem parte da minha vida. Por isso agradeço à todos de coração.

*Não importa quanto a vida possa ser ruim,
sempre existe algo que você pode fazer, e triunfar.
Enquanto há vida, há esperança.
(Stephen Hawking)*

RESUMO

Nos últimos anos, o uso de tecnologias de informação e comunicação (TIC) aplicado no contexto educacional tem se tornado cada vez mais frequente. Diversos países adaptaram suas abordagens educacionais para promover e suportar a utilização dessas tecnologias tanto em cursos presenciais quanto em cursos a distância. No contexto de Educação à distância, o número de usuários tem crescido consideravelmente, o que torna os ambientes mais complexos, tendo em vista que esses irão gerar uma grande quantidade de atividades para serem avaliadas, tornando o processo de avaliação manual uma tarefa árdua ao professor. Esse problema cresce exponencialmente se quisermos incluir avaliações abertas (escritas), uma vez que às suas correções aumentaria ainda mais o trabalho do professor, devido ao fato de que o processamento por máquina deste tipo de avaliação é uma opção trabalhosa e pouco efetiva, o que implica que as correções sejam feitas por eles individualmente, acarretando na sobrecarga de trabalho e conseqüentemente um alto custo para correção da produção textual. Com o crescimento do número de estudantes, um maior número de atividades serão feitas a cada momento e a correção pelos professores manualmente se tornaria rapidamente inviável. Existe uma técnica muito usada na literatura chamada avaliação por pares que possibilita a inclusão de avaliações escritas de tal maneira que não seja necessário aumentar os custos vinculados às correções destas, uma vez que os próprios alunos atuam no processo de correção, o que ganha uma independência e liberdade ao professor. No entanto, a aplicação dessa técnica geralmente gera uma insatisfação no aluno, o que pode levar a um fornecimento de um resultado incorreto e enganoso, o que indica que o aspecto motivacional gera comportamentos inadequados por parte dos alunos, comprometendo a aprendizagem e o sistema de avaliação. Dessa forma, é preciso um mecanismo para engajar os participantes do modelo de avaliação por pares, através da utilização de técnicas que tenham a capacidade de "influenciar" positivamente o estado emocional/cognitivo do aluno. Nesse sentido, esse trabalho propõe um modelo de avaliação por pares juntamente com técnicas de gamificação. Com o modelo proposto é possível unir as vantagens de ambas as técnicas, diminuindo a sobrecarga do professor, economizando seu tempo e esforço para o que realmente é necessário, e conseqüentemente, diminuindo o custo associado as correções, sem o viés da desmotivação dos alunos no processo, o que possibilita a expansão de cursos para centenas ou até mesmo milhares de usuários. Através de experimentos realizados, foi possível concluir que as notas obtidas com o modelo proposto se equipararam às dos professores, utilizando menos tempo para obtê-las e a um custo aproximadamente 72% menor. Além disso, a gamificação influenciou de forma positiva, aumentando em 11.76% o número de cadastros e em 64.28% o número de acessos. Além disso, a quantidade de redações realizadas aumentou 10.53% e a quantidade de redações corrigidas cresceu em 20%.

Palavras-chaves: Avaliação por pares. Gamificação. Ambientes online.

ABSTRACT

In recent years, the use of information and communication technologies (ICT) applied in educational environments has become increasingly usual. Several countries have adapted their educational approaches to promote and support the use of these technologies in both classroom courses as in distance learning courses. In the context of Distance learning, the number of users has grown considerably, making the environments more complex, given that these will generate a lot of activities to be evaluated, making the manual evaluation process a chore to teacher. This problem grows exponentially if we want include open assessments (written), since their corrections would further increase the teacher's work, due to the fact that processing by machine of this type of evaluation is a laborious and less effective option, which implies that corrections are made by them individually, resulting in overwork and consequently high costs for correction of text production. With the increasing number of students, more activities will be made every now and the correction by the teachers manually would quickly become infeasible. There is a technique widely used in the literature called peer assessment that enables the inclusion of the written evaluations in such a way that it is not necessary to increase the costs linked to these corrections, once the students themselves would act in the correction process, gaining independence and free to the teacher. However, the application of this technique usually generates a student dissatisfaction, which can lead to incorrect and misleading results, which indicates that the motivational aspect generates inappropriate behaviour by students, compromising learning and assessment system. Thus, we need a mechanism to engage the participants in the peer assessment model, using techniques that have the ability to "influence" a positive emotional/cognitive state on the student. In this sense, this work proposes an peer assessment model with gamification techniques. With the proposed model, it is possible to unite the advantages of both techniques, reducing the overwork of the teacher, saving their time and effort to what is really necessary, and consequently reducing the cost associated with the corrections without the bias of demotivation of students in process, which allows the expansion of the courses for hundreds or even thousands of users. Through experiments, we concluded that the grades obtained with the proposed model were equivalent to those of teachers, using less time to to obtain them and at a cost approximately 72% less. Furthermore, gamification had a positive impact by increasing the register number in 11.76% and 64.28% in the number of logins. Moreover, the amount of realized activities increased 10.53% and the amount of corrected activities increased by 20%.

Keywords: Peer Assessment. Gamification. Online environments.

LISTA DE ILUSTRAÇÕES

Figura 1 – Número de matrículas em ambientes a distância	17
Figura 2 – Subelementos das categorias de gamificação	30
Figura 3 – STI como uma ferramenta de integração multidisciplinar	34
Figura 4 – Arquitetura tradicional de um sistema tutor inteligente	36
Figura 5 – Ontologias integradas	37
Figura 6 – Tipos de ontologias, de acordo com sua generalidade	38
Figura 7 – Ferramentas existentes para criação de ontologias	40
Figura 8 – Ferramentas existentes para criação de ontologias	41
Figura 9 – Percentual de artigos por base bibliográfica	51
Figura 10 – Fluxo de remoção de artigos da revisão sistemática	52
Figura 11 – Quantidade de artigos por ano de publicação	53
Figura 12 – Tipos dos artigos selecionados para análise	53
Figura 13 – Quantidade de artigos referentes ao contexto aplicado.	55
Figura 14 – Quantidade de artigos referentes ao nível educacional envolvido.	55
Figura 15 – Estatísticas dos artigos observados com relação ao desempenho dos alunos	57
Figura 16 – Estatísticas dos artigos observados com relação ao auxílio aos professores	58
Figura 17 – Estatísticas dos artigos observados com relação às dificuldades envolvidas	59
Figura 18 – Fluxo avaliação em um ambiente educacional + modelo	65
Figura 19 – Atividades referentes a etapa de uso do modelo de avaliação por pares	66
Figura 20 – Fluxo de execução do modelo em detalhes	67
Figura 21 – Etapas do modelo	71
Figura 22 – Elementos de gamificação do MeuTutor	77
Figura 23 – Tela do MeuTutor ao errar uma questão - vídeos recomendados	78
Figura 24 – Árvore do aluno para a disciplina de Matemática	79
Figura 25 – Estudos em grupo - Experiência social do MeuTutor	80
Figura 26 – Compartilhamento das ações do usuário no Facebook	81
Figura 27 – Visão geral da arquitetura do MeuTutor	81
Figura 28 – Diagrama de entidade relacionamento para um problema discursivo	83
Figura 29 – Diagrama de entidade relacionamento para um formulário do sistema	84
Figura 30 – Diagrama de entidade relacionamento para a resposta do usuário	85
Figura 31 – Diagrama de entidade relacionamento para as informações do usuário	86
Figura 32 – Diagrama de entidade relacionamento para gamificação	86
Figura 33 – Diagrama de entidade relacionamento para informações de gamificação do usuário	87
Figura 34 – Classes da Ontologia Peer Assessment.owl	89
Figura 35 – Diagrama de componentes da implementação do modelo proposto	90

Figura 36 – Tela de listagem de temas de redação do MeuTutor	93
Figura 37 – Tela de cadastro de redações do MeuTutor	93
Figura 38 – Tela inicial de redação para um aluno no MeuTutor	94
Figura 39 – Tela proposta de uma redação para um aluno no MeuTutor	95
Figura 40 – Tela onde um aluno corrige a redação de outro	96
Figura 41 – Tela relatório final de um aluno	97
Figura 42 – Formulário disponibilizado ao especialista para avaliação de redação fora do ambiente	110
Figura 43 – Formulário disponibilizado ao especialista para avaliação de redação fora do ambiente - Parte 2	111
Figura 44 – Histograma com a representação das frequências em todas as métricas . . .	112
Figura 45 – Diagrama de caixa com comparativo da métrica nota para os tratamentos T1, T2 e T3	114
Figura 46 – Diagramas de caixa com a diferença da métrica nota nos tratamentos T1 x T2	115
Figura 47 – Histograma para a métrica nota no tratamento T1	116
Figura 48 – Histograma para a métrica nota no tratamento T2	117
Figura 49 – Histograma para a métrica nota no tratamento T3	117
Figura 50 – Histograma para a métrica nota na combinação T1 (azul) x T2 (vermelho) .	118
Figura 51 – Diagrama de caixa com comparativo da métrica tempo para os tratamentos T1, T2 e T3	119
Figura 52 – Histograma para a métrica tempo no tratamento T1	120
Figura 53 – Histograma para a métrica tempo no tratamento T2	121
Figura 54 – Histograma para a métrica tempo no tratamento T3	121
Figura 55 – Diagrama de caixa com comparativo da métrica custo entre os tratamentos T1, T2 e T3	123
Figura 56 – Histograma para a métrica custo no tratamento T1	124
Figura 57 – Histograma para a métrica custo no tratamento 3	124
Figura 58 – Intervalo de confiança da métrica nota para os tratamentos T1, T2 e T3 . .	127
Figura 59 – Intervalo de confiança da métrica tempo para os tratamentos T1, T2 e T3 .	129
Figura 60 – Intervalo de confiança da métrica custo para os tratamentos T1, T2 e T3 . .	131
Figura 61 – Diagrama de caixa com comparativo da métrica Acesso para os tratamentos T1 e T2	146
Figura 62 – Histograma para a métrica acesso (A) no tratamento T1	147
Figura 63 – Histograma para a métrica acesso (A) no tratamento T2	148
Figura 64 – Histograma para a métrica acesso(A) na combinação T1 (azul) x T2 (vermelho)	149
Figura 65 – Diagrama de caixa com comparativo da métrica RF para os tratamentos T1 e T2	150
Figura 66 – Histograma para a métrica Redação Feita (RF) no tratamento T1	150
Figura 67 – Histograma para a métrica Redação Feita (RF) no tratamento T2	151

Figura 68 – Histograma para a métrica Redação Feita (RF) na combinação T1 (azul) x T2 (vermelho)	152
Figura 69 – Diagrama de caixa com comparativo da métrica Redação Corrigida (RC) para os tratamentos T1 e T2	152
Figura 70 – Histograma para a métrica Redação Corrigida (RC) no tratamento T1 . . .	153
Figura 71 – Histograma para a métrica Redação Corrigida (RC) no tratamento T2 . . .	154
Figura 72 – Intervalo de confiança da métrica acesso (A) para os tratamentos T1 e T2 .	155
Figura 73 – Intervalo de confiança da métrica Redações Feitas (RF) para os tratamentos T1 e T2	157
Figura 74 – Intervalo de confiança da métrica Redações Corrigidas (RC) para os tratamentos T1 e T2	159
Figura 75 – Classes da Ontologia Learner.owl	178
Figura 76 – Classes da Ontologia Resource.owl	179
Figura 77 – Classes da Ontologia Gamificação.owl	180
Figura 78 – Proposta de redação criada para a unidade de experimento	187
Figura 79 – Formulário de avaliação criado com base nos parâmetros de correção do ENEM189	

LISTA DE TABELAS

Tabela 1 – Formulário de extração dos dados.	50
Tabela 2 – Resultado da execução da String nas bases científicas selecionadas.	50
Tabela 3 – Quantidade de publicações por países	54
Tabela 4 – Finalidade de uso de avaliação por pares	56
Tabela 5 – Loops de atividades definidos no modelo de gamificação.	75
Tabela 6 – Definição formal das hipóteses de pesquisa.	103
Tabela 7 – Níveis dos fatores.	104
Tabela 8 – Definição dos tratamentos.	106
Tabela 9 – Custo, nota e tempo dos tipos de avaliações (professor ou pares de alunos) e tipos de ambiente (MeuTutor ou Questionário on line).	111
Tabela 10 – Medidas de dispersão das amostras do experimento com pares de alunos. . .	113
Tabela 11 – Sumarização dos dados relativos a variável Nota (N).	118
Tabela 12 – Sumarização dos dados relativos a variável Tempo (T).	122
Tabela 13 – Sumarização dos dados relativos a variável Custo (C).	125
Tabela 14 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Nota	127
Tabela 15 – Resultado da aplicação do T-test com os dados da métrica Nota	128
Tabela 16 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Tempo	129
Tabela 17 – Resultado da aplicação dos testes estatísticos com os dados da métrica Tempo	130
Tabela 18 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Custo	132
Tabela 19 – Resultado da aplicação do teste de Wilcoxon com os dados da métrica Custo	133
Tabela 20 – Níveis dos fatores do experimento de gamificação	140
Tabela 21 – Definição formal das hipóteses de pesquisa do experimento de gamificação .	140
Tabela 22 – Definição dos tratamentos do experimento de gamificação	141
Tabela 23 – Sumarização dos dados relativos a variável Acesso (A).	144
Tabela 24 – Detalhamento da quantidade de alunos que fizeram/corrigiram x redações .	145
Tabela 25 – Sumarização dos dados relativos a variável Acesso (A).	148
Tabela 26 – Sumarização dos dados relativos a variável Redações Feitas (RF).	151
Tabela 27 – Sumarização dos dados relativos a variável Redações Corrigidas (RC). . . .	154
Tabela 28 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Acesso	156
Tabela 29 – Resultado da aplicação dos testes estatísticos com os dados da métrica Acesso	156
Tabela 30 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Redações Feitas (RF)	158

Tabela 31 – Resultado da aplicação dos testes estatísticos com os dados da métrica Redação Feita (RF)	158
Tabela 32 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Redações Corrigidas (RC)	159
Tabela 33 – Resultado da aplicação dos testes estatísticos com os dados da métrica Redação Corrigida (RC)	160

LISTA DE ABREVIATURAS E SIGLAS

- AAAL** Anywhere, Anytime, Anyperson learning/Aprendizagem a qualquer hora em qualquer lugar para qualquer pessoa
- ACM** Association for Computing Machinery/Associação para
- AVA** Ambiente virtual de aprendizagem
- CSCL** Computer-supported collaborative learning/Aprendizagem colaborativa suportada por computador
- DL** Description logic/Lógica de descrição
- DTO** Data transfer object/Objeto de Transferência de Dados
- EAD** Educação a distância
- ENEM** Exame nacional do ensino médio
- ER** Modelo de entidade relacionamento
- GQM** Goal, question, metric/ Objetivos, questões e métricas
- INEP** Instituto Nacional de Estudos e Pesquisas Educacionais
- LPS** Linha de produto de software
- MOOC** Massive open online courses/Cursos massivos online e abertos
- OWL** Web Ontology Language/Linguagem de ontologia para Web
- PHP** Hypertext Preprocessor/Pré-Processador de hiper texto
- RDF-S** Resource Description Framework schema/Esquema de arcabouço de descrição de recursos
- RDF** Resource Description Framework/Arcabouço de descrição de recursos
- SOAP** Simple Object Access protocol/ Protocolo de acesso a objetos simples
- STI** Sistema tutor inteligente
- TIC** Tecnologia da informação e comunicação
- WSDL** Web Services Description Language/ Linguagem de descrição de serviços web
- XML** Extensible Markup Language/ Linguagem de marcação extensível

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Motivação e contextualização do trabalho	16
1.2	Problemática	20
1.3	Objetivos	22
1.4	Escopo	22
1.5	Contribuições do trabalho	23
1.6	Organização da dissertação	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Avaliação por pares	25
2.1.1	Vantagens da avaliação por pares	26
2.1.2	Limitações da avaliação por pares	27
2.2	Gamificação	28
2.2.1	Elementos da gamificação	29
2.2.2	Benefícios da gamificação	31
2.2.3	Criando um plano de gamificação	32
2.3	Sistemas tutores inteligentes	33
2.3.1	Características de um STI	34
2.3.2	Arquitetura dos Sistemas Tutores Inteligentes	35
2.4	Ontologias	37
2.4.1	Benefícios de ontologias	38
2.4.2	OWL e protégé	39
3	TRABALHOS RELACIONADOS	42
3.1	Protocolo	43
3.1.1	Descrição geral	43
3.1.2	Estratégia de busca	45
3.1.3	Seleção de Estudos	47
3.1.3.1	Processo de Seleção dos Estudos Primários	48
3.1.3.2	Resumo da coleta de dados	49
3.2	Execução da revisão sistemática	50
3.3	Análise de dados obtidos	52
3.4	Trabalhos Relacionados	60
4	A PROPOSTA	63
4.1	O modelo de avaliação por pares	63

4.1.1	Fluxo de avaliação em ambientes educacionais online x modelo proposto . . .	63
4.1.2	Modelo proposto	65
4.1.2.1	A gamificação no modelo	72
4.2	Implementação do modelo de avaliação por pares	75
4.2.1	O ambiente educacional MeuTutor	75
4.2.1.1	Arquitetura do MeuTutor ENEM	79
4.2.2	Implementação do modelo	82
4.2.3	Integração do modelo ao ambiente MeuTutor	92
5	VALIDAÇÃO - EXPERIMENTO PROCESSO AVALIAÇÃO PARES	98
5.1	Projeto de experimento	98
5.1.1	Definição do Problema	99
5.1.2	Objetivos da Investigação	101
5.1.3	Planejamento do experimento	101
5.1.3.1	Questões de pesquisa e hipóteses	102
5.1.3.2	Fatores e variáveis de resposta	104
5.1.3.3	Unidades de Experimento	104
5.1.3.4	Seleção do Projeto Experimental	105
5.1.3.5	Plano de execução	105
5.1.3.6	Análise de ameaças à validade	107
5.2	Narrativa de execução do experimento	107
5.3	Análise dos dados	109
5.3.1	Análise preliminar dos dados	111
5.3.2	Análise descritiva dos dados	114
5.3.3	Verificação das hipóteses	125
5.4	Principais conclusões	136
6	VALIDAÇÃO - EXPERIMENTO GAMIFICAÇÃO	137
6.1	Projeto de experimento	137
6.1.1	Planejamento Experimento	138
6.2	Execução do experimento	143
6.3	Análise dos dados	144
6.3.1	Análise preliminar dos dados	144
6.3.2	Análise descritiva dos dados	145
6.3.3	Verificação das hipóteses	154
6.4	Principais conclusões	160
7	CONCLUSÕES E TRABALHOS FUTUROS	162
	REFERÊNCIAS	166

APÊNDICE A – LISTA COM OS ARTIGOS SELECIONADOS PARA REVISÃO SISTEMÁTICA	173
APÊNDICE B – MODELADEM USANDO ONTOLOGIAS	178
APÊNDICE C – EXEMPLO INTERFACE WS PROVIDA PELO MODELO - IWRITINGEVALUATIONWS	181
APÊNDICE D – PROPOSTA DE REDAÇÃO CRIADA PELO ES- PECIALISTA	186
APÊNDICE E – FORMULÁRIO DE AVALIAÇÃO PADRÃO PARA REDAÇÕES DO ENEM	188
APÊNDICE F – ARTIGOS PUBLICADOS/PREVISTOS AO LONGO DO MESTRADO	190

1 INTRODUÇÃO

A aplicação de tecnologias da informação e comunicação (TIC) na educação tem sido cada vez mais destacada e leitmotiv para o processo de ensino-aprendizagem. Nos últimos anos, diversos países têm adaptado suas abordagens educacionais para promover e suportar a utilização de tecnologias tanto em cursos presenciais quanto em cursos a distância. No contexto de Educação a Distância (EAD) - uma modalidade de educação mediada por tecnologias em que discentes e docentes estão separados espacial e/ou temporalmente, ou seja, não estão fisicamente presentes em um ambiente presencial de ensino-aprendizagem (MORAN, 2008) - o uso destas ferramentas se torna de fundamental importância para manter uma comunicação de múltiplas vias entre o aluno e o professor.

Atualmente existem diversas tecnologias que apóiam esse tipo de educação, como por exemplo, os sistemas tutores inteligentes (STIs)(SLEEMAN; BROWN, 1982), os ambientes virtuais de aprendizagem (AVA) (ex: Moodle¹, SOLAR, TelEduc², Blackboard³, Sakai⁴), os ambientes de aprendizagem colaborativa suportada por computador⁵(MIYAKE, 2007) e mais recentemente, os cursos online abertos e massivos⁶(WULF et al., 2014) (ex: EDX⁷, Coursera⁸, Udacity⁹). As necessidades envolvidas nestes ambientes constituem uma mudança no paradigma de aprendizagem, onde passamos a perceber uma descentralização do processo de ensino e aprendizado, antes vinculado às aulas presenciais. Isso vem ocorrendo através de um processo mediado pela tecnologia, que busca permitir o aprendizado a partir de qualquer lugar, a qualquer momento e para qualquer pessoa, AAAL¹⁰ (BITTENCOURT et al., 2008).

1.1 Motivação e contextualização do trabalho

Com o surgimento cada vez maior de ambientes EAD, o número de usuários tem crescido consideravelmente, atingindo o número de 5,8 milhões de matrículas em 2012 no Brasil segundo Censo EAD.BR 2012, divulgado pela Associação Brasileira de Educação a Distância (Abed), de acordo com a figura 1. A grande quantidade de ambientes que surgiram ao longo do tempo tornou possível que qualquer pessoa com uma conexão à internet se inscreva em cursos de nível universitário, geralmente gratuitos, o que explica o

¹ Moodle, disponível em: <https://moodle.org/>

² Teleduc, disponível em: <http://www.teleduc.org.br/>

³ Blackboard, disponível em: <http://www.blackboard.com/>

⁴ Sakai, disponível em: <http://www.sakaiproject.org/>

⁵ Tradução do inglês: Computer-supported collaborative learning - CSCL

⁶ Tradução do inglês: Massive Open Online Course - MOOC

⁷ EDX, disponível em: <https://www.edx.org/>

⁸ Coursera, disponível em: <https://www.coursera.org/>

⁹ Udacity, disponível em: <https://www.udacity.com/>

¹⁰ Sigla em Inglês para: Anywhere, Anytime, Anyone Learning

grande aumento da popularidade dos ambientes de educação online no Brasil e no mundo (EAD).

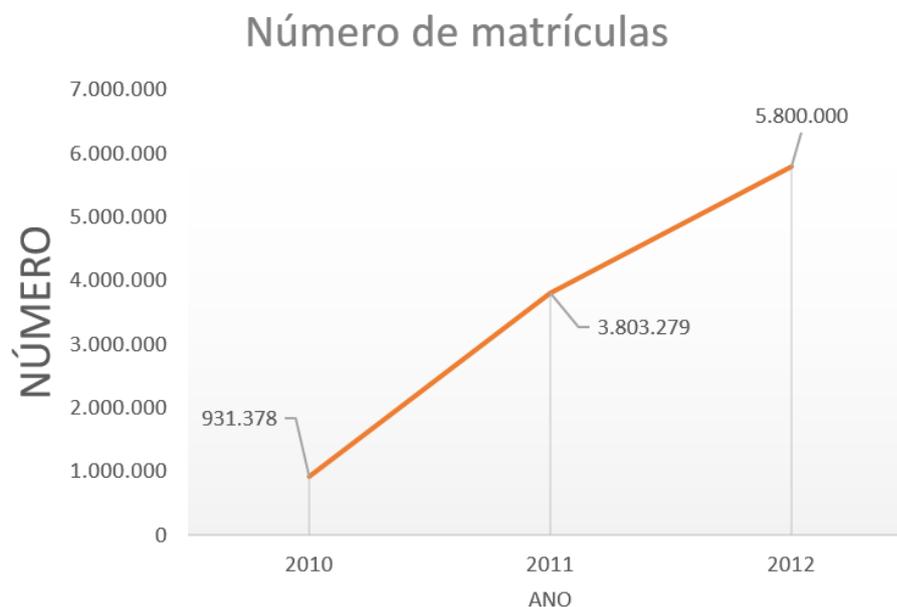


Figura 1 – Número de matrículas em ambientes a distância

Fonte: Censo EAD.BR 2012

Por um lado, o alto número de matrículas nestes ambientes indicam que houve um grande aumento da oferta de cursos online, permitindo um maior “Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento”, uma das metas do documento intitulado “Grandes Desafios da Computação 2006 - 2016” (CARVALHO; LEON et al., 2006), criado pela Sociedade Brasileira de Computação.

Por outro lado, o aumento significativo do número de usuários torna os ambientes mais complexos, uma vez que esses usuários, ao utilizarem os ambientes, irão gerar uma grande quantidade de dados a serem processados. Diante do cenário de milhões de usuários e a grande quantidade de dados disponíveis, a avaliação manual desses dados torna-se uma tarefa árdua ao professor/tutor destes ambientes. Neste sentido, torna-se fundamental o uso de TICs que tenham a capacidade de processar estes dados de forma independente, bem como apoiar o professor no processo de construção do conhecimento de forma a atender as necessidades individuais de cada aluno.

Com isso, enquanto novas tecnologias da Web permitem formas escaláveis para fornecer conteúdo de vídeo conferência, implementar fóruns sociais e acompanhar o progresso do aluno nestes ambientes, continuamos limitados em nossa capacidade de avaliar e dar feedback para trabalhos complexos dos estudantes e muitas vezes abertos (avaliações escritas), tais como provas matemáticas, projeto de problemas e redações (PIECH et al., 2013).

Diante de tal dificuldade, estes ambientes educacionais online frequentemente oferecem soluções mais simples aos seus usuários, como por exemplo, questões objetivas como questões de múltipla escolha, verdadeiro ou falso, preencher lacunas ou relacionar colunas, uma vez que este tipo de problema pode ser processado de forma automática pelo computador. Porém, no contexto de avaliações escritas aplicadas nestes ambientes a distância, a tarefa de personalização da aprendizagem geraria uma grande sobrecarga no professor.

Isso é ocasionado uma vez que o processamento por máquina deste tipo de avaliação se torna uma opção trabalhosa (em contraste com o processamento de provas com questões de múltipla escolha, por exemplo), implicando que as correções destas atividades sejam feitas pelos professores individualmente, o que acarreta um alto custo para correção da produção textual. Com o crescimento do número de estudantes nestes ambientes e consequentemente, um maior número de atividades sendo feitas a cada momento, a correção delas pelos professores/tutores se tornaria rapidamente inviável.

Para suprir este problema pode-se utilizar um conceito chamado de avaliação por pares, do inglês "peer assessment". Avaliação por pares é um processo pelo qual os alunos ou seus pares atribuem notas ou testes com base em parâmetros de referência pré-definidos pelo professor (SADLER; GOOD, 2006). A prática é utilizada para salvar o tempo dos professores e melhorar a compreensão dos estudantes a respeito dos materiais do curso, bem como melhorar as suas habilidades metacognitivas (MALEHORN, 1994), aumentando a sua produtividade.

Os Modelos de Avaliação por pares - que historicamente tem sido usados para logística, melhorar habilidades pedagógicas, metacognitivas e, até mesmo, benefícios afetivos ((SADLER; GOOD, 2006)) - oferecem uma solução promissora que pode escalar a classificação de serviços complexos em cursos com dezenas ou até mesmo centenas de milhares de estudantes, permitindo a correção das atividades, sem necessariamente aumentar os custos. O seu uso traz diversos benefícios aos alunos, tais como: aumento da responsabilidade e autonomia; faz com que os alunos se esforcem para uma compreensão mais avançada e profunda do assunto, competências e processos; traz o aluno para o papel ativo na aprendizagem; envolvem os alunos na reflexão crítica e desenvolve nos alunos uma melhor compreensão de sua própria subjetividade e julgamento(READING, 2014).

No contexto de ambientes de EAD, o sistema gerencia as correções feitas entre os próprios estudantes, isto é, os próprios alunos corrigem as atividades escritas de seus pares. Existem diversas abordagens de avaliação por pares publicados na literatura, como as abordagens formativas (ORSMOND*; MERRY; CALLAGHAN, 2004), modelos probabilísticos (PIECH et al., 2013) e até mesmo modelos usando redes bayesianas (WANG; VASSILEVA, 2003). Esta técnica aplicada neste contexto é especialmente promissora como desestabilizadora do papel passivo do aluno, de forma que este assuma responsabilidade por sua aprendizagem, buscando o aprimoramento do processo de aprendizagem através de sua participação ativa.

Como consequência, o uso desta técnica resulta em diversos benefícios educacionais, como um aumento da aprendizagem do aluno, e principalmente a redução de sobrecarga do professor (SADLER; GOOD, 2006). No entanto, o uso de modelos de avaliação por pares aplicados em ambientes educacionais online com o propósito de correções de atividades escritas ainda é limitado e um dos grandes problemas encontrados ao utilizá-la é a presença de comportamentos inadequados por parte dos alunos que comprometem a aprendizagem e o sistema de avaliação. O surgimento deste tipo de atitude ocorre por diversos fatores cognitivos e emocionais como tédio, desmotivação e necessidade de obter resultados rapidamente (BAKER et al., 2008). Para resolver este problema é necessário a utilização de técnicas e modelos que tenham a capacidade de "influenciar" positivamente o estado emocional/cognitivo do aluno, sem necessariamente aumentar o custo de implantação.

Neste sentido, é necessário incorporar no modelo de avaliação por pares uma técnica que tem este objetivo - a gamificação. Gamificação é o uso de mecânica, ideias e estética de jogos (e.x. contexto, feedback rápido, competição, fases, conquistas, pontos e etc.), para engajar pessoas, motivar ações, promover aprendizado e solucionar problemas (KAPP, 2012). Este termo é comumente utilizado para expressar o uso de elementos de jogos (enredo, pontuação, níveis, missões, troféus e rankings) em ambientes que não são de jogos (ambiente educacionais) para motivar ou influenciar as pessoas a realizarem uma determinada atividade.

Esta técnica foi escolhida devido a sua tendência de possuir um custo mais baixo de implementação se comparado a outras técnicas como, por exemplo, jogos. Isso acontece porque apenas alguns elementos de jogos são incluídos. Além disso, alguns trabalhos como apresentados em (CUNHA; GASPARINI; BERKENBROCK, 2013a) indicam que gamificação traz diversos benefícios em ambientes de aprendizagem colaborativa. Com isso, para incluir um suporte a avaliações discursivas de forma eficiente nesses ambientes, se torna necessário a criação de um modelo de avaliações por pares juntamente com técnicas de gamificação que poderá ser integrado nesses ambientes de educação online, possibilitando a inclusão de avaliações escritas, sem necessariamente aumentar os custos envolvidos, beneficiando tanto os professores quando os alunos envolvidos no processo.

O trabalho proposto supõe que o uso combinado de técnicas de avaliação por pares juntamente com gamificação torna o modelo proposto mais poderoso e completo, evitando e/ou diminuindo a presença de comportamentos inadequados por parte dos alunos que comprometem a aprendizagem e o sistema de avaliação. Os elementos de gamificação aplicados no modelo tendem a influenciar positivamente o estado do aluno, incentivando-os a participar do processo de ensino-aprendizagem proposto através das recompensas obtidas (pontos, níveis, troféus, entre outros).

1.2 Problemática

Os ambientes de aprendizagem são instrumentos que tentam garantir uma aprendizagem significativa, tendo como objetivos claros a utilização dos mesmos por parte de instituições que oferecem educação a distância ou ensino online (MESSA, 2010). O uso de soluções tecnológicas vem auxiliando o processo de ensino-aprendizagem e em consequência na melhoria da qualidade do ensino. Com isso, o número de cursos oferecidos a distância vem aumentando significativamente.

Porém, com esse crescimento do número de usuários, a quantidade de dados disponíveis (avaliações respondidas) nestes ambientes cresce exponencialmente. Assim, para garantir respostas rápidas e que realmente ajudem o aluno em suas necessidades, o trabalho do professor/tutor aumenta gradativamente a medida que o número de usuários cresce, se tornando rapidamente inviável (sobrecarregando suas atividades). Neste sentido, torna-se fundamental o uso de tecnologias que tenham a capacidade de processar estes dados de forma independente, bem como apoiar o professor no processo de construção do conhecimento de forma a atender as necessidades individuais de cada aluno.

Esta grande quantidade de dados vindos das interações dos usuários nos ambientes de aprendizagem e as formas de analisar, processar e enviar informações úteis aos alunos para sua aprendizagem, geradas a partir desses dados de maneira simples é o problema que identificamos. Para que se possa obter informações relevantes das interações dos usuários que utilizam esses ambientes é preciso que o professor seja capaz de analisar os dados (respostas das atividades) individualmente de cada aluno de forma manual. No entanto, a grande quantidade de dados torna impossível que isso seja realizado manualmente, e em tempo hábil de oferecer o resultado desejado.

Ao se trabalhar com avaliações escritas aplicadas nestes ambientes a distância, a tarefa de correção manual gera uma sobrecarga no professor, pois, diferente de outras atividades tradicionais (questões de múltipla escolha), as atividades escritas não podem ser corrigidas de uma maneira automática, implicando que as correções destas atividades sejam feitas pelos professores individualmente, acarretando em um alto custo para correção da produção textual.

A solução mais simples seria aumentar o número de professores/tutores nestes ambientes e conseqüentemente diminuir o problema, porém o número de usuários tende a crescer exponencialmente. No Brasil, os ambientes online que tem o mesmo propósito utilizam os métodos tradicionais para correção das atividades e assim, na tentativa de reduzir os custos, limitam o número de atividades que o estudante pode fazer no ambiente.

Com base nesses questionamentos, surge a nossa primeira pergunta.

Q1- Como podemos incluir avaliações escritas em ambientes educacionais online de uma maneira eficiente?

Este é o era o problema inicial do trabalho proposto. O fato de incluirmos avaliações escritas nestes ambientes nos traz uma série de problemas, como o alto custo da correção

textual e uma maior sobrecarga do professor, devido à atribuição de novas responsabilidades a ele. Neste sentido, é preciso a utilização de algum mecanismo que possibilite esta inclusão de tal maneira que não seja necessário aumentar o custo vinculado às correções das redações e que estas correções sejam feitas de maneira mais automática e independente do número de alunos no ambiente de aprendizagem. Levando isto em consideração, é essencial que haja algum nível de automatização de modo a processar os dados das avaliações.

Neste sentido, a solução proposta visa criar um mecanismo de avaliação por pares e aplicá-lo no contexto de provas discursivas nestes ambientes. Usando este conceito em um ambiente educacional EAD, o sistema ficará responsável pelo gerenciamento das correções que serão feitas de uma maneira semi-automatizada pelos próprios estudantes, isto é, os próprios alunos corrigem as atividades escritas de seus pares.

Com isso, o ambiente permitirá que seus alunos estudem e pratiquem através de avaliações escritas no próprio ambiente. Usando avaliação por pares é possível desatrelar o custo das correções do número de usuários, e sem necessariamente aumentar o trabalho do professor, diminuindo os problemas citados anteriormente.

O uso da técnica tem benefícios educacionais e reduz a sobrecarga do professor. Porém, um dos grandes problemas encontrados ao utilizá-la é a presença de comportamentos inadequados por parte dos alunos, que comprometem a aprendizagem e o sistema de avaliação. Esse tipo de comportamento está associado ao fator motivacional dos alunos que participam do processo. Neste sentido, surge o nosso principal problema de pesquisa.

Q2- Como engajar e motivar a participação dos alunos dentro do modelo de avaliação por pares?

O fato de atribuir novas responsabilidades ao aluno ocasiona um comportamento inadequado ou não esperado pelo sistema. O surgimento deste tipo de atitude ocorre por diversos fatores cognitivos e emocionais, como por exemplo, tédio, desmotivação e necessidade de obter resultados rápidos (BAKER et al., 2008).

Neste sentido, é necessário a utilização de técnicas que tenham a capacidade de "influenciar" positivamente o estado emocional/cognitivo do aluno, levando-o a agir de uma maneira mais proveitosa a si mesmo e ao modelo de avaliação do sistema. Neste sentido, a solução proposta utiliza uma técnica ainda pouco explorada na educação que tem este objetivo - a gamificação.

Com a gamificação, é possível fazer o uso da mecânica, ideias e estética de jogos (contexto, feedback, competição, fases, pontos) para influenciar o aluno, com o objetivo de engajá-lo, motivando ações benéficas, promovendo seu aprendizado e sua capacidade de analisar e solucionar problemas.

Nesse sentido, a solução proposta visa resolver as duas questões apresentadas, solucionando o problema de negócio, envolvendo o uso de avaliação por pares junto com técnicas de gamificação. Esses elementos foram escolhidos devido aos seus benefícios. Na seção

seguinte definiremos os objetivos dessa dissertação.

1.3 Objetivos

O grande objetivo deste trabalho (problema de negócio) é **possibilitar a inclusão de avaliações discursivas em ambientes de educação online de maneira eficiente**, onde não seja necessário aumentar o custo vinculado as correções das atividades, sendo estas feitas através de um processo semi-automatizado (onde os próprios alunos participam do processo) e independente do número de alunos no ambiente de aprendizagem (número que tende a aumentar significativamente). Com isso, o ambiente permitirá que seus alunos estudem e pratiquem os conteúdos através deste tipo de questão, tornando o ambiente mais flexível.

Os objetivos específicos deste trabalho são:

- Realizar um levantamento do estado da arte no que diz respeito ao uso de avaliações por pares juntamente com técnicas de gamificação aplicadas no contexto de avaliações escritas em ambientes educacionais online;
- Discutir e propor um modelo de avaliação por pares flexível que poderá ser utilizado em ambientes de educação a distância (EAD);
- Implementar o modelo proposto e integrá-lo em um ambiente educacional, a fim de que se possa analisar seu comportamento em ação com usuários reais do ambiente;
- Avaliar o modelo proposto e implementado, apresentando as suas vantagens e desvantagens, bem como seus pontos fortes, limitações e possíveis trabalhos futuros.

1.4 Escopo

O trabalho apresentado ao longo dessa dissertação trata sobre a proposição, criação e avaliação de um modelo de avaliação por pares juntamente com técnicas de gamificação, a ser aplicado no contexto de avaliações escritas (questões discursivas, redações, entre outras aplicações) em ambientes educacionais online. Situa-se na linha de pesquisa de Engenharia de Sistemas Computacionais do Mestrado em Informática, do Instituto de Computação da Universidade Federal de Alagoas.

Esta linha de pesquisa trata da aplicação e desenvolvimento de modelos, algoritmos, métodos numéricos e/ou métodos formais para o tratamento de problemas relacionados a sistemas computacionais. O trabalho desenvolvido se caracteriza como interdisciplinar, envolvendo a integração entre diversas áreas da computação - arquiteturas peer-to-peer, trabalho cooperativo suportado por computador e aprendizagem colaborativa (segundo

a classificação da ACM (ACM, 2012)) - e educação, promovendo uma integração especialmente entre engenharia de software, modelagem de conhecimento e informática na educação.

Esta dissertação visa auxiliar professores e tutores de cursos mediados por computador, apresentando um modelo eficaz que poderá ser utilizado no contexto de avaliações escritas nestes ambientes de educação online, evitando uma possível sobrecarga em seu trabalho, além de possibilitar mais agilidade nas correções e promover uma maior aprendizagem ao aluno, tornando-o mais ativo em seu processo de aprendizagem.

Como apresentado, o trabalho apresenta a construção de um modelo de avaliação por pares eficiente com a finalidade de incluir avaliações discursivas em ambientes de aprendizagem online. Esse modelo, como será detalhado nos próximos capítulos, é apenas parte de um processo já existente nas avaliações dos estudantes nesses ambientes.

Com isso, algumas atividades como preparação do ambiente (cadastro de atividades), identificação das necessidades dos alunos e criação de um plano de avaliação, apesar de serem citadas nesse trabalho, não estão incluídas em nossa proposta. Além disso, algumas das etapas do modelo, como a seleção dos usuários para a correção e o cálculo da reputação e confiança foram criadas e construídas baseadas em trabalhos semelhantes existentes.

Dessa forma, o foco desse trabalho será na criação do modelo de avaliação por pares, sua implementação usando técnicas de engenharia de software adequadas e sua inclusão nos ambientes educacionais online.

1.5 Contribuições do trabalho

O presente trabalho contribui com a comunidade de Informática na Educação, apresentando um modelo de avaliação por pares com gamificação que poderá ser utilizado/incorporado em ambientes educacionais à distância, permitindo a inclusão de avaliações escritas de forma viável, e trazendo diversos benefícios ao professor/tutor e aos alunos envolvidos nestes ambientes, conforme apresentados nas seções anteriores.

Além da apresentação do modelo, outra contribuição do trabalho é a sua implementação e futura disponibilização para ser facilmente integrado aos ambientes educacionais. Essa implementação também foi integrada em um ambiente educacional. Alguns experimentos também foram executados para validar a eficácia do modelo proposto.

Além disso, outra contribuição do trabalho é a apresentação de uma revisão sistemática sobre o uso de avaliações por pares em ambientes educacionais online. Esta revisão teve como objetivo levantar os trabalhos já existentes na área. Neste sentido, as contribuições científicas deste trabalho poderão servir como base para novas pesquisas nas áreas envolvidas como arquiteturas peer-to-peer, trabalho cooperativo suportado por computador e aprendizagem colaborativa.

Ainda como contribuição do trabalho, alguns artefatos gerados ao longo do trabalho como os modelos de entidade relacionamentos apresentados, e suas implementações em

ontologias, principalmente a ontologia de Avaliação por pares criada, são considerados como contribuições secundárias do trabalho e podem ser utilizadas em outros projetos com outros propósitos.

1.6 Organização da dissertação

A dissertação aqui apresentada contém sete capítulos organizados e distribuídos da seguinte maneira:

- Capítulo 1: Neste capítulo são apresentados uma motivação, contextualização, problemática, objetivos e as contribuições do trabalho aqui proposto.
- Capítulo 2: São apresentados neste capítulo a fundamentação teórica que aborda os principais conhecimentos utilizados nessa dissertação, enfatizando as técnicas de avaliação por pares e gamificação. Além disso, são apresentados os conceitos de sistemas tutores inteligentes e ontologias.
- Capítulo 3: Apresenta de forma detalhada um estudo dos trabalhos relacionados através de uma revisão sistemática de literatura. Aqui serão apresentados desde o protocolo utilizado para a revisão, bem como o processo de obtenção dos artigos, extração dos dados e apresentação e análise dos resultados.
- Capítulo 4: Este capítulo apresenta, de forma detalhada o modelo de avaliação por pares proposto, bem como seu processo de implementação em um ambiente educacional escolhido.
- Capítulo 5: Finalmente, este capítulo apresenta o planejamento e o processo de execução de um experimento com o objetivo de avaliar o modelo proposto do ponto de vista do processo de avaliação por pares, mostrando a aplicação da proposta em um cenário real, e uma discussão acerca dos resultados alcançados.
- Capítulo 6: Esse capítulo mostra o planejamento e execução de um experimento com o objetivo de avaliar o impacto da gamificação aplicado no modelo apresentado.
- Capítulo 7: Por fim, este capítulo apresenta as conclusões acerca do trabalho apresentado, bem como alguns ramos de trabalhos futuros a serem seguidos.
- No final, serão apresentados as referências utilizadas na elaboração do trabalho e os apêndices.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo serão abordados os principais tópicos contendo os fundamentos teóricos dessa dissertação, necessários para uma apropriada compreensão da mesma. A fundamentação teórica apresentada ao longo desse capítulo servirá como uma base teórica para a análise e compreensão dos elementos do modelo apresentado, bem como o entendimento completo de sua implementação em um ambiente e processo de sua validação.

Nesse sentido, as seções foram distribuídas da forma como segue. Em primeiro momento, será apresentado os conceitos de avaliação por pares assim como as vantagens em seu uso na seção 2.1. Em seguida, na seção 2.2, será explicitado e detalhado o conceito de gamificação, os seus benefícios e os principais elementos da gamificação que serão usados de alguma maneira nesse trabalho. Além disso, será apresentado uma abordagem para a construção de um plano de gamificação. Finalmente, para uma melhor compreensão da implementação do modelo serão apresentados de forma breve os conceitos de sistemas tutores inteligentes na seção 2.3 e ontologias na seção 2.4.

2.1 Avaliação por pares

A avaliação por pares é um tipo de avaliação de alunos feita pelos próprios estudantes, onde são fornecidos feedback formativos e/ou classificação sumativa. Permite que os instrutores compartilhem a avaliação dos trabalhos com seus alunos. É uma forma de avaliação inovadora que visa melhorar a qualidade da aprendizagem e capacitar alunos (MCDOWELL; MOWL; GIBBS, 1995). Pode incluir o envolvimento do aluno, não só nas decisões finais acerca dos trabalhos feitos, mas também na definição prévia de critérios de avaliação e na seleção/criação de trabalhos e avaliações (BIGGS; TANG, 2011).

O objetivo primordial da avaliação por pares é fornecer feedback aos alunos e a qualidade mais significativa da avaliação de pares é que ela é poderosa (TOPPING, 2009). Os estudantes, ao comentar e julgar os trabalhos dos colegas, desempenham um papel vital na avaliação formativa, atingindo um dos resultados desejáveis da educação em geral, que é tornar maior a capacidade do aluno de fazer julgamentos.

Os Modelos de Avaliação por pares tem sido usados para logística, melhorar habilidades pedagógicas, metacognitivas e, até mesmo, benefícios afetivos (SADLER; GOOD, 2006). Relatórios dos tipos de avaliação em que a avaliação pelos pares é utilizada para fins sumativos incluem redação, habilidades clínicas, palestras e apresentações orais, projetos arquitetônicos, habilidades interpessoais, fotografia e atividades em pequenos grupos (FALCHIKOV, 1986).

A avaliação por pares envolve os alunos tomando a responsabilidade de avaliar o trabalho de seus colegas em relação aos critérios de avaliação definidos. Eles podem, portanto,

ser comprometidos em fornecer feedback aos seus pares. É uma maneira poderosa para os alunos atuarem como 'professor' e ganhar uma oportunidade de entender melhor os critérios de avaliação. Também transfere um certo domínio sobre o processo de avaliação para eles, assim, potencialmente aumentando a sua motivação e engajamento.

Em um processo com avaliação por pares, é muito importante que o avaliador (um aluno em questão) conheça bem os motivos e a atividade que será avaliada, de modo que ele consiga definir os objetivos da avaliação. Uma vez definido esses objetivos, o próximo passo é a definição dos critérios de avaliação, estes que deverão ser construídos de forma clara e simples.

É possível usar a avaliação por pares para avaliar os esforços e as contribuições para o trabalho em grupo através de uma ampla variedade de atividades individuais. É possível projetar a avaliação por pares para ser feita abertamente, incentivando a comparação e discussão, ou anonimamente, dependendo da tarefa de avaliação e de contexto. O ponto-chave é garantir que os participantes (tanto os alunos e seus colegas) entendam o propósito da avaliação de pares e o que se espera deles.

A avaliação por pares também promove a aprendizagem ao longo da vida, ajudando os alunos a avaliar a si próprio e os seus pares as realizações de forma realista, e não apenas incentivando-os sempre que contar com professor. Do ponto de vista computacional, oferecem uma solução promissora que pode escalar a classificação de serviços complexos em cursos com dezenas ou até mesmo centenas de milhares de estudantes, permitindo a correção das atividades, sem necessariamente aumentar os custos da tecnologia computacional.

2.1.1 Vantagens da avaliação por pares

São muitas as afirmações feitas a respeito de avaliação por pares, porém apenas na literatura atual há algumas pesquisas mais detalhadas sobre quais componentes específicos de avaliação por pares beneficia o aluno e em que medida. Estes benefícios são principalmente relacionados com os alunos cada vez mais envolvidos e um aumento na capacidade de assumir mais responsabilidade pela sua própria aprendizagem, isto é, torna-se mais ativo no processo de aprendizagem.

Os principais pontos fortes da avaliação por pares são a melhora da motivação, a promoção ao uso da linguagem da avaliação e a ajuda na melhora da comunicação sobre a aprendizagem entre professor e aluno. De forma menos convincente, os alunos são mais propensos a aceitar a crítica do seu trabalho quando se trata de um dos colegas, em vez de um professor (BLACK et al., 2003).

A avaliação por pares também pode trazer benefícios fundamentais a todos os intervenientes no processo de ensino/aprendizagem. Neste sentido é considerado que, com o uso desta prática avaliativa, ganha-se os seguintes benefícios (BLACK; WILIAM, 2009):

1. Os alunos aprendem a avaliar o seu trabalho e o trabalho dos outros, desenvolvendo sua capacidade de julgamento;
2. Os alunos começam a desenvolver hábitos e capacidades de colaboração nas aprendizagens;
3. Os alunos tornam-se participantes, e não "vítimas" no processo de avaliação.

Com isso, é possível afirmar que a avaliação por pares é um instrumento poderoso de avaliação educacional, permitindo alcançar objetivos pedagógicos diversos. Uma outra vantagem, especialmente quando realizadas no âmbito de critérios de avaliação externos genéricos, é uma compreensão mais profunda dos critérios de avaliação. Além disso, o seu uso traz diversos outros benefícios aos alunos, tais como: aumento da responsabilidade e autonomia; faz com que os alunos se esforcem para uma compreensão mais avançada e profunda do assunto, competências e processos; traz o aluno para o papel ativo na aprendizagem; envolvem os alunos na reflexão crítica e desenvolve nos alunos uma melhor compreensão de sua própria subjetividade e julgamento (READING, 2014).

Por outro lado, o uso de avaliação por pares também traz benefícios ao professor, como por exemplo, uma diminuição da sobrecarga de trabalho, uma vez que as correções das atividades serão feitas pelos próprios alunos, cabendo ao professor apenas gerenciar esse processo. A revisão por pares também pode ser usada como uma forma de ajudar os instrutores a melhorar o ensino e aprendizagem em seus cursos.

2.1.2 Limitações da avaliação por pares

A técnica de avaliação por pares tem sido muito utilizada nos diversos contextos apresentados. A revisão por pares é tida pela grande maioria dos pesquisadores como o mecanismo mais efetivo e eficaz para garantir a qualidade, confiabilidade, integridade e consistência da literatura acadêmica. As limitações e falhas do processo, principalmente em relação à fraude e plágio, embora reconhecidas, não diminuem seu largo emprego, mesmo porque não se conhece outro método mais eficiente (MULLIGAN; HALL; RAPHAEL, 2013).

Um detalhado estudo conduzido com 4.037 pesquisadores em todo o mundo com reconhecida atividade como revisores, apontou que a grande maioria considera a revisão por pares essencial à comunicação científica, e nove entre dez afirmaram que o processo adiciona qualidade a seus próprios manuscritos, sendo o duplo-cego o sistema considerado mais efetivo por 76% dos respondentes. O sistema simples-cego também foi considerado efetivo por 45% dos entrevistados, e 47% apoiam a revisão pós-publicação como forma de complementar o processo (MULLIGAN; HALL; RAPHAEL, 2013).

Entretanto, o processo de revisão pelos pares não protege contra má-conduta. Ele pode identificar erros, mas assenta na honestidade dos autores e, por isso, pode falhar

no reconhecimento de investigação fraudulenta deliberada. Com o objectivo de reduzir a ocorrência destas situações, são várias as organizações que têm vindo a produzir Boas Práticas de investigação (SCOTT, 2007).

2.2 Gamificação

A técnica de gamificação está fortemente ligada a capacidade emocional dos participantes envolvidos. A cognição, a memória e a capacidade de tomar decisões do indivíduo estão intrinsicamente ligados às emoções (CYTOWIC, 2002). Ou seja, estudantes que se sentem ansiosos, chateados ou deprimidos não assimilam as informações corretamente e, por causa disso, acabam apresentando comportamentos impróprios que dificultam a aprendizagem (BAKER et al., 2008). Em contrapartida, alunos que se sentem motivados, desafiados e intrigados tendem a obter melhores resultados (VANLEHN et al., 2011).

Ao se sentirem ameaçados, como por exemplo, quando submetidos a necessidade de alcançar resultados rápidos, ou até mesmo por fatores comportamentais do aluno, como desmotivação, falta de interesse ou tédio, os alunos começam a ter alguns comportamentos inadequados que comprometem a aprendizagem. Esses comportamentos são conhecidos como "gaming the system", expressão utilizada para o ato de trapacear o sistema, ou seja, externalizar comportamentos inadequados no sistema para apenas cumprir uma tarefa exigida (ANDRADE et al., 2013), interferindo diretamente no uso adequado desses sistemas.

A partir desses estudos, percebe-se que se torna necessário o uso de técnicas que tentem evitar estes comportamentos inadequados e ao mesmo tempo motive, desafie e intrigue os alunos, instigando-os a terem melhores resultados. Uma das técnicas que tem esse objetivo é a gamificação. Durante os últimos anos, a gamificação tem sido um tópico de tendências e um assunto muito discutido, como forma de apoiar o envolvimento dos usuários e melhorar os padrões positivos na utilização de um serviço (DETERDING et al., 2011) (HUOTARI; HAMARI, 2012), obtendo benefícios tais como o aumento da atividade do usuário, de sua interação social e melhorando a qualidade e produtividade de suas ações no sistema (HAMARI; LEHDONVIRTA, 2010). Na definição de gamificação a seguir, é destacado o seu papel em invocar as mesmas experiências psicológicas como jogos (geralmente) fazem (HUOTARI; HAMARI, 2012).

Gamificação é o uso de mecânica, ideias e estética de jogos (como por exemplo, contexto, feedback rápido, competição, fases, conquistas, pontos e etc.), para engajar pessoas, motivar ações, promover aprendizado e solucionar problemas (KAPP, 2012). Vale ressaltar que o termo gamificação não significa "aprender por meio de jogos", umas das interpretações de forma errada. O uso correto deste termo refere-se apenas à utilização de elementos e técnicas de design de jogos em situações fora do contexto de jogos, a fim de obter **maior participação e envolvimento** das pessoas em um determinado assunto ou

contexto. Essa abordagem pode então ser aplicada em ambientes empresariais, escolas, administração pública e até em atividades do cotidiano (ANDRADE et al., 2013).

Estudos sobre o uso de jogos e elementos de jogos em educação vêm sendo realizados há algumas décadas (MALONE; LEPPER, 1987) (CORDOVA; LEPPER, 1996) (GEE, 2003) (SHAFFER, 2006) (KLOPFER, 2008) , porém nos últimos anos o interesse pelo tema vem aumentando em ritmo acelerado (KAPP, 2012). A gamificação no contexto educacional tem como principal objetivo o aumento do engajamento dos usuários, através do uso de técnicas semelhantes às usadas em jogos como placares e feedbacks (FLATLA et al., 2011), fazendo com que os usuários se sintam no controle de suas ações e se motivem com as tarefas de aprendizagem a serem cumpridas (PAVLUS, 2010). O uso correto destes elementos determina o sucesso ou o fracasso de um jogo ou de um ambiente gamificado (KAPP, 2012).

2.2.1 Elementos da gamificação

A gamificação engloba diversos elementos que tornam possível simular os aspectos de um jogo ¹ em um ambiente diferente de um jogo. O termo jogo é bastante ambíguo e apresenta diversos significados nas mais diversas aplicações. Um jogo pode ser definido como um sistema no qual os jogadores se envolvem em um conflito artificial, definido por regras e que possui um resultado quantificável (SALEN; ZIMMERMAN, 2004). Segundo esta definição, um jogo possui três elementos principais: desafio, as regras e o resultado (feedback) de suas ações.

Contudo, há diversos outros elementos dentro de um jogo. Podemos classificar os elementos da gamificação em quatro categorias: Mecânica, Histórias, Estética e Tecnologia (SCHELL, 2014). Estas categorias são:

- **Mecânica:** Tem como objetivo definir as regras e ações (procedimentos) do jogo que levarão os jogadores a atingir o seu principal objetivo
- **História:** Descreve o enredo, isto é, a sequência de eventos (acontecimentos) que ocorrerão durante o jogo, podendo ser simples e linear ou complexa e cheia de detalhes
- **Estética:** Descreve as características audiovisuais do jogo, tais como: seu visual, o design, trilhas sonoras, sons de ações e os demais recursos gráficos presentes que irão afetar diretamente a experiência do jogador durante o jogo
- **Tecnologia:** Define os materiais a serem utilizados e as interações que tornarão a interatividade com o jogo (i.e. jogabilidade) possível

¹ Do inglês Game

Cada categoria descrita acima possui uma serie de subelementos envolvidos para que se torne possível a construção de um jogo complexo. A figura 2 apresenta de forma resumida alguns desses subelementos.



Figura 2 – Subelementos das categorias de gamificação

Fonte: (ANDRADE et al., 2013)

Alguns desses elementos (os mais importantes) são detalhados abaixo (KAPP, 2012):

- **Objetivos:** São as metas que um jogador deverá alcançar ao longo do jogo. Estas metas devem ser bastante claras e quantificáveis, sem margem para ambiguidade.
- **Enredo:** É a história do jogo propriamente dita. Proporciona relevância e significado para as ações dos jogadores, levando-os a um mundo novo.
- **Regras:** Definem os comportamentos aceitáveis para o jogo como um todo, isto é, definem desde o número de jogadores, por exemplo, até os meios para alcançar os objetivos do jogo. Tudo o que pode ou não pode ser feito no jogo deverá ser definido por uma regra. Estas podem ser de quatro tipos diferentes: operacional, fundamental/constitutiva, implícita e/ou instrutiva.
- **Feedback:** Tem como objetivo levar o jogador a ter comportamentos, pensamentos e/ou ações corretas e desejadas. É um recurso típico de qualquer jogo e tem ação rápida e imediata.
- **Recompensa:** É um dos principais elementos da gamificação, uma vez que estão diretamente ligadas a motivação dos jogadores. Possui um papel importante na estrutura do jogo e fornece um incentivo valioso aos jogadores, levando-os a um comportamento desejado.

Dentre as recompensas que podem ser alcançadas pelos jogadores, podemos destacar algumas que são bastante utilizadas. Alguns desses principais elementos que podemos destacar são (ZICHERMANN; CUNNINGHAM, 2011):

- **Sistema de pontos:** De acordo com as tarefas que o usuário realiza, o mesmo é recompensado com uma quantidade determinada de pontos.
- **Níveis:** tem como objetivo mostrar ao usuário seu progresso dentro do sistema, geralmente é utilizado em conjunto com os pontos.
- **Rankings:** uma maneira de visualizar o progresso dos outros usuários e criar um senso de competição dentro do sistema.
- **Medalhas/Conquistas:** elementos gráficos que o usuário recebe por realizar tarefas específicas (geralmente tarefas mais direcionadas aos objetivos)
- **Desafios e missões:** tarefas específicas que o usuário deve realizar dentro de um sistema, sendo recompensado de alguma maneira por isso (pontos e medalhas). Cria um senso de desafio para o usuário do sistema

É importante destacar que estes elementos apresentados nesta seção são apenas alguns dos principais elementos que podem compor um jogo. Um jogo não necessariamente compõe todos esses elementos e, portanto, um ambiente gamificado não deve possuir todos os elementos. Cada elemento apresentado aqui apresenta vantagens/desvantagens e podem ser utilizados na criação de uma situação mais motivadora.

Vale ressaltar também que utilizar Gamificação não significa, apenas, inserir elementos de jogos de maneira aleatória em qualquer tipo de situação. Para que se obtenha sucesso ao gamificar um ambiente, deve-se realizar uma análise e uma modelagem do impacto de cada elemento e propor técnicas e mudanças nas arquiteturas atuais, de forma que se possa utilizá-las de maneira adequada, inteligente e eficaz (ANDRADE et al., 2013).

2.2.2 Benefícios da gamificação

O fato de tornar um ambiente gamificado traz diversos benefícios que dependerão de como (e quais) os elementos de gamificação são integrados e utilizados no ambiente. Diversos estudos avaliam o incremento de gamificação em seus ambientes (LI et al., 2013) (DENNY, 2013)(NIKKILA et al., 2013). Na grande maioria, a gamificação incentivou a participação e o engajamento dos usuários. A utilização de gamificação em sistemas colaborativos pode aumentar o engajamento do usuário, principalmente aumentando o desenvolvimento de atividades dentro do sistema (CUNHA; GASPARINI; BERKENBROCK, 2013b). Segundo o mesmo estudo, a utilização de técnicas de gamificação também contribui com outros aspectos da experiência do usuário, tais como a satisfação, o prazer, a

diversão, o entretenimento, o interesse, a motivação, a sensação de desafio, entre outras características.

Por outro lado, mesmo com estes trabalhos indicando os benefícios do uso da gamificação, seu uso deve ser feito com bastante planejamento, uma vez que há riscos em sua utilização, como a redução da motivação intrínseca do usuário pela motivação extrínseca (NICHOLSON, 2012), como por exemplo, o usuário utilizar o sistema apenas para receber recompensas (motivação extrínseca), em vez de utilizá-lo pois o sistema irá trazer algum benefício (motivação intrínseca).

No contexto de ambientes educacionais, uma das grandes vantagens no uso de gamificação é o seu feedback rápido, isto é, é possível proporcionar um sistema no qual os estudantes consigam visualizar o efeito de suas ações e aprendizagens de forma rápida. Além disso, com a experiência de gamificação, os usuários sentem que suas ações fazem sentido dentro de uma causa maior, ou seja, mesmo estes não esteja vendo efetivamente a melhora na sua aprendizagem, eles sentem que seus objetivos contribuem para algo maior e mais importante.

2.2.3 Criando um plano de gamificação

A criação de um plano de gamificação exige que se pense em muitos detalhes para que se tenha um sucesso em sua implantação. Esses detalhes devem ser bem planejados e bem fundamentados. O conjunto de todos eles são incluídos em um documento chamado plano de gamificação. Existem diversas formas de como criar esse plano, porém neste trabalho será apresentado apenas o framework de gamificação 6D, disponibilizado em um curso do Coursera ², por Kevin Werbach's da Universidade da Pensilvânia.

Basicamente, para se construir um plano de gamificação basta seguir e responder 6 questões fundamentais (as chamadas 6D). Cada questão visa a atender aspectos específicos de gamificação que levarão ao seu sucesso. As questões definidas (originalmente em inglês) e seus principais questionamentos são (WERBACH; HUNTER, 2012):

1. ***Define Business Objectives:*** Esta questão visa a responder a seguinte pergunta: “Porque você está gamificando?”, ao mesmo tempo que responde “Qual é o seu objetivo ao utilizar gamificação?”. Deve-se enfatizar o objetivo final ao invés de detalhar os meios;
2. ***Delineate target behaviors:*** Esta questão se preocupa com o comportamento dos utilizadores da gamificação. Responde as seguintes perguntas “O que você quer que seus jogadores façam?” e “Quais são as métricas que permitirá que você possa medir esse comportamento?”. Vale ressaltar que esses comportamentos devem promover seus objetivos de negócios;

² Plataforma online que oferece cursos a distância à milhares de estudantes em diversos países

3. ***Describe your players:*** Esta questão se preocupa com os tipos de usuários de gamificação. Responde perguntas do tipo “Quem são as pessoas que irão participar da sua atividade gamificada?” e “Qual é a sua relação entre elas?”;
4. ***Devise your activity loops:*** Esta questão se preocupa com o como. Deve-se explorar em maior detalhe como você vai motivar os seus jogadores usando loops engajados e de progressão. Em primeiro lugar, deve-se descrever os tipos de feedbacks que o sistema vai oferecer aos jogadores para incentivar ainda mais suas ações, sem deixar de explicar como esse feedback vai trabalhar para motivá-los;
5. ***Don't forget the fun:*** Embora mais abstrato do que alguns dos outros elementos, é necessário garantir que o seu sistema gamificado seja tão divertido quanto os outros aspectos;
6. ***Deploy the appropriate tools:*** Por fim, você já deve ter identificado vários dos elementos do jogo e outras especificidades do seu sistema gamificado. Responde à pergunta “Quais são alguns dos elementos do jogo envolvidos e qual será a sua experiência para os jogadores?”.

2.3 Sistemas tutores inteligentes

Os Sistemas Tutores Inteligentes (STI) têm uma história interessante, originada na área de Inteligência Artificial (IA). Existem diversas definições na literatura a cerca de STI. Podemos definir os STI como programas de computador que utilizam técnicas da IA para representar o conhecimento e levar a termo uma interação com o aluno (SLEEMAN; BROWN, 1982). De acordo com (GIRAFFA, 1999), o STI é um sistema que incorpora técnicas de Inteligência Artificial a fim de tentar criar um ambiente que leve em consideração os diversos estilos cognitivos dos alunos. Existe também uma definição mais detalhada ao considerar os STI como programas que, interagindo com o aluno, modificam suas bases de conhecimento (aprendem), percebem as intervenções do aluno, e possuem a capacidade de adaptar as estratégias de ensino de acordo com o desenrolar do diálogo (VICARI, 1989).

Os sistemas tutores inteligentes se apresenta como uma ferramenta com enorme potencial do ponto de vista educacional, uma vez que englobam a capacidade de se adaptar às necessidades de cada aluno, aumentando consideravelmente a interação aluno-sistema, o que é de extrema necessidade para garantir a qualidade da aprendizagem. Assim, nesse sentido, um STI deve possuir em sua estrutura básica características que envolvam as seguintes questões: O que ensinar? Como ensinar? E a quem ensinar?

Para representar essas questões é necessário um entendimento interdisciplinar, envolvendo conhecimentos de ciência da computação, psicologia e educação. Um STI entra como uma ferramenta que possibilita a integração entre essas disciplinas, promovendo um

ganho considerável na aprendizagem. A Figura 3 ilustra o posicionamento de STIs nas disciplinas que envolvem as questões citadas acima. Estes suprem quase que completamente a cobertura da área de inteligência artificial e educação (WOOLF, 2010).

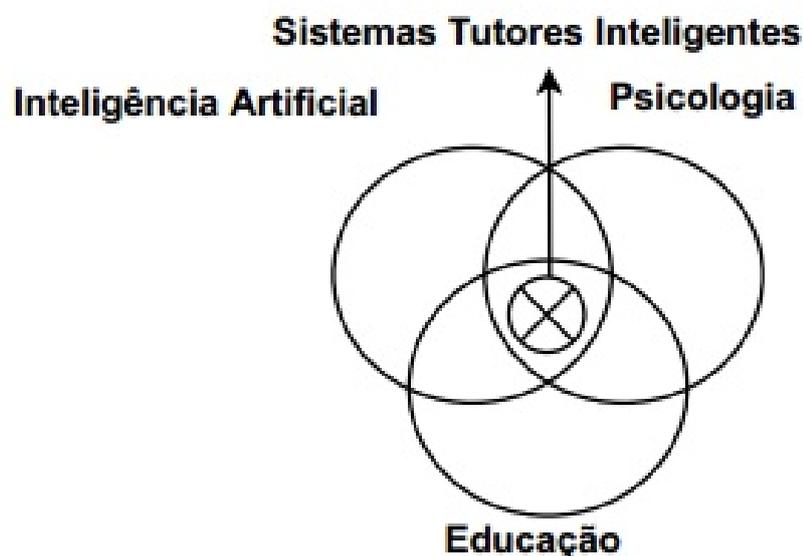


Figura 3 – STI como uma ferramenta de integração multidisciplinar

Fonte: Adaptado de (WOOLF, 2010)

Cada disciplina, isoladamente, tem um propósito bem definido. A Inteligência artificial trata questões sobre inteligência e por conseguinte aprendizagem do aluno envolvido no sistema. A Psicologia, por sua vez, particularmente a subárea de ciência cognitiva, trata como as pessoas pensam e aprendem. Por fim, a educação foca em como apoiar o ensino de uma melhor forma. Por este motivo, o desenvolvimento de um STI requer um esforço concentrado e uma diversidade de conhecimentos e técnicas, envolvendo uma equipe multidisciplinar, composta por profissionais da Informática, Pedagogia, Comunicação, Psicologia e outras áreas, visto que a IA é uma área tipicamente interdisciplinar (VICARI; GIRAFFA, 2003).

2.3.1 Características de um STI

Para um sistema ser considerado um sistema tutor inteligente ele deve apresentar algumas características fundamentais. Essas características são advindas das diversas definições de um STI. Um STI deve passar em três “testes” antes de ser considerado inteligente (JONASSEN; WANG, 1993):

1. Conteúdo do tema ou especialidade deve ser codificado de modo que o sistema possa acessar as informações, fazer inferências ou resolver problemas;

2. Sistema deve ser capaz de avaliar a aquisição deste conhecimento pelo aluno;
3. As estratégias tutoriais devem ser projetadas para reduzir a discrepância entre o conhecimento do especialista e o conhecimento do aluno.

Além disso, um STI deve possuir as seguintes características(LOINAZ, 2001):

- O conhecimento do domínio está restrito e claramente articulado.
- Possuem conhecimento do aluno que lhes permite dirigir e adaptar o ensino.
- A seqüência do ensino não está predeterminada pelo designer instrucional.
- Realizam processos de diagnóstico mais adaptados ao aluno e mais detalhados.
- A comunicação Tutor-Aluno melhora, permitindo que o aluno realize perguntas ao tutor.

Assim, de uma forma genérica, os STI se caracterizam por representar separadamente a matéria que se ensina (modelo do domínio) e as estratégias para ensiná-la (modelo pedagógico). Por outro lado, caracterizam o aluno (através do modelo do aluno) com o objetivo de obter um ensino individualizado. Outra característica marcante é a necessidade da interface de comunicação ser um módulo bem planejado, de fácil manipulação, e que favoreça o processo de comunicação tutor-aluno.

Neste sentido, para ser inteligente, um STI precisa ser flexível, ter a capacidade de aprender com o ambiente e atualizar o seu conhecimento. Além de transmitir conhecimento específico, deve inferir sobre como gerenciar o processo de ensino-aprendizagem durante a interação com o aluno. Por propiciarem um maior nível de interação entre este e o sistema, os STI têm se mostrado eficientes para a melhoria do desempenho e da motivação dos alunos (COSTA, 1999).

2.3.2 Arquitetura dos Sistemas Tutores Inteligentes

O principal objetivo dos Sistemas Tutores Inteligentes é proporcionar um ensino adaptado a cada aluno, tentando se aproximar do comportamento de um professor humano na sala de aula. Contudo, para se viabilizar a sua construção, faz-se necessário que o sistema tutor tenha embutido uma grande quantidade de código relacionado a diversos aspectos bem distintos. Estes, como descrito anteriormente, buscam responder as três questões básicas: o que ensinar, como ensinar e a quem ensinar.

Tais questões deságuam em modelos como descritos a seguir (Modelo do Domínio, Modelo do Aprendiz e Modelo Pedagógico) (WOOLF, 2010). A integração desses modelos juntamente com uma interface para o aluno, formam a arquitetura tradicional de um sistema tutor inteligente, conforme ilustrado na Figura 4.

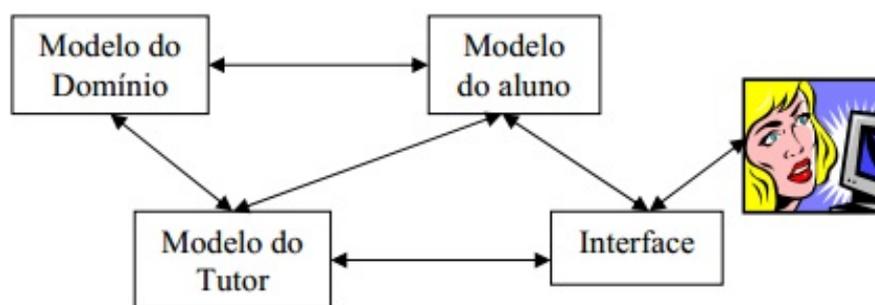


Figura 4 – Arquitetura tradicional de um sistema tutor inteligente

Fonte: Adaptado de (WOOLF, 2010)

Esta arquitetura é denominada clássica e também conhecida como função tripartida ou arquitetura tradicional de STI. O termo tripartido se refere às funções associadas aos modelos do tutor, do aluno e do domínio. Esses modelos têm funções específicas, tais como:

- **Modelo do Domínio:** Este modelo é responsável por descrever o conteúdo que o tutor está habilitado a ensinar, respondendo a questões sobre o que ensinar. Ele deve ser uma representação do conhecimento especializado em um domínio específico. Dessa forma, deve representar os fatos, procedimentos ou métodos que especialistas usam para realizar tarefas ou resolver problemas;
- **Modelo do Estudante:** Este modelo é responsável por modelar o comportamento e informações do estudante que vai se submeter ao processo de ensino de um tutor em algum domínio, de forma que possa haver um raciocínio automatizado em relação ao mesmo. Modelos do aprendiz tipicamente representam o comportamento do estudante, incluindo respostas do estudante, ações, resultados das ações e resultados intermediários. O comportamento do estudante reflete o conhecimento do estudante, assim como os erros mais comuns;
- **Modelo Pedagógico:** Representa as estratégias e táticas de ensino, respondendo a questões sobre como ensinar, ele detém o conhecimento de como vai ser ensinado, ou seja, ele elabora o sequenciamento das atividades que serão passadas para o estudante a partir de informações obtidas pelas informações do modelo do aprendiz e do domínio.

Podemos ver na arquitetura apresentada na Figura 4, que ela forneceu grandes avanços na modelagem dos ambientes educacionais (principalmente em STI), pois houve uma separação do domínio da sua forma de manipulação (estratégia de aprendizagem). Com isso, as estratégias podem ser implementadas a parte e associadas ao modelo de domínio juntamente com o modelo do estudante.

As ontologias podem ser classificadas em diversos tipos, a depender de seu objetivo final. De acordo com o seu nível de generalidade, elas podem ser classificadas em ontologias de nível superior (top level), ontologias de domínio, ontologias de tarefas e ontologias de aplicação, como ilustrado na Figura 6.

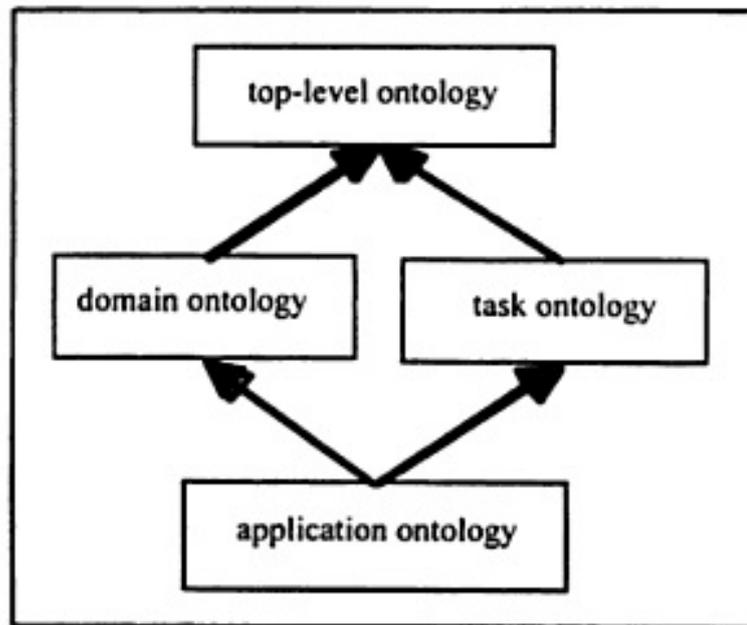


Figura 6 – Tipos de ontologias, de acordo com sua generalidade

Fonte: (GUARINO, 1998)

As ontologias de nível superior descrevem conceitos muito gerais como espaço, tempo, matéria, objeto, evento, ação, entre outros. Eles são independentes do problema particular ou de domínio. As ontologias de domínio e tarefas descrevem, respectivamente, o vocabulário relacionado a um domínio genérico (como a medicina, ou automóveis) ou a uma tarefa genérica ou atividade (como diagnosticar ou venda). Por fim, as ontologias de aplicação descrevem os conceitos em função de um domínio específico e tarefa em particular, sendo assim, muitas vezes especializações de ambas as ontologias relacionadas.

2.4.1 Benefícios de ontologias

O uso de ontologias para modelar e formalizar o conhecimento traz diversos benefícios aos sistemas envolvidos, conforme as vantagens já citadas anteriormente. Algumas razões para o desenvolvimento de uma ontologia são (MCCUINNESS, 2005):

- Automação - Para compartilhar o entendimento comum da estrutura de informações entre as pessoas e os agentes de software: se diferentes páginas da web ou aplicativos compartilham e publicam a mesma ontologia básica dos termos que eles usam, então os agentes de computador pode extrair e agregar informações desses lugares

diferentes e usar essas informações agregadas para responder a consultas do usuário ou como entrada para outras aplicações;

- Reuso - Para permitir a reutilização do conhecimento do domínio: se um grupo de pesquisadores/desenvolvedores cria uma ontologia detalhada, outros podem simplesmente reutilizá-lo para os seus domínios;
- Formalização - Para analisar o domínio do conhecimento: análise formal dos termos é extremamente valioso ao tentar reutilizar ontologias existentes e alargá-los.

Assim, o uso de ontologias tem vários benefícios, incluindo uma maior automação, uma vez que o conhecimento é representado formalmente, permitindo a sua compreensão pelas máquinas. Ontologias têm sido utilizadas em várias áreas da Ciência da Computação, dentre elas: Recuperação de informações na Internet; Processamento de Linguagem Natural; Gestão do Conhecimento; Web-Semântica; Educação.

Além de serem utilizadas em diversas áreas, o uso de ontologias traz diversas outras vantagens, como por exemplo: a comunicação (as ontologias possibilitam a comunicação entre pessoas acerca de determinado conhecimento, pois permitem raciocínio e entendimento sobre um domínio) e Representação de Conhecimento (As ontologias formam um vocabulário de consenso que permite representar conhecimento de um domínio em seu nível mais alto de abstração).

2.4.2 OWL e protégé

O conceito de ontologia é muito abstrato do ponto de vista computacional. Para criar e representar uma ontologia precisamos de uma linguagem de modelagem específica que a represente e uma ferramenta apropriada que facilite sua criação. Neste sentido, temos escolhido a linguagem OWL como forma de representação de uma ontologia e a ferramenta protégé será utilizada em sua criação.

A Web Ontology Language (OWL) apareceu para resolver um problema de limitação de poder expressivo existente nas linguagens RDF e RDF Schema. OWL pertence a uma família de linguagens de representação do conhecimento para autoria de ontologias. É uma recomendação da W3C desde Fevereiro de 2004 e pode ser usada por aplicações que necessitam processar o conteúdo de informação, em vez de apenas apresentá-la para os seres humanos.

OWL facilita uma maior facilidade de interpretação por máquina de conteúdo da Web do que a suportada pelo XML, RDF e RDF Schema (RDF-S), fornecendo vocabulário adicional juntamente com uma semântica formal. OWL se baseia na lógica descrição (DL) e é composta por três sublinguagens que oferecem vantagens importantes em determinados cenários de aplicação.

As ontologias podem ser descritas utilizando a linguagem OWL. No entanto, a criação de uma ontologia diretamente em OWL é uma tarefa difícil e suscetível a erros. Para

facilitar a criação de ontologias, algumas ferramentas surgiram com interfaces gráficas muito amigáveis. Existem várias ferramentas para a criação de ontologias. A Figura 7 apresenta uma lista com várias ferramentas mais utilizadas pela comunidade (CARDOSO, 2007).

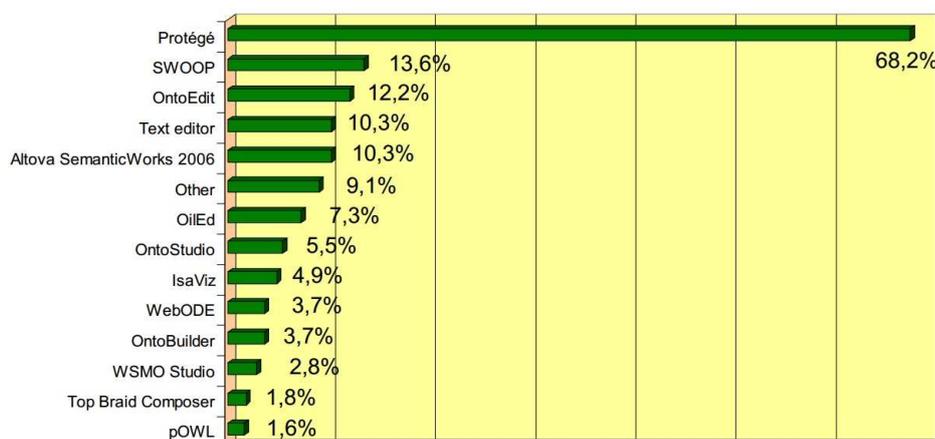


Figura 7 – Ferramentas existentes para criação de ontologias

Fonte: (CARDOSO, 2007)

Deste modo, a ferramenta que foi escolhida para ser utilizada neste trabalho foi a Protégé. Protégé é uma plataforma livre, de código aberto que fornece uma comunidade de usuários em crescimento, com um conjunto de ferramentas para a construção de modelos de domínio e aplicações baseados no conhecimento com ontologias (SCIENCES, 2013) ³.

Protege é baseado em Java, é extensível, e proporciona um ambiente plug-and-play que faz com que seja uma base flexível para prototipagem rápida e desenvolvimento de aplicativos.

O editor Protégé-OWL permite aos usuários criar ontologias para a Web Semântica, em particular na linguagem OWL. Uma ontologia OWL inclui descrições de classes, propriedades e suas instâncias (SMITH; WELTY; MCGUINNESS, 2004) ⁴. Algumas telas podem ser encontradas em seu site, como a apresentada na Figura 8.

A arquitetura do Protégé pode ser adaptada para construir aplicações baseadas em ontologias simples e complexas. Os desenvolvedores podem integrar a saída do Protégé com sistemas de regras ou outros solucionadores de problemas para construir uma ampla gama de sistemas inteligentes. Além disso, a sua equipe desenvolvedora e a vasta comunidade do Protégé está sempre disponível a ajudá-lo.

A construção de ontologias de domínio envolve, além da definição de seu domínio, escopo, linguagem e ferramenta que será utilizada, uma metodologia que guiará o pro-

³ Disponível em <http://protege.stanford.edu>

⁴ Disponível em <http://www.w3.org/TR/owl-guide/>

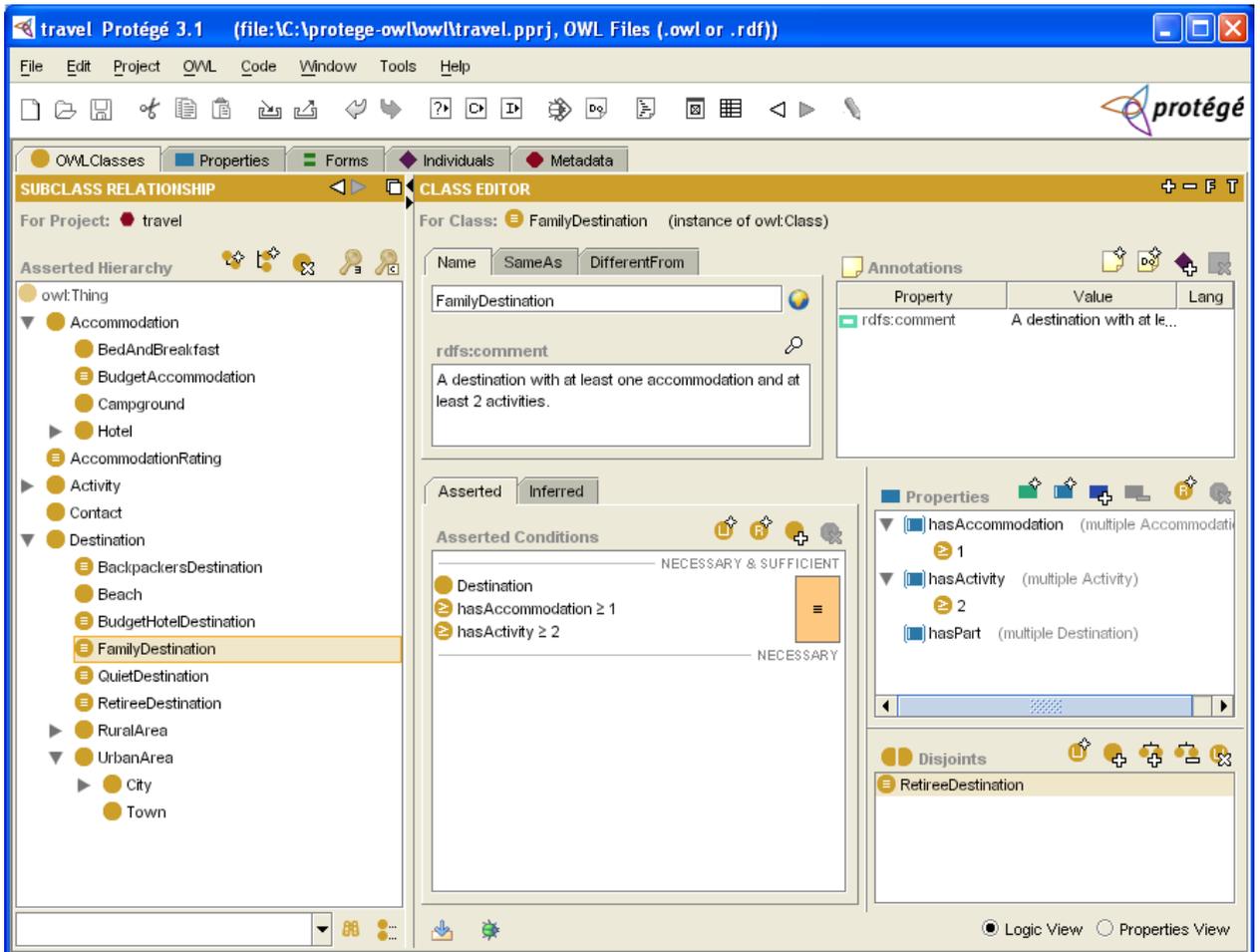


Figura 8 – Ferramentas existentes para criação de ontologias

Fonte: (SCIENCES, 2013)

cesso de construção. Por estar fora do escopo deste trabalho, não serão apresentadas as metodologias existentes para construção de ontologias.

3 TRABALHOS RELACIONADOS

Assim como todas as pesquisas científicas, em um trabalho de dissertação também é necessário inicialmente realizar um estudo sobre os trabalhos semelhantes já realizados e disponíveis na literatura. Existem diversas maneiras para realizar um estudo desse tipo, em geral chamado de revisão de literatura. Contudo, para que uma determinada revisão de literatura possua um valor científico, é importante que ele seja realizado através de um processo sistemático que envolve diversas etapas. Assim, o estudo dos relacionados a este trabalho foi realizado utilizando um processo conhecido como revisão sistemática (GALVÃO; SAWADA; TREVIZAN, 2004). Neste sentido, este capítulo tem como objetivo apresentar a revisão sistemática realizada, bem como apresentar seus resultados e os principais trabalhos relacionados.

Uma revisão sistemática, assim como outros tipos de estudo de revisão, é uma forma de pesquisa que utiliza como fonte de dados a literatura sobre um determinado tema (SAMPAIO; MANCINI, 2007). É um meio de identificar, avaliar e interpretar todas as pesquisas disponíveis relevantes para uma determinada questão de pesquisa, área temática ou fenômeno de interesse e tem o objetivo de sintetizar um conjunto de estudos existentes de maneira justa, sem que os resultados obtidos sejam tendenciosos (KEELE, 2007).

As revisões sistemáticas devem ser realizadas em conformidade com uma estratégia de busca pré-definida. Durante uma revisão sistemática os pesquisadores devem identificar e relatar quais estudos que estão relacionados a sua pergunta de pesquisa, bem como devem identificar e relatar as pesquisas que não estão relacionadas a sua pergunta.

A principal vantagem de uma revisão sistemática é que sua metodologia é bem definida, e com isso se torna menos provável que os resultados sejam tendenciosos, apesar de não proteger contra o viés de publicação nos estudos primários. A principal desvantagem de revisões sistemáticas da literatura, por sua vez, é que elas exigem um esforço consideravelmente maior do que as tradicionais revisões de literatura (KEELE, 2007).

Para avaliar a relevância da proposta apresentada foi iniciada a realização de uma revisão sistemática. A principal razão para realizar esta revisão sistemática é **fornecer uma visão geral dos trabalhos apresentados na literatura que envolve o uso de técnicas de avaliação por pares aplicadas na área de educação online**. Além do nosso objetivo primário, podemos listar os seguintes objetivos secundários:

- Resumir os resultados de pesquisa apresentados na literatura;
- Identificar os principais assuntos de pesquisa já avaliados pela literatura
- Identificar conceitos conflitantes sobre assuntos de pesquisa
- Definir o estado da arte sobre o tema apresentado.

Uma revisão sistemática da literatura, por ter uma metodologia bem definida, se difere da revisão da literatura clássica em relação ao contexto da estratégia de pesquisa, bem como no que diz respeito à utilização de critérios de inclusão e exclusão previamente definidos, o que permite a construção de uma revisão sistemática e clara da literatura (KEELE, 2007).

O processo de revisão sistemática inicia-se com a definição do protocolo de pesquisa. Este é responsável pela definição das questões de investigação e os métodos que serão utilizados para conduzir o processo de revisão. Portanto, o protocolo pode ser visto como o objeto fundamental da revisão sistemática. Na tentativa de reduzir o viés causado pelo pesquisador e para fazer com que a revisão sistemática seja um processo replicável, é muito importante que o protocolo seja definido antes do início dos trabalhos.

Sendo assim, o primeiro passo foi construir o protocolo para a revisão sistemática, apresentado na subseção 3.1. Em seguida, o processo de revisão sistemática é executado seguindo as definições pré-especificadas no protocolo. O processo de execução da revisão sistemática é apresentado na subseção 3.2. Por fim, é apresentada a análise dos dados obtidos com a revisão na subseção 3.3.

3.1 Protocolo

Esta seção tem como objetivo apresentar o protocolo da revisão sistemática apresentada. É o primeiro passo e é importante que algumas questões estejam claras em sua definição, tais como a definição dos principais objetivos de pesquisa, a definição da(s) pergunta(s) de pesquisa, a criação da(s) String(s) de busca, a definição das técnicas que serão usadas para avaliar os estudos e o planejamento e a seleção dos engenhos de busca e dos artigos.

Todas essas informações são essenciais e definirão o sucesso e a reprodutibilidade da revisão sistemática. Diante disso, as subseções a seguir definem estas informações do protocolo da revisão apresentada. A subseção 3.1.1 apresenta uma descrição geral do protocolo, seus objetivos, problemas abordados e as questões de pesquisa definidas. A subseção 3.1.2, por sua vez, define a estratégia de busca a ser utilizada, a String de busca e quais bases bibliográficas serão utilizadas como fonte de estudos e finalmente, os critérios de seleção (inclusão e exclusão) dos estudos são apresentados na subseção (3.1.3).

3.1.1 Descrição geral

A primeira etapa para a realização de uma revisão sistemática é definir os objetivos da pesquisa e o problema a ser abordado. Para isso é elicitado um conjunto de questões de pesquisas que serão usadas ao longo da revisão. Com isso, a revisão apresenta as seguintes características:

Tema: Uma revisão sistemática sobre o uso de avaliação por pares em ambientes educacionais online.

Objetivo: O principal objetivo desta revisão sistemática é fornecer uma visão geral dos trabalhos apresentados na literatura que envolve o uso de técnicas de avaliação por pares aplicadas em ambientes educacionais online.

Problema (high level question): Como as abordagens de avaliação por pares tem sido usadas em ambientes educacionais online?¹

O próximo passo é definir as questões de pesquisa, baseado no problema e no objetivo da revisão. As seguintes questões foram definidas:

- **QP1: Em quais contextos e níveis educacionais a aplicação de avaliação por pares tem sido mais aplicada em ambientes educacionais online?**

A resposta a esta questão permite identificar quais são os contextos (saúde, educação, lazer, entre outros) e os níveis educacionais (básico, fundamental, médio e universitário) do ambiente online no qual o uso de avaliação por pares tem sido aplicado.

- **QP2: Com que finalidade as técnicas de avaliação por pares tem sido usadas nos ambientes educacionais online?**

Esta questão tem como objetivo saber a finalidade do uso de avaliação por pares nos ambientes online. A finalidade pode ser, por exemplo, corrigir problemas de múltipla escolha, corrigir testes, atividades discursivas e/ou simplesmente avaliar as características de outro aluno, entre outras possíveis.

- **QP3: Há evidências que indicam que o uso de abordagens de avaliação por pares melhora o desempenho dos alunos em ambientes de aprendizagem online?**

Esta questão pretende analisar documentos e identificar se eles fornecem alguma evidência de que o uso de abordagens de avaliação por pares melhora o desempenho dos alunos em ambientes de aprendizagem online. Desempenho nesse caso refere-se ao nível de aprendizagem do aluno e o quanto este consegue aprender. Essas evidências devem envolver resultados positivos e negativos, incluindo ou não avaliação empírica. Em caso afirmativo, o quanto melhora o desempenho e em que aspecto.

- **QP4: Há evidências que indicam que o uso de abordagens de avaliação por pares auxilia o trabalho dos professores?**

Esta questão pretende analisar documentos e identificar se eles fornecem alguma evidência de que o uso de abordagens de avaliação por pares auxilia o trabalho dos

¹ Este tipo de pergunta se caracteriza como exploratória, buscando o entendimento/esclarecimento das características do uso de avaliação por pares em ambientes educacionais online. Ao mesmo tempo é descritiva e classificatória, com um formato “Como X é?”

professores em ambientes de aprendizagem online. Essas evidências devem envolver resultados positivos e negativos, incluindo ou não avaliação empírica. Em caso afirmativo, como o uso auxilia os professores.

- **QP5: Há/Houve dificuldades em se utilizar uma abordagem de avaliação por pares nestes ambientes online?**

Esta questão tem como objetivo identificar possíveis dificuldades apontadas nos estudos a respeito do uso de avaliação por pares em ambientes educacionais online.

Essas questões apresentadas acima são a base fundamental da revisão sistemática apresentada. Elas nos darão informações suficientes para analisar como é o uso de avaliação por pares e sua finalidade nesses ambientes (QP1 e QP2), seus benefícios tanto aos professores como alunos (QP3 e QP4), bem como possíveis problemas na sua utilização (QP5).

3.1.2 Estratégia de busca

Antes de iniciarmos o estudo sistemático é necessário definir uma estratégia de busca para os estudos primários, isto é, estudos empíricos que investigam uma pergunta de pesquisa específica (KEELE, 2007)). Basicamente, a estratégia é composta por duas etapas: a primeira define as palavras-chave e a semântica da pesquisa; a segunda, por sua vez, define quais bibliotecas digitais, revistas e conferências serão utilizadas como fonte de estudos. Essas duas etapas serão apresentadas a seguir.

Na primeira etapa, inicialmente, é preciso definir quais são as palavras-chave para a revisão proposta. Estas palavras podem ser identificadas a partir da definição das questões de pesquisa, bem como da identificação de um conjunto de sinônimos para elas. As palavras-chave (e seus sinônimos) definidos através das questões de pesquisa são apresentadas a seguir:

- Peer Assessment (Peer Review, Co-assessment, Peer Grading, Self Assessment);
- Online Learning Environment (Massive Open Online Courses, Intelligent Tutoring System, Collaborative Learning, CSCL, Online Courses, LMS, Learning management system)
- Student (Learner);
- Teacher (Tutor).

A partir disso, é possível construir a String ² de busca, através da combinação das palavras-chave e de seus sinônimos, usando os operadores lógicos (OU) ³ entre os sinônimos

² Termo de programação para se referir a um texto

³ Tradução em português do operador lógico OR

identificados, enquanto que o operador (E) ⁴ será usado entre as palavras-chave. Como resultado, para a pergunta de pesquisa foi definida a seguinte String de busca a partir das palavras-chave e de seus sinônimos:

String = (“peer review” OR “peer assessment” OR “co assessment” OR “peer grading” OR “self assessment”) AND (“massive open online courses” OR “online learning environment” OR “intelligent tutoring system” OR “collaborative learning” OR “CSCL” OR “online courses” OR “LMS” OR “Learning management system”) AND ((student OR learner) OR (teacher OR tutor))

A limitação para professores e estudantes foi proposital, pois o trabalho proposto apenas envolve esses dois tipos de usuários. Após as definições gerais da revisão sistemática, a definição do problema e a construção da String de busca, o próximo passo é definir em quais bases bibliográficas a pesquisa será realizada. É necessário que a base possua algum mecanismo automático para obtenção dos estudos indexados. Além disso, é necessário que a base atenda os seguintes critérios para serem incluídas nesta revisão:

- Disponibilidade dos artigos nas bases de dados bibliográficas;
- Disponibilidade de busca a partir de palavras-chave (no qual será feita a partir da String de busca);
- Alta relevância da base de dados bibliográfica.

Diante dos critérios de inclusão apresentados, as bases de dados bibliográficas digitais que os atendem e foram selecionadas e incluídas na pesquisa são:

- ACM Digital Library (<http://portal.acm.org>)
- IEEE Xplore (<http://ieeexplore.ieee.org>)
- Elsevier (Science Direct) (<http://www.sciencedirect.com>)
- Elsevier Scopus (<http://www.scopus.com>)
- SpringerLink (<http://link.springer.com>)
- Compendex (<http://www.engineeringvillage.com>)
- ISI Web of Knowledge (<http://wokinfo.com>)

⁴ Tradução em português do operador lógico AND

3.1.3 Seleção de Estudos

O próximo passo da definição do protocolo da revisão sistemática, visando melhorar os resultados obtidos, consiste na criação de alguns critérios de seleção para os artigos que serão selecionados (inclusão, exclusão e qualidade) com base na pergunta de pesquisa, String de busca e bases bibliográficas a serem utilizadas (KEELE, 2007). Estes critérios são os parâmetros de referência no qual os artigos extraídos serão avaliados, podendo estes serem incluídos ou excluídos da revisão sistemática.

Um estudo deve ser incluído na revisão sistemática baseado em sua relevância com relação a(s) questões de pesquisa propostas. Sendo assim, apenas serão incluídos artigos que descrevam pesquisas relacionadas ao tema abordado. Os critérios de inclusão definidos foram:

- Estudos revisados por especialistas que fornecem respostas para as perguntas de pesquisa;
- Estudos que apresentam alguma abordagem para a aplicação de técnicas de avaliação por pares em ambientes de aprendizagem online;
- Estudos que incidem sobre o uso de técnicas de avaliação por pares em ambientes educacionais on-line com o objetivo de: melhorar a experiência de aprendizagem dos alunos e/ou auxiliar os professores neste tipo de ambiente, diminuindo sua sobrecarga de trabalho;
- Estudos que fornecem alguma evidência empírica sobre as vantagens/desvantagens/-problemas do uso de técnicas de avaliação por pares em ambientes educacionais online
- Estudos sobre o uso de técnicas de avaliação por pares em questões discursivas (atividades escritas);
- Estudos publicados a partir do ano de 2004 (01 de janeiro de 2004)(10 anos);
- Caso existam vários artigos que relatem o mesmo estudo, apenas o mais recente deverá ser selecionado. De forma análoga, se um artigo apresenta mais de um estudo, estes serão tratados/avaliados individualmente.

Por sua vez, os critérios de exclusão são aqueles que irão retirar os estudos da revisão sistemática. Geralmente são estudos que, por algum motivo, não apresentam relevância suficiente para a revisão sistemática. Os critérios de exclusão definidos foram:

- Estudos não completos e que apresentam lacunas no trabalho e/ou não apresentam fundamentação teórica adequada;

- Estudos que não incluem resultados experimentais;
- Estudos publicados como short-papers e/ou estudos secundários, como outras revisões sistemáticas e surveys e capítulos de livro;
- Artigos de similares (quando dois ou mais artigos têm conteúdo muito parecido, iremos manter apenas o estudo mais recente);
- Estudos que não foram revisados por especialistas (peer review);
- Estudos que não estão relacionados com as questões de pesquisa apresentadas;
- Estudos publicados em língua estrangeira, diferente de inglês;
- Estudos publicados antes do ano de 2004 (publicados até 31 de dezembro de 2003);

Por fim, precisamos definir os critérios de qualidade. Estes, por sua vez, ajudam a selecionar e identificar os principais estudos em relação às suas qualidades e seu potencial em relação às questões de pesquisa. Os principais critérios de qualidade definidos foram:

- Número alto de citações do trabalho;
- Meio onde o trabalho foi publicado, em especial conferências e journals;
- Maior grau de relevância do problema tratado no trabalho;
- Alta qualidade dos experimentos apresentados, bem como seus resultados e as discussões geradas;
- Maior tipicidade da pesquisa, onde as soluções apresentadas podem ser extraídas para outras áreas do conhecimento.

3.1.3.1 Processo de Seleção dos Estudos Primários

Após a definição da pergunta de pesquisa, da estratégia de busca dos estudos primários, e dos critérios de inclusão e exclusão, é necessário definir o processo de seleção dos trabalhos primários, o qual é apresentado a seguir. Em primeiro lugar, deverá ser realizada as buscas nas bases de dados especificadas. Os resultados deverão ser extraídos, geralmente em arquivos bibtex com as referências, e importados para uma ferramenta de gerenciamento de revisão sistemática chamada Start.

StArt (State of the Art through Systematic Review) tem como objetivo dar suporte ao pesquisador, apoiando a aplicação da técnica de revisão sistemática. A versão atual da ferramenta StArt apoia as três etapas do processo, sendo que no Planejamento, o pesquisador preenche o protocolo. Na Execução, o pesquisador adiciona e avalia os artigos que a compõem e extrai informações daqueles que são relevantes ao tópico de pesquisa

abordado. Na Sumarização, são apresentados gráficos e tabelas que dão uma visão geral e auxiliam a descrever o estado da arte do tema pesquisado (ZAMBONI et al., 2010).

Após a seleção dos candidatos a estudos primários, estes deverão ser analisados de forma a excluir os artigos no que dizem respeito a nossa pergunta de pesquisa e aos critérios de exclusão. Para isso, será realizada uma análise dos estudos contidos na ferramenta com base em uma estratégia de busca pré-definida. Os estágios do processo de exclusão de estudos podem ser vistos a seguir:

1. Excluir estudos com base nos seus títulos;
2. Excluir estudos com base nos seus resumos (abstracts);
3. Excluir estudos com base na sua Introdução e Conclusão;
4. Obter os artigos primários e analisar criticamente os estudos.

Durante o Passo 1, devem ser excluídos apenas os estudos que claramente não estão relacionados ao tema da pesquisa. Assim, artigos cujo os títulos indicam claramente que estão fora do escopo da pesquisa devem ser excluídos. No entanto, nem sempre os títulos indicam claramente sobre o que o artigo está tratando. Nesses casos, os artigos devem ser incluídos para revisão na próxima etapa (Passo 2). O mesmo fato deve acontecer caso a análise dos resumos não indique claramente o que está sendo feito por um determinado estudo. Este processo irá reduzir drasticamente a quantidade de estudos irrelevantes.

Após a análise dos resumos, os artigos que foram aprovados serão novamente submetidos para análise no Passo 3, onde uma visão geral e a introdução e a conclusão dos artigos serão avaliadas. Ao final do processo, os estudos selecionados e que não foram excluídos nas etapas anteriores serão submetidos a análise de dados (Passo4), onde as principais informações relevantes a pesquisa serão extraídas.

3.1.3.2 Resumo da coleta de dados

Após a definição da pesquisa e os processos de seleção, deverá ser realizado um resumo detalhado dos estudos selecionados com o objetivo de representar o conhecimento gerado pela pesquisa. Este resumo trata de questões relativas a pergunta de pesquisa investigada. Com isso, será possível obter conclusões sobre as questões de pesquisa através da obtenção dos dados.

O processo de extração de dados deverá ser realizado através da leitura do texto completo de cada um dos trabalhos selecionados nas etapas anteriores. Durante esta fase, os dados serão extraídos de cada um dos artigos incluídos nesta análise sistemática de acordo com um formulário de extração pré-definido, ilustrado na Tabela 1. Essa forma permite registrar os detalhes completos dos artigos em análise e ser mais específico sobre como cada um deles abordou as questões da nossa pesquisa.

Tabela 1 – Formulário de extração dos dados.

Número	Dados do estudo	Descrição
1	Identificador do estudo	Identificador único para o estudo
2	Data da extração	
3	Autores, ano, título e país	
4	Tipo do artigo	Journal, conferência, revistas
5	Contexto de aplicação	Indústria, Academia
6	Método de pesquisa (baseado em (EASTERBROOK et al., 2008))	Experimento controlado, estudo de caso, survey, etnografia, ação de pesquisa, cenário ilustrativo, não aplicável
7	QP1- Contexto de aplicação	Saúde, Educação, Lazer, Jogos ..
8	QP1- Nível educacional	Básico, Fundamental, Médio e Universitário
9	QP2- Finalidade do uso	Corrigir problemas, exercícios Aplicação de provas, atividades discursivas e/ou simplesmente avaliar outro aluno, etc
10	QP3- Evidência desempenho	Positiva/Negativa com ou sem avaliação empírica
11	QP4- Evidência auxílio professor	Positiva/Negativa com ou sem avaliação empírica
12	QP5- Evidência dificuldade	Positiva/Negativa com ou sem avaliação empírica

Fonte: elaborado pelo autor.

3.2 Execução da revisão sistemática

Após a definição do protocolo da revisão sistemática, o próximo passo é a sua execução. Conforme o processo definido na seção 3.1.3.1, o primeiro passo é a obtenção dos estudos, a partir das bases bibliográficas definidas e a String de busca. Este processo foi feito manualmente usando as ferramentas de busca providas por cada uma das bases. O número de artigos obtidos em cada uma das bases e o total está representado na Tabela 2.

Tabela 2 – Resultado da execução da String nas bases científicas selecionadas.

Número	Base de artigos	Número de Artigos (com filtro)
1	ACM Digital Library	312
2	IEEE Xplore	821
3	Elsevier (Science Direct)	272
4	Scopus	818
5	Springer Link	760
6	Compendex	69
7	ISI Web of Knowledge	62
	TOTAIS	3114

Fonte: elaborado pelo autor.

Vale ressaltar que, para estas buscas, foram usados os filtros de artigos a partir do ano de 2004 e na área de ciência da computação. A figura 9 apresenta a relação entre a quantidade de artigos das bases. Todos os artigos obtidos foram incluídos na ferramenta Start.

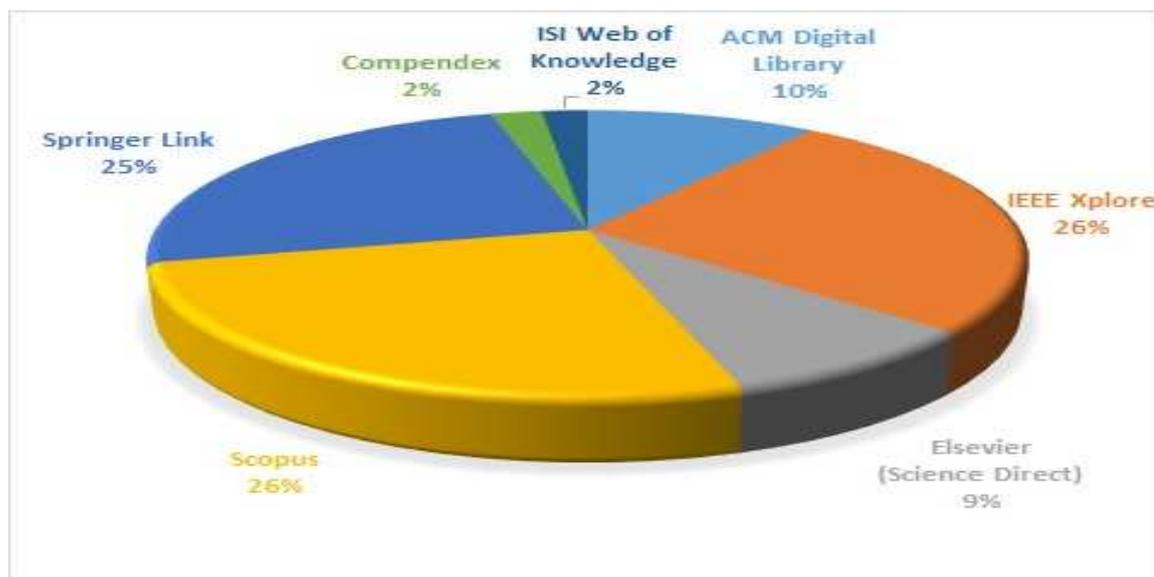


Figura 9 – Percentual de artigos por base bibliográfica

Fonte: elaborado pelo autor.

Após a seleção e inclusão dos candidatos a estudos primários na ferramenta Start (3114 títulos), estes deverão ser analisados com a intenção de excluir os artigos não relevantes. Em primeiro lugar, a ferramenta Start identificou automaticamente 112 estudos duplicados, reduzindo o número total para 3002 estudos. Em seguida, os artigos foram analisados em relação aos seus títulos (passo 1 da seção 3.1.3.1). Este processo, que geralmente reduz drasticamente a quantidade de artigos irrelevantes, excluiu 2831 estudos, deixando apenas 283 selecionados.

Os artigos aprovados pelo título foram então analisados com relação aos seus resumos (passo 2 da seção 3.1.3.1). Os artigos considerados irrelevantes foram excluídos. Destes 283 estudos, 209 foram excluídos no total, sendo 38 duplicados, 1 secundário, 9 shortpapers, 2 não avaliados por pares, 23 que apresentam domínio muito específico para a revisão proposta, 131 que não aplica avaliação por pares em ambientes educacionais online, 2 publicados informalmente e 3 capítulos de livro. Com isso, 74 artigos foram para a próxima etapa.

Os artigos que tiverem seus resumos aprovados prosseguiram para a próxima etapa. Nesta etapa foram avaliados todos os 74 artigos com base em sua introdução e conclusão (passo 3 da seção 3.1.3.1), onde foram excluídos 30 artigos (1 duplicado, 1 shortpaper, 7 que apresentam domínio muito específico para a revisão proposta e 21 que não aplica avaliação por pares em ambientes educacionais online), restando apenas 44 artigos a serem analisados e usados na próxima etapa de extração dos dados.

Em resumo, usando o processo de seleção definido na seção 3.1.3.1, 3070 artigos foram excluídos, restando apenas 44 estudos. A lista contendo os artigos selecionados está no Apêndice A. Os resultados de cada etapa do processo estão ilustrados, de forma mais

clara, na Figura 10. O próximo passo é extrair os dados destes artigos selecionados.

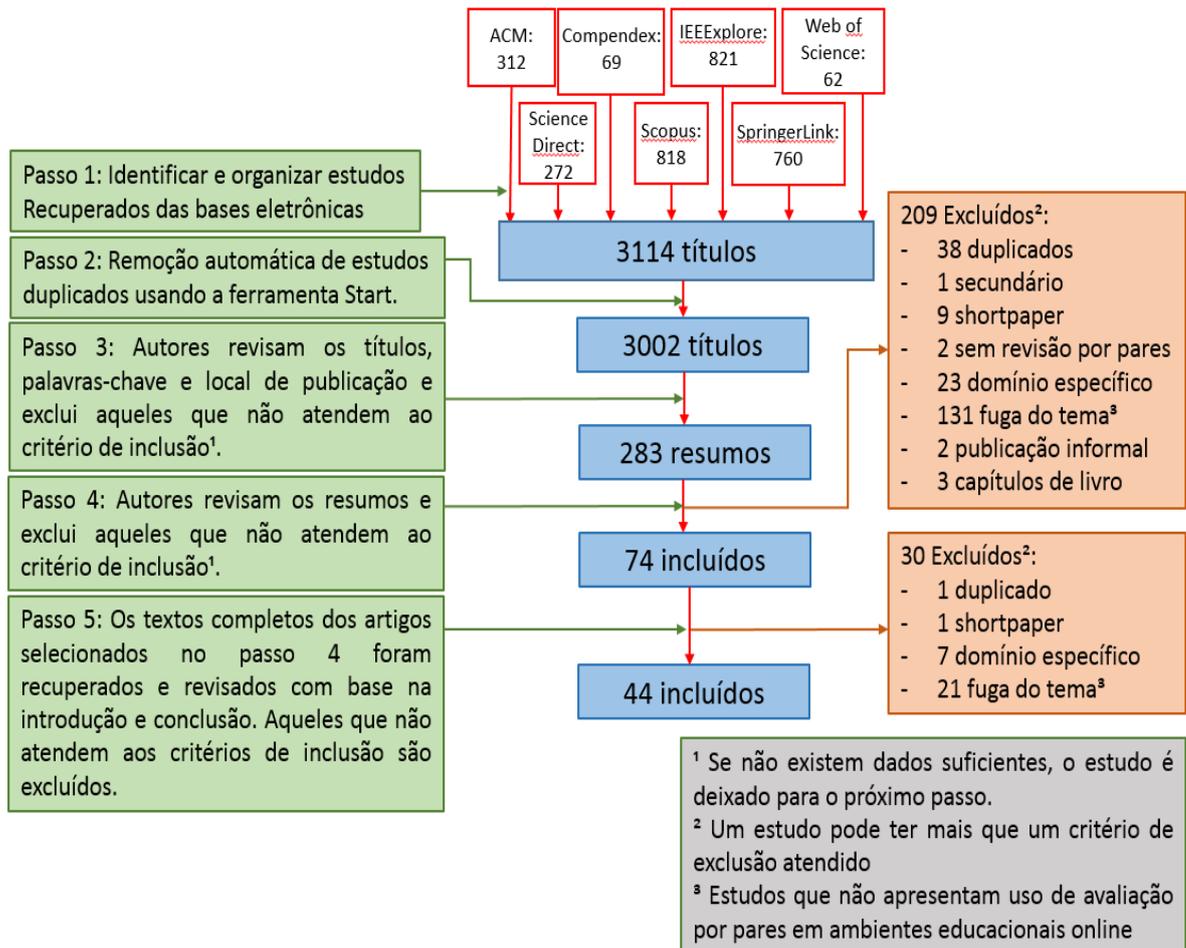


Figura 10 – Fluxo de remoção de artigos da revisão sistemática

Fonte: elaborado pelo autor.

3.3 Análise de dados obtidos

Um total de 44 artigos preencheram os critérios de inclusão e de qualidade (e não foram excluídos por nenhum critério de exclusão). Sendo assim, seus dados foram extraídos e analisados. Antes de apresentar os resultados e as análises para cada pergunta da pesquisa, revisamos os resultados da avaliação a fim de dar uma breve descrição das características gerais dos estudos selecionados.

Esta seção apresenta as características gerais dos artigos incluídos na revisão. A seguir, será descrito em detalhes estas informações. A primeira característica analisada foi com relação ao ano de publicação. Os artigos revisados foram publicados entre 2004 e 2014 (em um período de 10 anos). A figura 11 apresenta a quantidade de artigos analisados por ano de publicação.

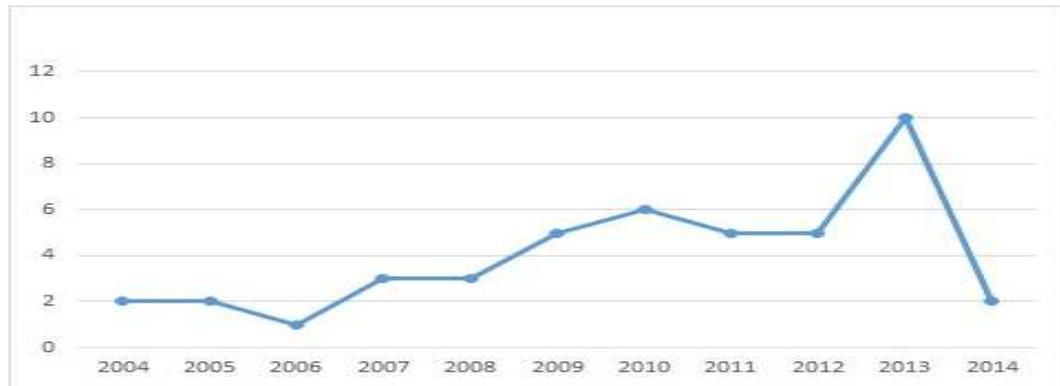


Figura 11 – Quantidade de artigos por ano de publicação

Fonte: elaborado pelo autor.

Podemos observar que os pesquisadores estão atualmente preocupados com o tema, mas é difícil afirmar que existe algum tipo de tendência, apesar do crescimento nos últimos anos. Também é importante notar que, como o processo de busca desta revisão foi realizada na metade do ano de 2014, uma diminuição do número de publicações seria esperada em 2014.

A segunda característica analisada foi com relação ao tipo de artigo. Os artigos incluídos nesta revisão podem ser publicações em conferências (workshops) ou journals. A maioria deles são de documentos publicados em conferências/workshops (70,45%; 31 artigos), seguido por publicações em journals (29,55%; 13 artigos), conforme figura 12.



Figura 12 – Tipos dos artigos selecionados para análise

Fonte: elaborado pelo autor.

A terceira e última característica geral analisada foi com relação aos países de publicação dos artigos selecionados. Os dados dessa característica referem-se ao país do primeiro

autor do trabalho, se esta informação estiver no artigo. Caso contrário, refere-se ao país do evento/journal no qual o artigo foi publicado. A Tabela 3 apresenta os países bem como a quantidade de publicações que foram realizadas em cada um deles.

Tabela 3 – Quantidade de publicações por países

País	Quantidade de publicações
Taiwan	6
EUA	5
Itália	4
Holanda	3
Austrália	3
Alemanha	2
Portugal, Reino Unido, Dinamarca, Japão, Serbia, Áustria, Espanha, Uganda, Grécia, Singapore, Israel, França, China, Canada, Indiana	1

Fonte: elaborado pelo autor.

Podemos observar na tabela que houve uma grande distribuição geográfica, atingindo um alto número de países de diversos continentes. Contudo, países como Taiwan, EUA, Itália, Holanda e Austrália se destacaram por possuir um número maior do que três publicações. Em seguida, a Alemanha atingiu duas publicações, seguida pelo restante dos países com apenas uma publicação. No total, seis artigos não foram classificados devido à falta dessa informação nos metadados.

Após a análise dos dados estatísticos gerais, vamos a análise dos dados referentes às questões de pesquisa envolvidas na revisão sistemática. Dessa forma, a primeira questão de pesquisa analisada foi **QP1: Em quais contextos e níveis educacionais a aplicação de avaliação por pares tem sido mais investigada em ambientes educacionais online?**.

Para responder essa pergunta precisamos analisar tanto os dados do contexto quanto os dados do nível educacional. Do ponto de vista do contexto temos que todos os trabalhos selecionados foram referentes à Educação, sendo que 32 (72,7%) deles eram trabalhos envolvendo academia (universidade), enquanto que apenas 5 (11,4%) deles eram trabalhos voltados a indústria. No total, 7 (15,9%) artigos não deixaram claro o contexto no qual estava inserido. Esses dados estão ilustrados na Figura 13.

Por outro lado, com relação ao nível educacional temos que 15 (34,1%) artigos eram trabalhos com estudantes de nível superior enquanto que apenas 6 (13,6%) artigos envolviam alunos de nível médio. Outros 5 artigos (11,4%) eram voltados a estudantes de língua estrangeira e apenas 1 (2,3%) envolvia curso profissional. Somam 17 (38,6%) artigos onde não havia nível educacional específico ou não deixa claro qual nível educacional utilizaram. Esses dados estão ilustrados na Figura 14.

Ao analisarmos os dados apresentados, podemos afirmar que a resposta para a pergunta QP1 seria que **a aplicação de avaliação por pares tem sido mais investigada**

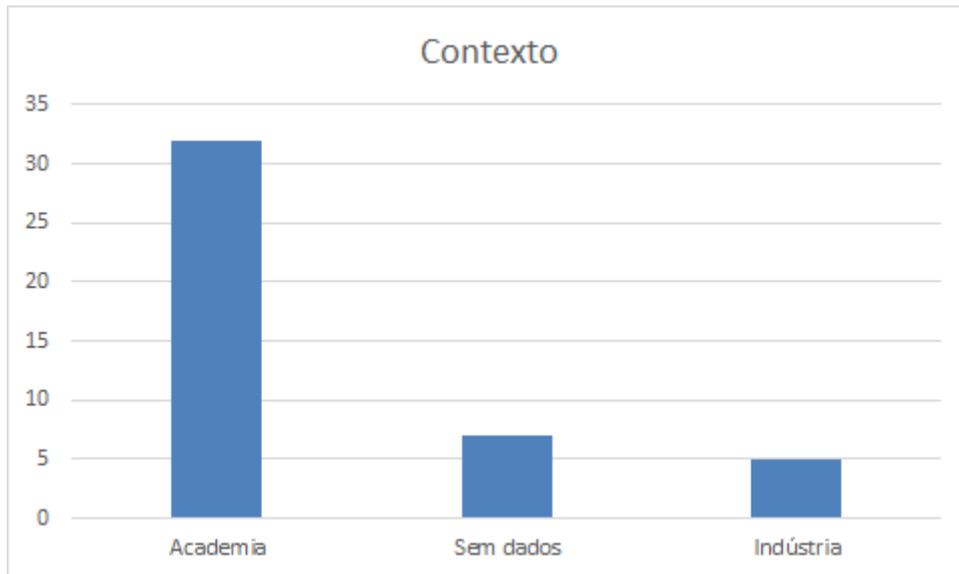


Figura 13 – Quantidade de artigos referentes ao contexto aplicado.

Fonte: elaborado pelo autor.

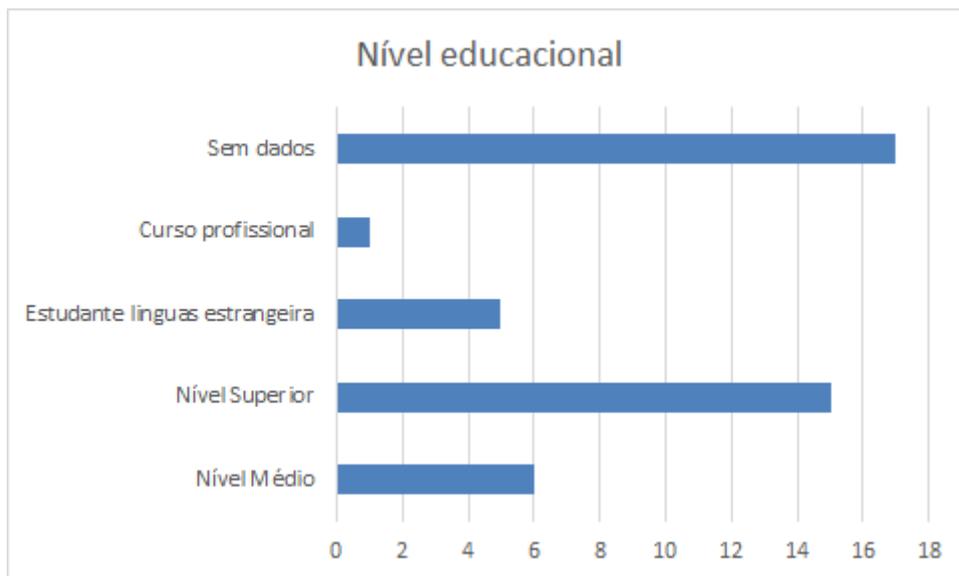


Figura 14 – Quantidade de artigos referentes ao nível educacional envolvido.

Fonte: elaborado pelo autor.

em ambientes educacionais online sob um contexto acadêmico trabalhado em cursos de nível superior.

Esse resultado indica que as pesquisas e utilização de técnicas de avaliação por pares está muito restrito a academia e não a indústria. Isso nos leva a questionar se realmente a utilização dessa técnica é viável do ponto de vista industrial sob o ponto de vista de alguns critérios adotados na indústria (baixo custo, facilidade de implementação e implantação, ganhos para os usuários, entre outros). Além disso, o uso em nível superior pode nos

indicar uma necessidade de mais conhecimento prévio para sua utilização, isto é, será que a técnica pode ser utilizada para estudantes de ensino fundamental ou médio sem maiores problemas? Alguns estudos indicam que é possível utilizá-la em qualquer nível educacional.

A segunda questão de pesquisa analisada foi **QP2: Com que finalidade as técnicas de avaliação por pares tem sido usadas nos ambientes educacionais online?**. Após a leitura dos artigos ficou claro que a técnica de avaliação por pares é muito utilizada com diversas finalidades diferentes. Dessa forma, foi necessário um agrupamento das finalidades. A tabela 4 apresenta as principais finalidades de uso de avaliação por pares apresentadas nos artigos estudados.

Tabela 4 – Finalidade de uso de avaliação por pares

Finalidade de uso	Quantidade
Artigos sem dados	9
Permitir a criação de cursos online para professores	6
Melhorar as habilidades em grupos de trabalho	4
Avaliação de textos escritos (artigos científicos, língua estrangeira)	9
Avaliação de atividades específicas de um curso	12
Construir um framework de sistema de avaliação por pares	4

Fonte: elaborado pelo autor.

Podemos observar na tabela que a principal finalidade de uso é a **avaliação de atividades específicas de um curso**, incluindo, por exemplo, a correção de um código de programação, avaliação de um portfólio, avaliações de mensagens de voz e cartas, entre outras. Em segundo lugar, destacou-se a avaliação de textos escritos, principalmente a **avaliação de atividades de língua estrangeira e redações**. Finalmente, em terceiro, veio a utilização de avaliação por pares para permitir que **professores criem seus cursos online** com mais facilidade. Finalmente, as últimas finalidades eram **a criação de grupos de trabalho mais eficientes** e **a construção de um framework para um sistema de avaliação por pares**.

Esse resultado já era esperado, tendo em vista que a avaliação por pares surgiu da necessidade de diminuir os custos associados a correções de atividades mais trabalhosas e que de certa forma exigia um conhecimento mais específico. Por esse motivo, as correções de atividades específicas de curso superior é a melhor opção para se utilizar tais técnicas de avaliação por pares.

Seguindo com a avaliação das questões de pesquisa, a terceira questão era **QP3: Há evidências que indicam que o uso de abordagens de avaliação por pares melhora o desempenho dos alunos em ambientes de aprendizagem online?**. Os resultados observados para essa pergunta foram bem interessantes. A Figura 15 apresenta a estatística observada com relação aos artigos.

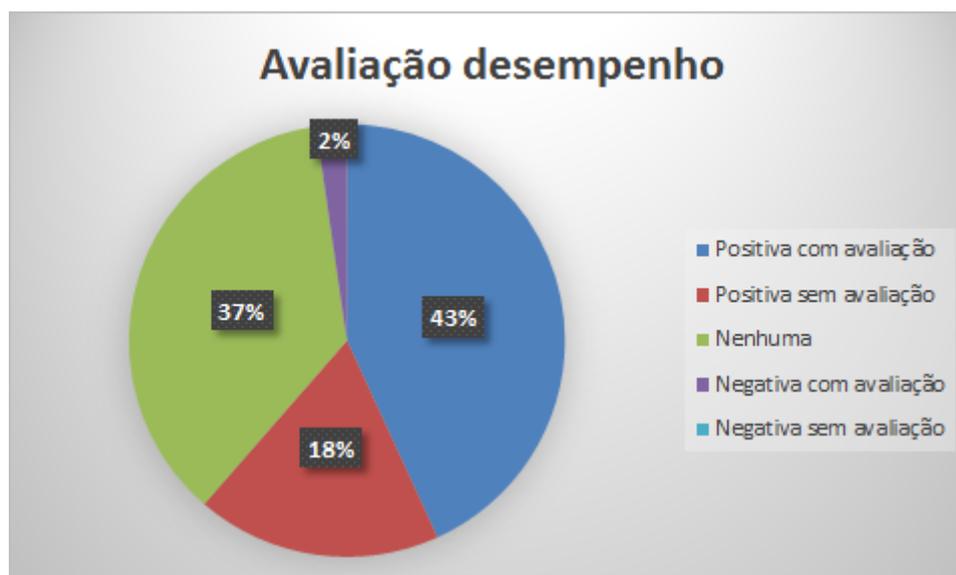


Figura 15 – Estatísticas dos artigos observados com relação ao desempenho dos alunos

Fonte: elaborado pelo autor.

Podemos observar na figura que em 19 (43%) artigos houveram evidências de uma melhora no desempenho dos alunos ao se utilizar avaliações por pares comprovadas com experimentos científicos. Dentre os aspectos melhorados podemos citar: melhora o desempenho e atitudes dos alunos em um grupo CSCL; levaria a um melhor desenvolvimento da equipe, mais satisfação grupo, os níveis mais baixos de conflito no grupo; atitudes mais positivas em relação à colaboração; ajuda a reforçar as habilidades transferíveis, como a colaboração e comunicação; melhora a confiança e desempenho dos alunos; melhora o interesse nas leituras; ajuda a impulsionar a aprendizagem; ganha experiência em leitura crítica; Melhora a capacidade de julgamento; entre outras.

Por outro lado, houveram 8 (18%) avaliações positivas sem avaliação. Dessas podemos citar: tem demonstrado ganhos de aprendizagem impressionantes no ensino de matemática e ciências; ganho de um sentimento de realização; promove e melhora o processo de aprendizagem; desenvolver habilidades genéricas de trabalho em equipe; diminuição das tensões entre os diferentes grupos culturais; melhorar a prática da escrita estilo acadêmica; entre outras.

Finalmente, em 16 (37%) artigos não houveram nenhuma evidência de melhora nem piora com relação ao desempenho dos alunos na utilização da prática. E houve apenas 1 (2%) artigo cujo avaliação foi negativa do uso de avaliação por pares, no qual foi comprovado empiricamente que os comentários dos colegas não têm influência significativa sobre os níveis de reflexão. Com isso, podemos responder a pergunta QP3 afirmando que **existem evidências que indicam que o uso de abordagens de avaliação por pares melhora o desempenho (performance) dos alunos em ambientes de aprendizagem em mais de 60% dos artigos avaliados.**

A próxima questão avaliada foi **QP4: Há evidências que indicam que o uso de abordagens de avaliação por pares auxilia o trabalho dos professores?**. Nessa questão houveram poucos artigos que trouxeram pesquisas concretas. A Figura 16 apresenta a estatística observada com relação aos artigos.

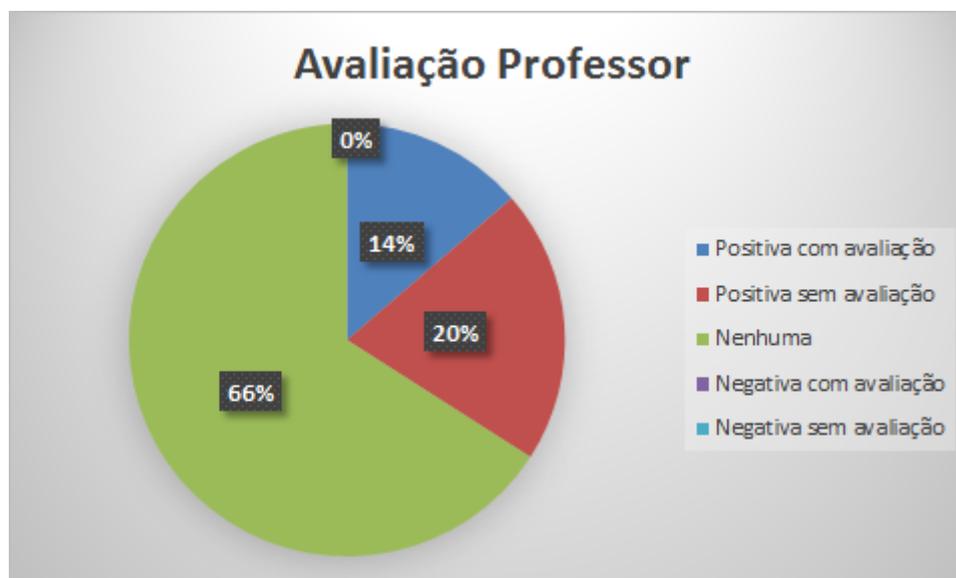


Figura 16 – Estatísticas dos artigos observados com relação ao auxílio aos professores

Fonte: elaborado pelo autor.

Podemos observar que a grande maioria dos artigos (cerca de 29 artigos ou 66%), não apresentaram nenhuma evidência positiva e/ou negativa com relação aos benefícios do uso de avaliação por pares aos professores, contudo, nenhum artigo apresentou evidências negativas.

Por outro lado, cerca de 9 (20%) artigos apresentaram evidências positivas sem avaliação empírica. Dentre essas evidências, podemos citar: aliviar o trabalho do professor nas atividades abertas; Diminuir a sobrecarga do professor; permitir aos professores para avaliar competências de trabalho em equipe de forma mais completa e mais eficiente em relação ao tempo; redução de tarefa demorada; Redução do tempo e esforço; entre outras.

Finalmente, apenas 6 (14%) artigos apresentaram avaliação empírica comprovando que existem evidências positivas na utilização de avaliação por pares ao professor. Dentre essas evidências, podemos citar: cumprir e facilitar de forma bem-sucedida; Permitir a criação de cursos onlines de forma fácil; reduzir as cargas de trabalho de atribuição de notas; suporte ao professor; entre outras.

Mesmo havendo poucos artigos que exploram os benefícios do uso de avaliação por pares aos professores (cerca de 15 ou 34%), ainda assim podemos observar que existem muitos benefícios e portanto, podemos responder a pergunta QP4 afirmando que **existem evidências que indicam que o uso de abordagens de avaliação por pares auxilia o trabalhos dos professores em pelo menos 34% dos artigos avaliados.**

A última pergunta a ser analisada em nossa revisão sistemática é **QP5: Há/Houve dificuldades em se utilizar uma abordagem de avaliação por pares nestes ambientes online?** Nessa questão houveram ainda menos artigos que trouxeram pesquisas concretas, tornando a análise um pouco mais complicada. A Figura 17 apresenta a estatística observada com relação aos artigos.

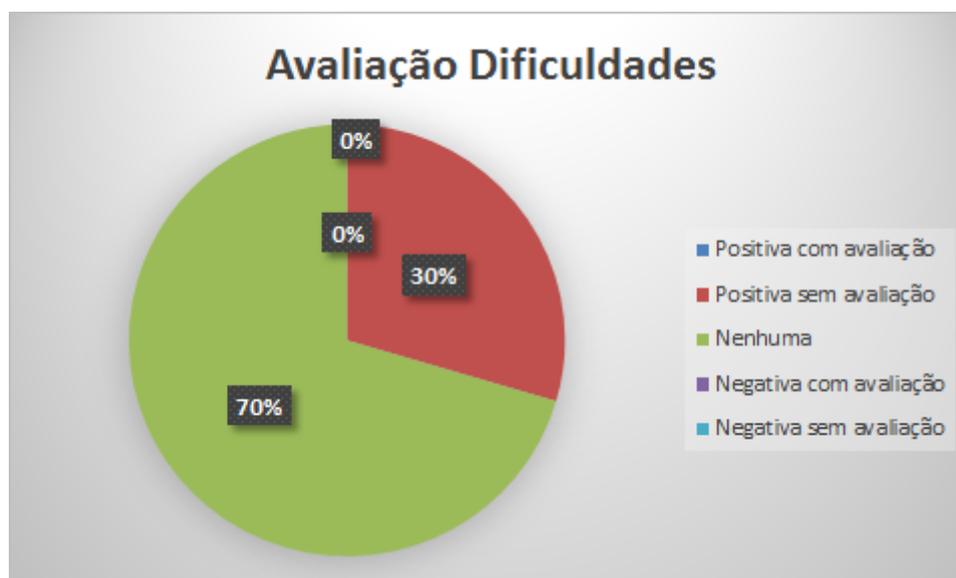


Figura 17 – Estatísticas dos artigos observados com relação às dificuldades envolvidas

Fonte: elaborado pelo autor.

Podemos observar que a grande maioria dos artigos (cerca de 31 artigos ou 70%), não apresentaram nenhuma evidência positiva e/ou negativa com relação as dificuldades envolvidas no uso de avaliação por pares, contudo, nenhum artigo apresentou evidências empíricas/não empíricas negativas. Além disso, nenhum artigo também apresentou evidências empíricas positivas.

Em 13 (30%) artigos, houveram evidências positivas não empíricas de que houveram dificuldades no processo de aplicação da técnica de avaliação por pares. Dentre as principais dificuldades, podemos citar:

- É uma solução de alto custo, devido à um grande processo de implementação e manutenção (HONEYCHURCH et al., 2013);
- Problemas na criação/utilização de questionários devido aos critérios adotados e, portanto, os professores devem tomar cuidado na criação deles. (ROSA; ESKE-NAZI, 2013);
- É importante passar mais tempo no desenvolvimento de competências sobre como entregar um bom feedback (DOMINGUEZ et al., 2012);

- Uma boa preparação e apoio orientações são essenciais para o seu sucesso (DOMINGUEZ et al., 2012);
- Dificuldade de avaliar certos critérios (CHANG; TSENG; LOU, 2012);
- Muito mais complexo orquestrar do que comentários similares feitos pelos professores (KNIGHT; STEINBACH, 2011);
- Quantidade de tempo de aula consumida pelo processo, a baixa motivação dos alunos traduzindo-se na incapacidade de permanecer na tarefa, e as dificuldades em captar resultados de aprendizagem para um processo (CUNHA; FIGUEIRA, 2009);
- Fornecimento aos seus pares de um feedback incorreto e enganoso durante a revisão das atividades (YANG; MENG, 2013);
- A necessidade de flexibilidade nas datas de vencimento (ROURKE et al., 2008);
- Os alunos entretanto relatam consistentemente insatisfação (FERMELIS; TUCKER; PALMER, 2007).

Essas dificuldades representam o ponto de partida do trabalho proposto, tendo em vista que grande parte delas, principalmente as últimas da lista se dá pelo aspecto motivacional dos alunos envolvidos, devido à baixa motivação e insatisfação deles. Com os dados apresentados, podemos afirmar que mesmo que poucos artigos tratem de dificuldades no processo de aplicação da técnica, ainda existem algumas, citadas acima, que podem trazer grandes prejuízos no futuro. Dessa forma, a resposta para a pergunta QP5 é que **houve dificuldades em se utilizar uma abordagem de avaliação por pares nos ambientes online em pelo menos 30% dos artigos analisados.**

3.4 Trabalhos Relacionados

Com a revisão sistemática realizada foi possível realizar um levantamento de alguns trabalhos relacionados com a dissertação proposta. Estes trabalhos relacionados apresentam, de alguma maneira, alguns aspectos e características semelhantes ao trabalho aqui proposto, ou seja, é tratado aqui sobre trabalhos que consideramos relacionados às ideias apresentadas nessa dissertação.

Os trabalhos discutidos aqui envolvem o uso de avaliação por pares em ambientes educacionais. Fomos capazes de encontrar trabalhos de grande significância. Vimos que o uso dessa técnica para avaliar trabalhos de características diferentes é bastante frequente. Podemos citar alguns trabalhos como em (DOMINGUEZ et al., 2012) onde é aplicado a técnica para avaliação de trabalhos em um curso de engenharia. (CHANG; TSENG; LOU, 2012) aplica com o objetivo de avaliar portfólios em estudantes do ensino médio que

curram computação. Finalmente em (KAWAI, 2006) é utilizado a técnica para avaliação de mensagens de voz e cartas para alunos estudantes de língua estrangeira.

No trabalho apresentado aqui nessa dissertação, utilizamos a técnica de avaliação por pares para correções de avaliações escritas em geral, independente de qual tipo de avaliação específica será realizada pelo professor.

Outros trabalhos apresentam algumas ferramentas próprias construídas com o objetivo de suportar a aplicação de avaliação por pares. Em (STERBINI; TEMPERINI, 2013), é apresentada uma ferramenta chamada OpenAnswer que tem como objetivo permitir o suporte a avaliações escritas usando técnicas de avaliações por pares, algo semelhante ao proposto aqui nesse trabalho.

Contudo, essa ferramenta não é integrada ao ambiente, o que implicaria que os usuários tivessem que utilizar mais uma plataforma só para as atividades de avaliações escritas. Em nossa proposta, pretendemos que o uso de avaliação por pares seja incluído direto no ambiente, sem essa necessidade de outros sistemas. Contudo, nossa técnica exige mais tempo para ser implantada do que essa ferramenta.

Outros trabalhos (CUNHA; FIGUEIRA, 2009), (GOULI; GOGOULOU; GRIGORIDOU, 2008), (MIAO; KOPER, 2007) e (TRAHASCH, 2004) também propõe ferramentas. Essas ferramentas suportam recursos além das funções básicas de envio de atividades/comentários dos alunos e classificação/comentário dos trabalhos e apresentação dos resultados, elas também suportam recursos como elaboração individual e colaborativo de atividades; avaliação das atividades de um ou de um grupo de alunos; a colaboração de autores e/ou assessores de forma síncrona ou assíncrona; a apresentação das atividades depois dos comentários dos avaliadores; uma variedade de estratégias para a criação do regime de avaliação aplicada no processo de revisão. São ferramentas poderosas, contudo precisam ser configuradas para utilização e não estão integradas aos ambientes educacionais, tendo os mesmos problemas apresentados em (STERBINI; TEMPERINI, 2013).

Dessa forma o trabalho aqui proposto não apresenta uma ferramenta que utiliza avaliação por pares e sim um modelo de avaliação por pares que poderá ser utilizado nos ambientes educacionais através da solução de implementação proposta aqui nesse trabalho. Claro que o fato de precisar de um processo de implantação do modelo dentro do ambiente, o tempo e custo para aplicar a solução aqui proposta é mais alto do que utilizar as ferramentas propostas acima, contudo, uma vez integrado o modelo ao ambiente, este se tornará mais poderoso e permitirá que mais atividades sejam realizadas de forma integrada.

Seguindo nesse caminho de construir modelos de avaliações por pares, existem alguns trabalhos. Em (KIGOZI et al., 2012), os autores projetaram e modelaram um processo de revisão baseado em avaliação por pares para aprendizagem colaborativa. O modelo é baseado no uso do estudante, apoiando-o pedagogicamente sua aprendizagem. Em

(TOSIC; NEJKOVIC, 2010), é proposto um novo método para avaliação por pares dos alunos com base no conceito de confiança. Outros trabalhos propõe modelos baseados em redes bayesianas como em (STERBINI; TEMPERINI, 2013) e (WANG; VASSILEVA, 2003). Todos esses modelos são implementados de forma própria e utilizados de acordo com suas necessidades, porém não tem como objetivo se integrar com os diversos ambientes educacionais e, portanto, se diferem desse trabalho nesse aspecto.

Finalmente, podemos citar o trabalho dos fundadores do Coursera, uma das maiores plataformas online de cursos do mundo, em seu artigo (PIECH et al., 2013). O Coursera utiliza fortemente o conceito de avaliação por pares para prover cursos online para milhares de pessoas em todo o mundo. Contudo, em seu trabalho, é apresentado apenas partes dos algoritmos de reputação e cálculo das notas finais, não propondo/não comentando sobre possíveis modelos de avaliação que eles utilizam, deixando seu trabalho limitado.

4 A PROPOSTA

A proposta apresentada nessa dissertação é um modelo de avaliação por pares juntamente com técnicas de gamificação com o objetivo de facilitar a inclusão de questões discursivas em ambientes educacionais online. Este modelo, como discutido nas seções anteriores, oferece uma solução promissora trazendo benefícios tanto aos professores (redução de uma sobrecarga de trabalho) bem como aos alunos (maior engajamento, aprendizagem, crescimento pessoal, entre outros), além de reduzir possíveis custos das correções textuais.

Nas subseções desse capítulo, será explicado com mais detalhes o conteúdo da proposta, iniciando com a apresentação do modelo em si e seu comportamento junto ao ambiente educacional na seção 4.1. Em seguida, na seção 4.2 será mostrada uma implementação do modelo em um ambiente educacional específico.

4.1 O modelo de avaliação por pares

Esta seção tem como objetivo descrever o modelo de avaliação por pares proposto. Porém, antes de falar sobre o modelo em si, precisamos entender um pouco sobre como se comporta o fluxo de avaliações em ambientes educacionais online e como o modelo proposto se comportaria dentro dele. Esse fluxo bem como o comportamento do modelo será descrito na seção 4.1.1. Em seguida o modelo proposto será apresentado na seção 4.1.2.

4.1.1 Fluxo de avaliação em ambientes educacionais online x modelo proposto

Um ambiente educacional tem como principal objetivo ensinar conceitos aos alunos, ou seja, ensinar alguma(s) disciplina(s) que, por sua vez, é composto de diversos assuntos. Um professor tem como função básica inicialmente a preparação do ambiente para que o aluno possa utilizá-lo. Essa preparação geralmente se resume na inclusão de materiais didáticos, questões, atividades e plano de avaliações. Geralmente os materiais inseridos são para suprir as dificuldades dos alunos. Após esse processo inicial, uma das principais tarefas dos professores, ao longo do curso, é identificar as necessidades individuais de seus alunos, oferecendo ajuda personalizada e direcionamento.

Essa identificação pode ser feita pelos professores e/ou algoritmos especializados e é importante para permitir que o plano de avaliação seja criado adequadamente. Geralmente é feita a partir da aplicação de provas (simulados) com questões de múltiplas escolhas de forma tradicional ou online (no próprio ambiente, se este suportar). Com os resultados dessa prova, o professor conseguirá analisar as dificuldades de seus estudantes. Após essa identificação, é necessário a criação de um plano de avaliação.

O plano de avaliação é onde as atividades e os recursos (as questões ou qualquer outro tipo de material) deverão ser criados e inseridos no ambiente, de forma planejada. As questões criadas deverão suprir as dificuldades identificadas na primeira etapa, para que haja um máximo de aprendizado. Após isso, o professor deve treinar/capacitar os alunos para usar o sistema. Após a inserção destas e a capacitação, os alunos irão aprender com os materiais disponibilizados e responder as questões disponíveis. O sistema corrige as atividades, apresentando um feedback tanto ao aluno quanto ao professor. Este, por sua vez, pode reavaliar seu plano de avaliação e alterá-lo se for necessário.

Contudo, quando se trata de questões discursivas (abertas), os ambientes educacionais online atuais não conseguem fornecer um feedback automático de maneira satisfatória e, portanto, todas as atividades desse tipo exigiriam um trabalho de correção manual do professor, gerando uma sobrecarga e um possível atraso no plano de avaliação. Neste sentido, o modelo proposto neste trabalho visa corrigir essas atividades de através de um processo semi-automático, eliminando o trabalho do professor, através das correções pelos próprios alunos.

A Figura 18 apresenta uma visão geral da inclusão do modelo proposto nesse fluxo de avaliação. Os primeiros três passos, conforme a Figura 18, são as etapas de planejamento, que antecedem o uso de avaliação por pares propriamente dito no ambiente. São essas atividades que irão deixar o ambiente pronto para o seu funcionamento. Os últimos três passos já pertencem a etapa de execução do planejamento.

Conforme explanado anteriormente, as primeiras etapas são bem semelhantes. A primeira etapa é a configuração do ambiente pelo professor. Esta inclui a criação de um curso (disciplina, assuntos), a inclusão dos materiais do curso, inserção dos alunos, entre outras atividades pelo professor. Em seguida, é necessário que haja uma identificação das necessidades dos estudantes, aplicando um teste inicialmente.

Na próxima etapa - plano de avaliação - é onde as atividades e os recursos (as questões discursivas ou qualquer outro tipo de material) deverão ser criados e inseridos no ambiente. Vale ressaltar que, no caso de avaliações escritas, para cada atividade ou recurso criado, é necessário criar um formulário de avaliação, que será disponibilizado para os alunos. É através desse formulário que os alunos irão avaliar seus pares.

Após o planejamento ser concluído, o próximo passo é explicar sobre o funcionamento do modelo de avaliação por pares aos alunos e sua importância, explicar os critérios de avaliação (que estão definidos no formulário de avaliação, criado na etapa anterior), bem como assegurar que os estudantes estejam aptos a usar o recurso, através de treinamentos e capacitações (geralmente feito em tutoriais explicativos).

A partir desse momento, os recursos já estão disponíveis no ambiente e os alunos já estão aptos a utilizá-los. Os recursos serão disponibilizados mediante a sequência definida no plano de avaliação. O processo de distribuição e correção será feito através do modelo de avaliação por pares proposto (a ser melhor detalhado na seção 4.1.2) e os resultados

Fluxo de avaliação ambientes online + modelo

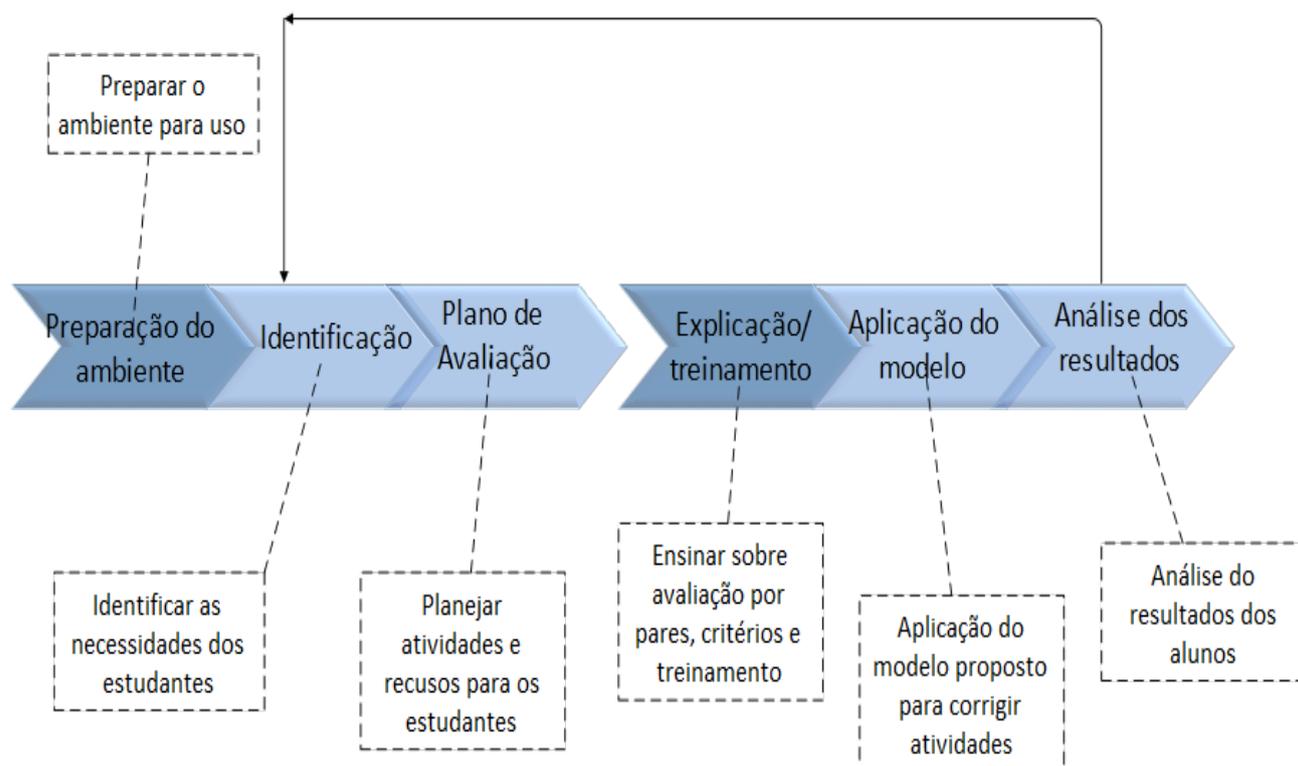


Figura 18 – Fluxo avaliação em um ambiente educacional + modelo

Fonte: elaborado pelo autor.

obtidos serão analisados pelo professor, a fim de que sejam identificadas as necessidades individuais de seus alunos, e este possa oferecer ajuda personalizada e direcionamentos, bem como alterações nos plano de avaliação.

4.1.2 Modelo proposto

Na subseção anterior, o modelo foi apresentado de um ponto de vista geral com base em seu comportamento dentro do fluxo educacional de ambientes online. Esta seção visa descrever o modelo propriamente dito, isto é, todas as etapas que o compõe de uma forma mais detalhada.

Na Figura 18, a quinta etapa era usar o modelo de avaliação por pares proposto como forma de avaliação em um ambiente online. Podemos detalhar esta etapa em diversas atividades que deverão serem feitas para efetivamente usar o modelo. Dessa forma, podemos quebrar a etapa em cinco outras atividades, conforme pode ser visto na figura 19.

O uso do modelo de avaliação por pares se inicia logo após a disponibilização dos recursos aos alunos no sistema e sua devida capacitação/treinamento a respeito do funcionamento da avaliação por pares. Os recursos são sequenciados de acordo com o plano de

Detalhe sobre o modelo dentro do fluxo de avaliação



Figura 19 – Atividades referentes a etapa de uso do modelo de avaliação por pares

Fonte: elaborado pelo autor.

avaliação. A primeira etapa consiste em definir o recurso atual que será utilizado. Esta decisão pode vir diretamente do plano de avaliação ou especificada pelo professor.

O recurso atual será então disponibilizado para os alunos, e estes por sua vez, irão respondê-los e enviar as suas respostas ao sistema, na segunda etapa - a submissão de atividades. O prazo dedicado a estas submissões é fixo e pré-definido pelo professor na etapa de plano de aula.

Em seguida, os próprios alunos receberão algumas atividades (de seus pares) e irão corrigi-las, usando como base o formulário de avaliação criado pelo professor para aquela atividade. O prazo para correção também é pré-definido pelo professor. É nessa terceira etapa, a revisão de atividades, que os alunos atribuem as notas aos seus pares, sem necessariamente conhecer o autor da atividade e nem as notas dos outros alunos.

Após a terceira etapa, o sistema deverá gerenciar as notas obtidas de cada avaliador em cada atividade e calcular a nota final a ser dada para aquela atividade do aluno, baseado nas notas obtidas pelos alunos, na quarta etapa. É nesta etapa também que será apresentada a nota final aos alunos em suas respectivas atividades, em forma de relatórios. Por fim, na quinta e última atividade, os alunos, com a posse do feedback dado pelo sistema, irão visualizar e identificar os seus pontos fracos e refletir-se sobre os erros cometidos naquela atividade.

Este ciclo continua para as outras atividades definidas no plano de avaliação, até

que todas as atividades já tenham sido feitas. Todo esse processo é mantido em seu funcionamento através da utilização de técnicas de gamificação, que tem como objetivo a motivação dos usuários presentes. Para entender melhor como o modelo se comporta dentro do sistema, a Figura 20 apresenta o fluxo de execução que deve ser incrementado no ambiente educacional para incluir o suporte a atividades discursivas, permitindo que os alunos possam utilizá-la, e como será o seu funcionamento no sistema.



Figura 20 – Fluxo de execução do modelo em detalhes

Fonte: elaborado pelo autor.

Como é possível ver na Figura 20, o primeiro passo é a criação das atividades discursivas feitas pelo professor juntamente com a criação e definição do formulário de avaliação, contendo os critérios no qual as atividades serão avaliadas pelos alunos. Uma atividade discursiva é composta de um enunciado e uma ou mais questões (alternativas abertas), onde o aluno irá responder. Um formulário de avaliação por sua vez é composto de um título e um conjunto de questões para avaliação da atividade e cada questão contém um ou mais critérios de avaliação. Esses critérios podem ser:

- **Booleano:** É um critério que somente aceita valores booleanos, isto é, verdadeiro ou falso. Útil para perguntas de eliminação;

- Grupo fixo: Tipo de critério semelhante ao booleano, porém aceita um determinado grupo de valores fixos. É um critério de seleção de valores;
- Intervalo: Neste tipo de critério, usado somente com números, é definido apenas um valor mínimo e um valor máximo permitido. Sendo assim, qualquer valor entre o mínimo e o máximo será permitido;
- Texto: Critério livre, aceitando qualquer valor em texto. Usado muito para comentários.

As atividades e os formulários deverão, em seguida, serem enviados (cadastrados) ao sistema pelo professor em uma interface específica para isso. O sistema por sua vez irá disponibilizar as atividades aos alunos de acordo com o plano de avaliação pré-definido.

A partir da disponibilização da atividade ao aluno para ser respondida, inicia-se o período de submissões - segundo passo. Esse período também está definido no plano de avaliação. Os estudantes responderão as atividades no próprio ambiente online. Para isso, o aluno deverá acessar o sistema e selecionar a atividade, que será disponibilizada juntamente com os outros recursos/atividades. Após a escolha da atividade (seja feita pelo sistema ou pelo usuário), o sistema a apresentará para o usuário, que irá responder com o seu texto.

Em seguida, o aluno submeterá sua resposta para o ambiente e este será responsável pelo gerenciamento da correção desta. No terceiro passo, em primeiro lugar, a gerência do sistema identifica os alunos que irão corrigir a atividade submetida e, a partir desse momento, o sistema enviará a atividade para estes alunos. Existem diversos trabalhos que indicam algoritmos para realizar essa distribuição (MOREIRA, 2014), (CAVALCANTI, 2008) e (SUNG et al., 2005). No final, esse processo de escolha de usuários deverá ser feito através do seguinte algoritmo (criado pelo autor):

1. Defina o número mínimo de correções necessárias, chamado de numMin;
2. Defina uma margem de folga para a busca de usuários, chamado de numFolg;
3. Defina um número máximo de atividades de correção pendentes para um único usuário, chamado de numMax;
4. Recuperar uma lista de usuários de tamanho (numMin + numFolg) que fizeram a mesma atividade e ainda possui avaliação pendente, em ordem crescente de atividades atribuídas;
5. Verificar a quantidade de avaliações pendentes para estes usuários (para não exceder o numMax);
6. Remover da lista de usuários aqueles que já possuem o número de avaliações pendentes numAvaPen \geq numMax;

7. Se o tamanho da lista for maior ou igual a numMin, retorne os numMin primeiros elementos da lista
8. Senão, recuperar usuários com maior probabilidade de corrigir (aqueles que tem muitas atividades feitas e corrigidas), até completar numMin.

Vale ressaltar que o número mínimo de correções necessárias (numMin do Passo 1) para uma atividade é muito dependente do tipo de atividade que está sendo realizada. Teoricamente quanto maior esse número, mais confiável serão os resultados da avaliação. Por outro lado, quanto maior for, mais trabalho será gerado aos pares, podendo ser prejudicial ao resultado final do modelo. Portanto existe um trad-off na escolha desse número. O número da margem de folga (numFolg do Passo 2) serve como uma reserva de usuários que serão recuperados para análise. Quanto mais usuários forem carregados, mais lento se tornará o algoritmo, porém garantirá maior precisão. Por sua vez, o número máximo de atividades de correção pendente a um único usuário (numMax do Passo 3) também deve ser escolhido com cuidado, para não sobrecarregar os alunos envolvidos.

Por padrão, esses números foram escolhidos de acordo com alguns estudos na literatura. O número mínimo escolhido foi 2. O número de folga foi 1 e o número máximo de atividades pendentes foi escolhido como 3. A justificativa foi de que com 2 correções já é possível avaliar os resultados se eles forem semelhantes. Caso haja divergência, um terceiro corretor é associado. A nota final será a média dos dois corretores que mais se aproximarem. Por outro lado, o número máximo de atividades pendentes foi escolhido como 3 pelo fato de que esperamos que um usuário corrija pelo menos 3 redações, sendo no mínimo recomendável de duas redações, para que o modelo funcione adequadamente.

Em seguida, no quarto passo, os estudantes, de acordo com a distribuição na etapa anterior, por sua vez, irão avaliar as respostas para as atividades apresentadas para serem corrigidas, baseado no formulário de avaliação pré-definido e disponibilizado pelo professor. O prazo definido para a revisão (correção) das atividades no sistema é chamado de período de avaliação. É neste período que os alunos avaliarão seus pares e enviarão as notas para o sistema.

Para atribuir sua nota, o estudante receberá as respostas de um determinado aluno e o formulário de avaliação com os critérios definidos. Em seguida, irá analisar a atividade sob o ponto de vista dos critérios (assim como um professor faria) e em seguida basta preencher o formulário com suas respostas. Esse processo será repetido enquanto houverem atividades existentes para a correção para aquele aluno. De acordo com o algoritmo apresentado, esse número tende a ser um média entre os alunos, para não sobrecarregar nenhum. Porém, é possível que haja pequenas variações.

No quinto passo, para se ter uma maior confiança na nota final, deverão ser avaliados a reputação geral de cada avaliador e o nível de competência de cada avaliador em cada critério. O cálculo da reputação geral de um avaliador deverá levar em consideração as

seguintes atividades realizadas por um aluno (onde + significa um aspecto positivo e - significa um aspecto negativo):

- + Acesso freqüente ao sistema;
- + Alto número de atividades discursivas realizadas;
- + Alto número de correções de atividades feitas;
- – Baixa freqüência de acesso ao sistema;
- – Alto número de correções pendentes
- – Alto número de atividades feitas com baixo número de correções

Para se calcular o nível de competência do avaliador em um determinado critério deve-se olhar seu histórico de atividades feitas que utilizaram aquele critério. Caso não possua dados, seu nível será considerado neutro. Caso possua dados no histórico, este deverá ser utilizado para determinar se o avaliador possui um aspecto positivo e/ou negativo a depender das notas dos critérios. A partir dos cálculos definidos, o sistema calculará a confiabilidade de cada avaliação, o nível de competência do avaliador nos critérios e poderá então calcular o resultado final.

Como o objetivo deste trabalho não é o cálculo da reputação e confiança dos usuários do sistema, mas este cálculo tem relevante importância no processo, buscamos estudos na literatura sobre esse tema. Como resultado, definimos que utilizaríamos o cálculo proposto por (PIECH et al., 2013), baseado em modelos probabilísticos. Seleccionamos este modelo de cálculo pois ele foi concebido com propósitos semelhantes ao do nosso modelo e atualmente é utilizado no Coursera, um dos ambientes de educação online com milhares de usuários. Também existem outros modelos de reputação que podem ser utilizados como os apresentados em (WANG; VASSILEVA, 2003), (WANG; VASSILEVA, 2005)

Em resumo, foi utilizado o cálculo PG1 para o cálculo de "Grader bias" e confiabilidade. Nesse caso, é assumido uma distribuição de probabilidade a priori para as variáveis latentes e utiliza a premissa de que o bias de um aluno individual pode ser diferente de 0, porém a média de muitos avaliadores sempre tende a zero. Em resumo, as formulas utilizadas são:

(Confiabilidade) $\tau_v \sim \mathcal{G}(\alpha_0, \beta_0)$ para todo avaliador v ,

(Bias) $b_v \sim \mathcal{N}(0, 1/\eta_0)$ para todo avaliador v ,

(Nota verdadeira) $s_u \sim \mathcal{N}(\mu_0, 1/\gamma_0)$ para todo usuário u e

(Nota observada) $z_u^v \sim \mathcal{N}(s_u + b_v, 1/\tau_v)$ para toda nota de par observada,

onde \mathcal{G} refere-se a distribuição gamma com hiperparâmetros α_0, β_0 fixados, enquanto que η_0 e γ_0 são hiperparâmetros para a distribuição a priori de biases e nota verdadeira,

respectivamente. Maiores informações de como essas formulas foram obtidas podem ser encontradas no site do artigo¹.

Finalmente, no sexto passo, o sistema deverá gerar um relatório contendo o resultado final da avaliação que será apresentado ao aluno que submeteu a atividade. Este relatório deverá ser detalhado, apresentando as notas de cada critério, bem como a nota final da atividade. É possível a inclusão de comentários e possíveis alterações nas notas por parte do professor antes de sua apresentação ao aluno. Este por sua vez irá identificar os pontos fracos e usar o sistema para aprender mais, através dos materiais do curso, realizando novas atividades e/ou corrigindo as atividades dos outros estudantes.

A partir das informações apresentadas acima a respeito do fluxo de execução e posicionamento do modelo dentro do fluxo educacional, podemos resumir o modelo apresentado em um conjunto de etapas necessárias para que o processo de avaliação por pares funcione adequadamente. Essas etapas estão ilustradas na Figura 21.

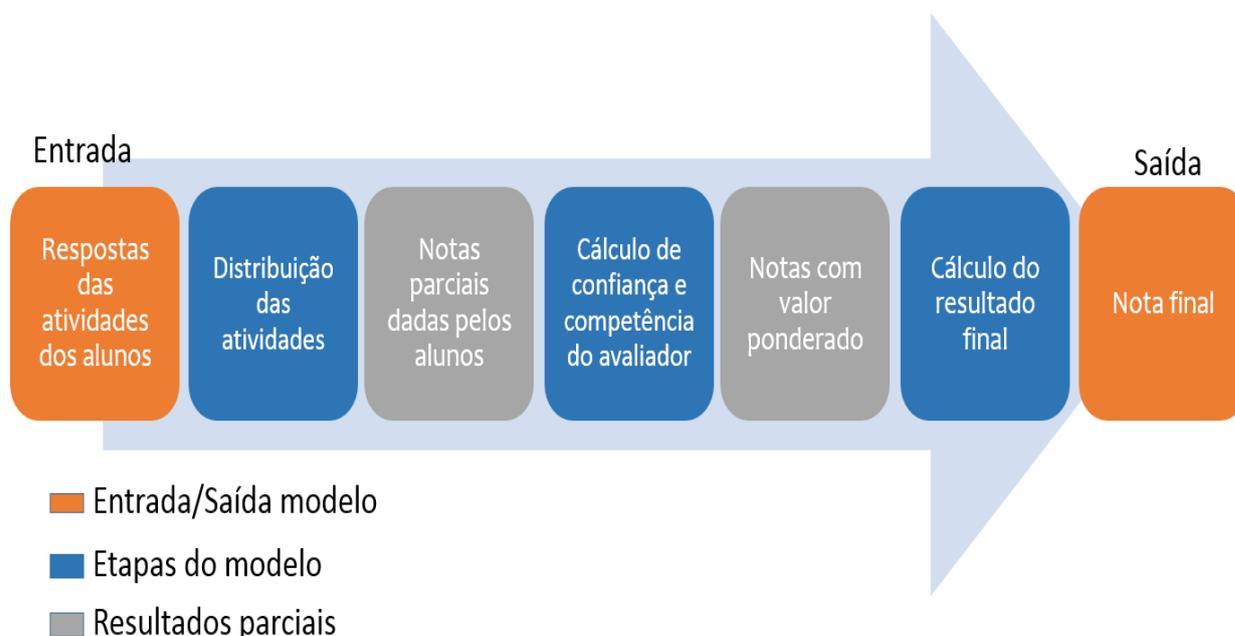


Figura 21 – Etapas do modelo

Fonte: elaborado pelo autor.

Podemos observar na figura apresentada que o modelo recebe como entrada o conjunto das respostas das atividades dos alunos. O primeiro passo é determinar, conforme as condições já apresentadas acima, a quem cada resposta de atividade será atribuída para correção (distribuição das atividades). Isso gerará, como resultado parcial, uma lista de notas parciais para cada resposta dada pelos alunos. Essa lista servirá como entrada para a próxima etapa do modelo, que tem como objetivo calcular a confiança e a competência do avaliador, atribuindo um peso para a nota. Nesse sentido, teremos como segundo

¹ Disponível em <http://web.stanford.edu/~cpiech/bio/papers/appendices/edm13_appendix.pdf>

resultado parcial a lista de notas com valor ponderado. Em sequência, a última etapa do modelo é calcular o resultado final para cada resposta de atividade dado as notas ponderadas, gerando assim, a saída do modelo com estes resultados finais.

4.1.2.1 A gamificação no modelo

As seções anteriores descreveram o modelo sob diversos aspectos, bem como os detalhes de cada etapa envolvida nele. Esta seção tem como objetivo descrever como a gamificação é incluída no modelo. Para incluir gamificação é necessário criar um plano de gamificação, conforme a seção 2.2.3. O plano de gamificação construído e apresentado abaixo foi baseado no framework de gamificação 6D, no qual foi apresentado na seção 2.2.3.

Em primeiro momento é preciso definir os objetivos de negócio da gamificação, isto é, qual é o nosso objetivo final da inclusão da gamificação no modelo proposto? Em primeiro lugar, estamos gamificando o modelo de avaliação por pares devido à uma limitação de motivação existente nele, o que poderia prejudicar os resultados. O principal objetivo da gamificação é fazer com que os alunos se sintam motivados a participar tanto realizando atividades quanto corrigindo-as. Além disso, como objetivos finais secundários, podemos citar:

1. Fazer os estudantes aprenderem mais;
2. Para motivar o aluno a continuar (ou de volta para) o sistema de solução de atividades;
3. A fim de proporcionar aos alunos uma maneira motivadora de estudar.

Após, precisamos delinear os comportamentos alvos (as coisas que nós queremos que os usuários façam), tendo como foco os objetivos finais pré-definidos. Vale ressaltar que esses comportamentos foram definidos por experiência do autor com gamificação, não tendo embasamento científico para tais decisões. Todos esses comportamentos são a tendência do que seria ideal para o usuário agir e assim beneficiar tanto ele quanto o modelo de avaliação por pares construído. Os comportamentos definidos são:

1. **Para o objetivo do item 1:**

- a) Estimular os estudantes a participar com mais frequência das atividades;
- b) Desafiar estudantes para conseguirem algum tipo de troféu.

O ato de participar com mais frequência das atividades a serem realizadas indica uma maior capacidade para o aluno aprender e conseqüentemente atingir o objetivo em questão. Desafiar tais estudantes para conseguirem alguns troféus, principalmente aqueles relacionados a atividades de aprendizagem irá incetivar os alunos a participarem mais.

2. Para o objetivo do item 2:

- a) Prover medalhas aos estudantes que participam das atividades;
- b) Prover troféus ou pontos aos alunos que responderem/corrigirem atividades;
- c) Promover um ranking parametrizado para os alunos.

O sistema de solução de atividades inclui realizar e corrigir redações dentro do ambiente. Essas tarefas tendem a se tornar cada vez mais um aspecto desmotivacional para os alunos. Nesse sentido, através da promoção de medalhas aos estudantes, troféus e enumerá-los em rankings aqueles que mais fizerem as atividades servirá como um aspecto motivacional para aqueles alunos que ficarem desmotivados, incluindo até aqueles que por algum motivo desistiram de realizar tais atividades (trazendo-o de volta ao sistema de avaliação, na tentativa de alcançar o objetivo 2).

3. Para o objetivo do item 3:

- a) Estimular a competição dos estudantes, com recompensas aos vencedores;
- b) Recompensar os estudantes por suas ações no sistema de forma inesperada.

O objetivo 2 tenta motivar o estudante a estudar. Como tendência os alunos em um sistema de avaliações sempre competem com seus amigos. Nesse sentido, temos como um comportamento alvo estimular a competição, tendo em vista que uma competição saudável leva a resultados melhores entre os alunos. Dessa forma, aqueles alunos que melhor se saírem na competição irão receber recompensas. Além disso, ações inesperadas no sistema também podem recompensar os alunos, o que irá motivar ainda mais a sua participação.

O terceiro passo do 6D é definir os jogadores. No nosso caso os usuários serão estudantes de cursos online. Em geral não possuem nenhum tipo de relacionamento entre si. Podemos classificar o perfil desses usuários segundo o modelo proposto em (KIM, 2014) como:

- **Exploradores:** são motivados pela aquisição de conhecimento, explorando fronteiras, encontrando brechas, e conhecendo as regras que regem o espaço. Eles gostam de acumular e mostrar conhecimento;
- **Criadores:** são motivados por oportunidades de auto-expressão. Criadores amam ferramentas e sistemas que lhes permitem personalizar a sua experiência, fazem a sua marca, e expressam sua singularidade. Eles gostam de personalizar fundos, fontes e avatares. Eles buscam status, reconhecimento e influência através da habilidade criativa;

- **Competidores:** são motivados por testar suas habilidades e ver como eles se comportam. Gostam de desenvolver suas habilidades, mostrar o seu talento, e saber onde estão dentro de um grupo. Eles valorizam maestria, aprendizado e construção de relacionamento via competição amigável;
- **Colaboradores:** são motivados por trabalhar com os outros para um objetivo maior. Eles gostam de "vencer juntos" e medir o sucesso como impacto coletivo. Desfrutam de participantes em grupos e equipes e formação de parcerias. Eles valorizam o trabalho em equipe, a aprendizagem compartilhada e construção de relacionamento via tarefas compartilhadas.

Os principais jogadores do nosso modelo seriamos exploradores e principalmente os colaboradores. Os exploradores por que eles tendem a realizar as todas as ações no sistema, uma vez que gostam de utilizar todos os recursos envolvidos. Os colaboradores, por sua vez, porque eles tem o instinto de ajudar seus amigos, o que é um ponto muito forte e necessário para que um sistema de avaliação por pares funcione de forma adequada.

Em seguida, precisamos definir os loops de atividade ². Basicamente um Loop de Atividade (LA) possui três componentes principais: motivação, ação e recompensa. A motivação consiste no objetivo do loop, isto é, a sua finalidade para ser criado. A ação consiste no comportamento desejado para o usuário. E finalmente, a recompensa é o que o usuário irá ganhar caso ele possua o comportamento desejado.

Um loop de atividade pode ser um Loop de Engajamento (LE) ³ ou um Loop Progressivo (LP) ⁴. A diferença é que o progressivo possui uma ideia temporal, onde o usuário vai realizando o loop aos poucos enquanto que o loop de engajamento o usuário o realiza na hora de forma inesperada. A Tabela 5 apresenta os loops definidos no nosso modelo de gamificação para atingir o objetivo. Vale ressaltar que o tipo refere-se ao tipo do loop criado e a motivação refere-se aos objetivos mais o comportamento desejado (que estão definidos anteriormente), por exemplo, 1A refere-se ao objetivo 1 comportamento "a".

A etapa 5 do framework 6D nos diz para não esquecer a diversão. Como forma de diversão temos incrementado os loops 7 e 10, onde os usuários ganharão pontos de forma inesperada por realizar algumas de suas ações no sistema além de estimular uma competição em si mesmo para obter melhores posições nos rankings e troféus e medalhas.

Finalmente, na última etapa é preciso definir os elementos da gamificação que serão utilizados. Como visto na Tabela 5, foram utilizados os seguintes elementos de gamificação: pontos, troféus, missões, rankings e medalhas. Todas as informações de como consegui-los estão detalhadas na mesma tabela.

² Do inglês "Activity Loop". Passo 4D do framework de gamificação

³ Do inglês "Engagement loop"

⁴ Do inglês "Progressive loop"

Tabela 5 – Loops de atividades definidos no modelo de gamificação.

Número	Tipo	Motivação	Ação	Recompensa
1	LP	1A	Realizar 3 atividades	Pontos
2	LP	1A	Corrigir 3 atividades	Pontos
3	LP	1A	Logar por 3 dias consecutivos	Pontos
4	LE	1A,2B	Responder atividade	Pontos
5	LE	1A,2B	Corrigir atividade	Pontos
6	LE	1A,2B	Logar	10 Pontos
7	LE	3B	Ação inesperada	Pontos
8	LP	1A,1B,2B	Responder 5 atividades	Troféu “O escritor”
9	LP	1A,1B,2B	Corrigir 5 atividades	Troféu “O avaliador”
10	LE	2C, 3A	Ficar entre os 10 primeiros no ranking	Troféu “No topo”
11	LE	2A	Ficar entre os três primeiros de uma atividade	Medalha Ouro, Prata ou Bronze

Fonte: elaborado pelo autor.

4.2 Implementação do modelo de avaliação por pares

A Seção 4.1 apresentou o modelo de avaliação por pares junto com gamificação proposto neste trabalho. Este modelo foi implementado em um ambiente educacional para que possa ser visto do ponto de vista computacional. Diante disso, esta seção tem como objetivo apresentar o processo de implementação do modelo proposto no ambiente educacional.

Em primeiro momento será apresentado o ambiente e suas necessidades na Seção 4.2.1. Em seguida, será descrito o modelo implementado na Seção 4.2.2. E finalmente, a integração do modelo com o ambiente será descrito na Seção 4.2.3.

4.2.1 O ambiente educacional MeuTutor

O modelo foi implementado e integrado ao ambiente educacional MeuTutor. Este foi escolhido como um bom cenário devido à sua flexibilidade de implementação e integração de novos componentes no sistema e por possuir um ambiente já preparado para gamificação, tornando mais fácil a integração do modelo ao sistema. Além disso, este ambiente tinha uma necessidade específica (com provas de redação) que será melhor detalhada na Seção 4.2.3.

O ambiente educacional MeuTutor ⁵ é um produto da empresa MeuTutor, sediada na incubadora de empresas da Universidade Federal de Alagoas. Podemos defini-lo, de acordo com as definições apresentadas na Seção 2.3, como um sistema tutor inteligente, que tem como objetivo acompanhar a aprendizagem dos alunos de forma personalizada, garantindo uma qualidade no ensino e melhorando o desempenho dos seus usuários. A versão escolhida a ser utilizada foi o MeuTutor-ENEM, que visa auxiliar alunos do ensino médio a se prepararem para o Exame Nacional do Ensino Médio (ENEM). Com isso,

⁵ A plataforma está disponível através do link: www.meututor.com.br

o ambiente oferece cursos referentes às disciplinas do ensino médio, como português, matemática, física, entre outras.

Como diferencial do MeuTutor, podemos citar: i) a utilização de teorias pedagógicas e computacionais para a personalização em larga escala da aprendizagem promovendo o ensino de qualidade, como a Teoria de Resposta ao Item (FLETCHER, 1994) e Mineração de Dados Educacionais (MERCERON; YACEF, 2005); ii) a gamificação da aprendizagem, de modo a tornar o estudante sempre motivado a aprender e interagir no ambiente; iii) a rápida disponibilização para diferentes domínios de aprendizagem e iv) a autogestão da aprendizagem utilizando conceitos de inteligência nos negócios.

O ambiente MeuTutor-ENEM foi construído tendo como ponto forte três bases: ambiente gamificado, aprendizagem personalizada e experiência social. O primeiro, como já explanado ao longo da dissertação tem como objetivo motivar o aluno a continuar seus estudos no ambiente. O objetivo principal da Gamificação é aumentar o engajamento dos usuários por meio do uso de técnicas semelhantes àquelas presentes em jogos (FLATLA et al., 2011), fazendo com que os usuários se sintam no controle de suas ações e se motivem com as tarefas (PAVLUS, 2010).

Com isso, é possível observar na Figura 22 alguns elementos de Gamificação (jogos) presentes na plataforma, que foi projetada como se fosse um jogo para motivar a aprendizagem e o estudo do aluno. A plataforma disponibiliza um mecanismo de pontos e níveis personalizados, para o aluno evoluir em cada disciplina e no geral (Na figura o aluno está no nível 5). Além de ganhar pontos por meio de interações com o sistema (a maioria das interações do estudante com o sistema é recompensada com pontos), o aluno também pode ganhar um bônus ao completar missões sugeridas. O sistema também oferece vários troféus que recompensam ações específicas no sistema, os estudantes irão descobrir que ações são essas e colecionar.

Melhor do que simplesmente ganhar pontos, passar de nível e colecionar troféus, é poder comparar todas as recompensas com os amigos. No MeuTutor-ENEM, o estudante pode comparar seu desempenho em cada disciplina com os seus amigos, além de competir com eles por meio de rankings.

Sob o segundo aspecto, a aprendizagem personalizada, no MeuTutor-ENEM, as questões e vídeo-aulas se adequam aos alunos. Através de uma base de questões no estilo ENEM, o estudante pode praticar seu conhecimento resolvendo vários problemas de cada assunto nas disciplinas disponíveis no sistema. À medida que ele resolve corretamente um problema, seu progresso no assunto cresce. Se errar a questão, uma vídeo-aula será recomendada para ele suprir sua deficiência.

A aprendizagem personalizada tem por objetivo aplicar métodos, técnicas e tecnologias com vistas a identificar limitações e necessidades específicas de um estudante ou de um grupo de estudantes de modo a personalizar a aprendizagem. Dentre as abordagens conhecidas para a personalização da aprendizagem, o MeuTutor-ENEM, além de

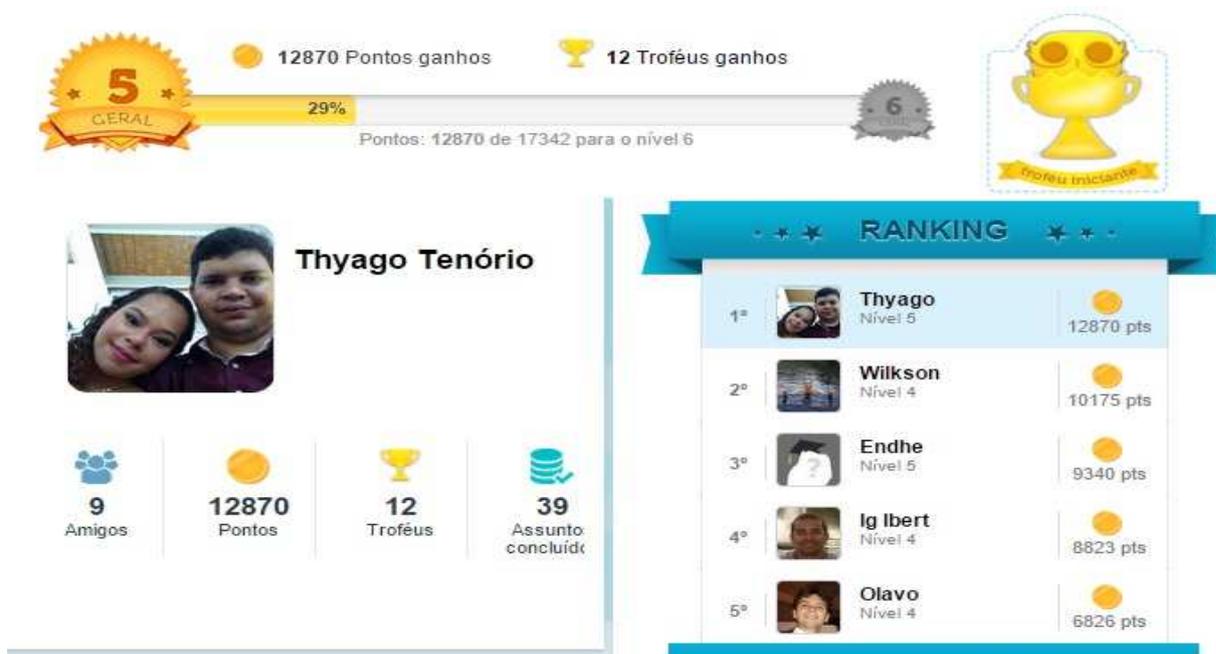


Figura 22 – Elementos de gamificação do MeiTutor

Fonte: elaborado pelo autor.

herdar algumas características por ser um STI, utiliza Mineração de Dados Educacionais (WITTEN; FRANK, 2005).

Com a utilização de técnicas de mineração de dados educacionais, o sistema consegue identificar e sanar alguns problemas educacionais dos alunos. Por exemplo, é possível verificar quais ações pedagógicas são ineficazes e identificar métodos que proporcionem melhores benefícios educacionais ao aluno (PAIVA¹ et al., 2012). No MeiTutor-ENEM, o aluno ao estudar na plataforma, tem seu perfil de interações no sistema minerado com o objetivo de verificar quais os seus pontos fracos (e fortes), identificando, assim, qual o próximo recurso educacional que deverá ser oferecido a ele: seja outro problema para desafiá-lo, seja um vídeo para suprir sua deficiência sobre determinado conceito. A Figura 23 ilustra o momento em que vídeos são recomendados para o aluno quando ele erra uma questão.

O MeiTutor-ENEM também oferece um sistema de provas personalizadas, onde o aluno pode escolher quais assuntos, de quais disciplinas, ele quer testar seu conhecimento. O sistema gera a prova e simula um exame real.

Em cada disciplina disponível no MeiTutor, existe uma árvore de assuntos representando o conhecimento que o aluno deve ter para estar preparado para o ENEM. A tela ilustrada na Figura 24 mostra a árvore de assuntos para a disciplina de Matemática. Quem dita o ritmo da aprendizagem no sistema é o estudante, praticando problemas o progresso dele no assunto aumenta e, ao aprender um assunto, ele consegue desbloquear novos assuntos para estudar. Além disso, através de relatórios no perfil, o estudante pode

The image shows a screenshot of the MeuTutor interface. On the left, there is a user profile for Thyago Tenório with a 'ASSINE JÁ' button. Below the profile are statistics: 9 Amigos, 12870 Pontos, 12 Troféus, and 39 Assuntos concluídos. A 'RANKING' section shows a list of users with their names, levels, and scores. On the right, there is a navigation bar for 'Matemática', 'Matemática Básica', and 'Praticar'. Below this, the 'Matemática Básica' section shows a question titled 'Questão - MultipleChoiceProblem_0_604132427' with a 'Reportar Erro' button and a progress indicator at 100.0%. The question text is: '(Inatel-MG) Nas últimas eleições presidenciais de um determinado país, em que 9% dos eleitores votaram em branco e 11% anularam o voto, o vencedor obteve 51% dos votos válidos. Não são considerados válidos os votos em branco e nulos. Pode-se afirmar que o vencedor, de fato, obteve de todos os eleitores um percentual de votos da ordem de:'. The options are A (47%), B (50%), C (41% << correta), D (44%), and E (38%). At the bottom, there is a 'Resposta Errada' message with a red 'X' icon and a recommendation for videos, along with a 'Próxima Questão' button.

Figura 23 – Tela do MeuTutor ao errar uma questão - vídeos recomendados

Fonte: elaborado pelo autor.

ver quantos assuntos ele completou, quantas disciplinas ele aprendeu, quantos pontos ganhou, quantos troféus conseguiu e muito mais.

Podemos observar que essa personalização automática da aprendizagem do MeuTutor-ENEM vem do uso de técnicas de sistemas tutores inteligentes. O MeuTutor-ENEM é um STI que usa diferentes técnicas para tutorar o aluno dentro da plataforma. Dentre essas técnicas, destacam-se a estratégia pedagógica baseada em algoritmos genéticos (MITCHELL, 1998) e a técnica knowledge tracing (KOBASA, 1994) para verificação da aquisição de conhecimento do aluno.

Por fim, no terceiro aspecto (experiência social), o MeuTutor mantém a preferência dos alunos tradicionais (discutir e estudar um determinado assunto ou conteúdo juntos, construindo conhecimento compartilhado), fornecendo recursos poderosos dentro do ambiente através de técnicas da aprendizagem colaborativa em sistemas computacionais.

Entre os recursos colaborativos providos aos estudantes está a ideia de grupos de estudo. Dessa forma, um aluno pode convidar outros amigos para praticar um determinado assunto. A Figura 25 mostra a tela de grupo de estudos, onde os alunos respondem questões juntos depois de debater e entrar em um determinado consenso.

Por outro lado, o aluno que está na Internet estudando, ou seja, no MeuTutor-ENEM, dificilmente só estará fazendo isso. Provavelmente, ele vai navegar em alguma rede social como o Facebook (acarretando na distração do estudo). Todavia, a plataforma MeuTutor-ENEM é totalmente integrada com o Facebook, permitindo que ele compartilhe suas conquistas na rede social, tornando o Facebook uma ferramenta aliada na hora de estudar.

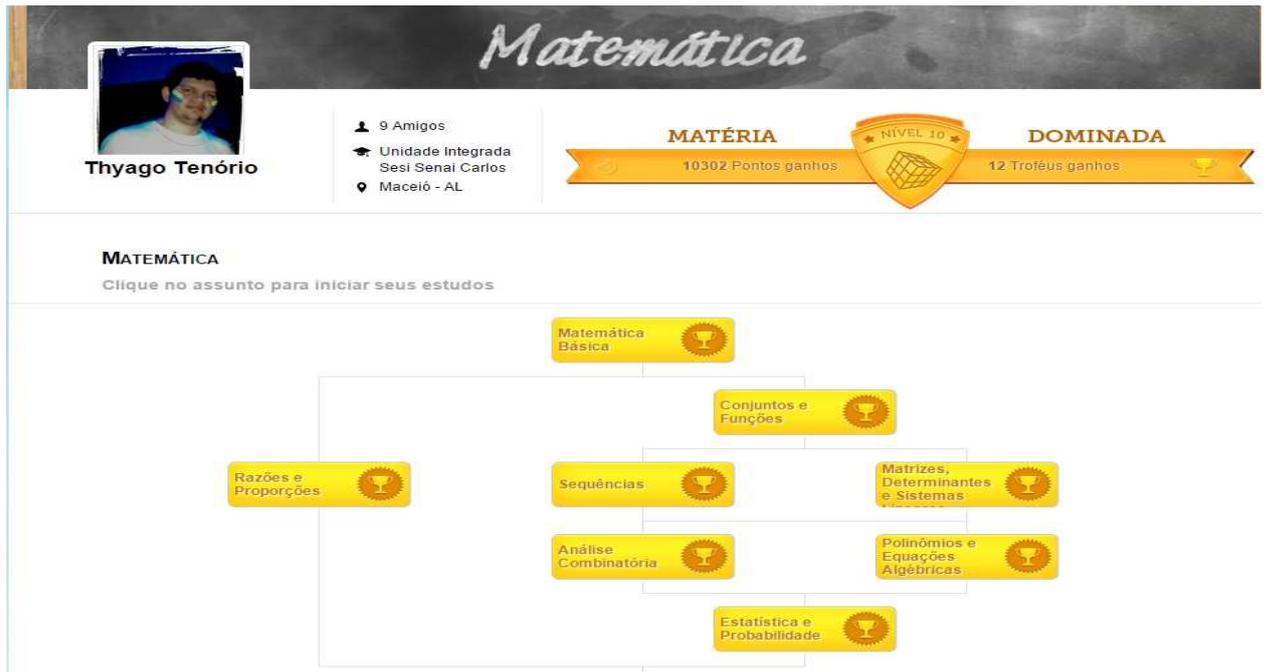


Figura 24 – Árvore do aluno para a disciplina de Matemática

Fonte: elaborado pelo autor.

A Figura 26 mostra o compartilhamento quando o aluno passa de nível na plataforma.

4.2.1.1 Arquitetura do MeuTutor ENEM

O ambiente educacional MeuTutor-ENEM foi construído a partir da instanciação de uma linha de produto de software (LPS)⁶ semântica, sendo portanto, um produto de uma LPS (SILVA et al., 2010). Nesse sentido, todas as suas funcionalidades internas são susceptíveis a variabilidades em sua implementação, o que torna o ambiente bastante flexível. Nesse sentido, por exemplo, o ambiente pode detectar que um aluno não está tendo ganhos consideráveis de aprendizagem ao usar uma determinada estratégia pedagógica e assim, em tempo de execução, ele pode modificar o componente responsável pela estratégia pedagógica para outra.

Com isso, toda a arquitetura do ambiente teve de ser projetada para suportar a necessidade de variabilidade de funcionalidades. Por este motivo, todo o sistema foi implementado baseado em componentes - unidades de softwares que encapsulam uma série de funcionalidades -, utilizando o modelo COSMOS (SILVA, 2003). Basicamente, um componente neste modelo possui interfaces requeridas, interfaces providas e implementação. A integração entre os componentes é feita através da ligação provida-requerida utilizando

⁶ Uma linha de produtos de software é um conjunto de sistemas de software que têm determinadas funcionalidades em comum, sendo desenvolvidos tendo esta mesma base e variando apenas alguns aspectos.

Potenciação
Questão 1

A figura ilustra uma ponte suspensa por estruturas metálicas em forma de arco de parábola. Os pontos A, B, C, D e E estão no mesmo nível da estrada e a distância entre quaisquer dois consecutivos é de 25m. Sabendo-se que os elementos de sustentação são todos perpendiculares ao plano da estrada e que a altura do elemento central CG é 20m, a altura de DH é:

A	17,5m
B	10,0m
C	7,5m
D	12,5m
E	15,0m

Responder questão

Membros deste grupo

- Olavo Holanda - 1ª posição - Resposta C
- Rodrigo Rodrigues - 5ª posição - Resposta C
- Endhe Elias - 3ª posição - Resposta C
- Wilson Eldon - 5ª posição - Resposta C
- Thyago Tenório - 4ª posição - Resposta C

ADICIONE MEMBROS

Chat Grupo de Estudos

Olavo Holanda: This is Photoshop's version of Lorem Ipsum. Proin gravida nibh vel velit auctor aliquet. Aenean sollicitudin, lorem quis bibendum auctor, nisi elit consequat ipsum, nec sagittis sem nibh id elit. Duis sed odio sit amet nibh vulputate cursus a sit amet mauris.

Rodrigo Rodrigues: This is Photoshop's version of Lorem Ipsum. Proin gravida nibh vel velit auctor aliquet.

Wilson Eldon: This is Photoshop's version of Lorem Ipsum.

escreva aqui... enviar

Figura 25 – Estudos em grupo - Experiência social do MeuTutor

Fonte: elaborado pelo autor.

um arquivo de configuração, isto é, um componente prover uma determinada interface que é requerida para outro componente.

Por questões de espaço e por não ser o foco deste trabalho, o diagrama de componentes da plataforma foi omitido, mas a visão geral da arquitetura pode ser vista na Figura 27.

Como podemos ver na figura, houve a utilização do padrão arquitetural Microkernel (BUSCHMANN; HENNEY; SCHIMDT, 2007). Basicamente o padrão consiste na separação das funcionalidades (componentes) do núcleo e os componentes satélites. Todos os componentes que não possuem variabilidade são colocados dentro do Microkernel da arquitetura. Esses componentes são chamados de obrigatórios, pois eles estão presentes em qualquer cenário de configuração do MeuTutor-ENEM (por exemplo, componentes como Login e Cadastro). Por outro lado, todos os componentes que podem sofrer variações ficam de fora do Microkernel (os satélites). Estes são opcionais e vão depender do conector para serem utilizados.

A função do conector é basicamente permitir a variabilidade entre os componentes, provendo um serviço de chaveamento entre qual componente vai ser utilizado em determinado momento, isto é, os conectores são responsáveis por chavear qual componente satélite será utilizado pelo microkernel. Por exemplo, um estudante visualiza uma questão de múltipla escolha que o sistema recomendou, mas não se adapta a este tipo de problema; neste caso, o conector responsável pelo componente de problemas pode mudar o satélite, que antes era múltipla escolha, para problemas de verdadeiro ou falso, fazendo a ligação necessária.

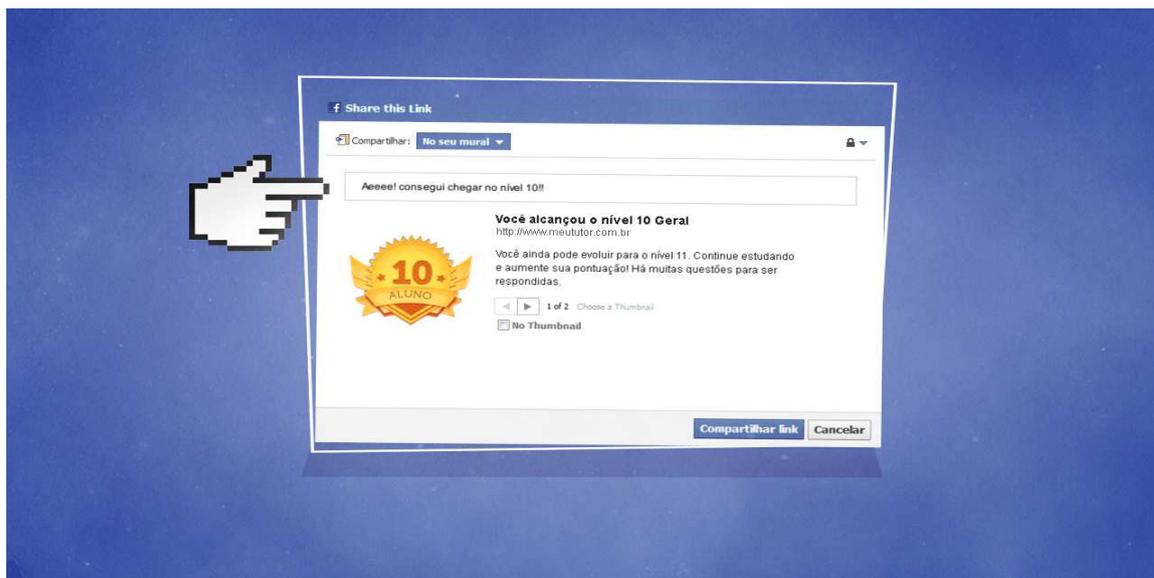


Figura 26 – Compartilhamento das ações do usuário no Facebook

Fonte: elaborado pelo autor.

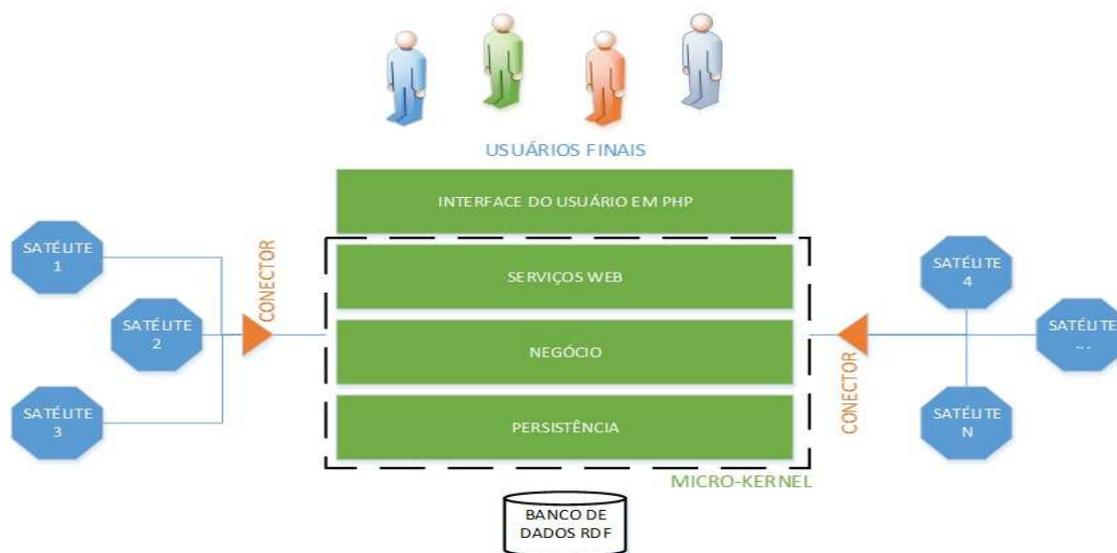


Figura 27 – Visão geral da arquitetura do MeuTutor

Fonte: (HOLANDA, 2014)

Podemos observar na figura que a implementação do Microkernel do ambiente utilizou outro padrão arquitetural: multicamada. Este tem como função separar as responsabilidades entre as diversas camadas da aplicação. Neste caso, o Microkernel foi dividido em três camadas:

- **Camada de Serviços Web:** Esta camada é responsável em implementar e prover todos os serviços web que o ambiente proporciona. Foi construída em Java e utiliza tecnologias de serviços web como WSDL e SOAP para prover serviços;

- **Camada de Negócio:** Encontram todas as funcionalidades que representam a lógica do sistema;
- **Camada de Persistência:** Responsável pelo acesso e comunicação com o banco de dados. O banco de dados utiliza tecnologias semânticas e é constituído de triplas RDF, modeladas por ontologias OWL.

Podemos notar que há uma camada acima de serviços web, a qual tem como funcionalidade criar a interface com o usuário do sistema. Esta camada foi implementada em PHP e utiliza os serviços providos pela camada de serviços web.

4.2.2 Implementação do modelo

Esta seção tem como objetivo descrever a implementação do modelo de avaliação por pares apresentado ao longo deste trabalho. A implementação do modelo sofreu algumas influências tecnológicas vindas do ambiente educacional no qual foi destinado. Entre essas influências estão o uso de ontologias para representar a sua modelagem, implementação do núcleo em Java utilizando o modelo COSMOS, arquitetura em camadas e fornecimento de interfaces de serviços web em WSDL/SOAP, onde serão fornecidas todas as funcionalidades do modelo.

Apesar de ser implementado usando as mesmas tecnologias do ambiente educacional MeuTutor-ENEM, vale ressaltar que o modelo ainda é independente do ambiente, uma vez que qualquer ambiente poderia fazer as chamadas dos serviços providos pelo modelo e adaptar os retornos (Objetos serializados) às suas necessidades. Vale ressaltar que esse trabalho de adaptação será apenas mostrado com relação a interface do MeuTutor-ENEM, o qual foi implementada em PHP.

Antes de mostrar a implementação dos componentes e dos serviços web referentes às camadas apresentadas na Seção 4.2.1.1, será apresentado, em primeiro momento, a modelagem do sistema, a qual se refere a camada de persistência. Para facilitar a compreensão dos elementos envolvidos e suas propriedades, foram criados vários diagramas de entidades e relacionamentos (ER) para representar os diversos elementos envolvidos.

Como explicado no modelo, o professor precisa, em primeiro momento, cadastrar a avaliação escrita e os formulários de avaliação que será usado pelos alunos na hora de avaliar. Neste sentido, será necessário a modelagem de uma questão discursiva bem como a modelagem do formulário de avaliação. Ao mesmo tempo, será necessário a modelagem para suportar as respostas dos usuários às questões bem como ao formulário de avaliação. A figura 28 apresenta o diagrama de entidade relacionamento referente a uma questão do modelo.

Podemos observar na figura que um problema discursivo possui alguns atributos como nível (nível recomendado pelo problema), dificuldade (fácil, médio e difícil), se está ou

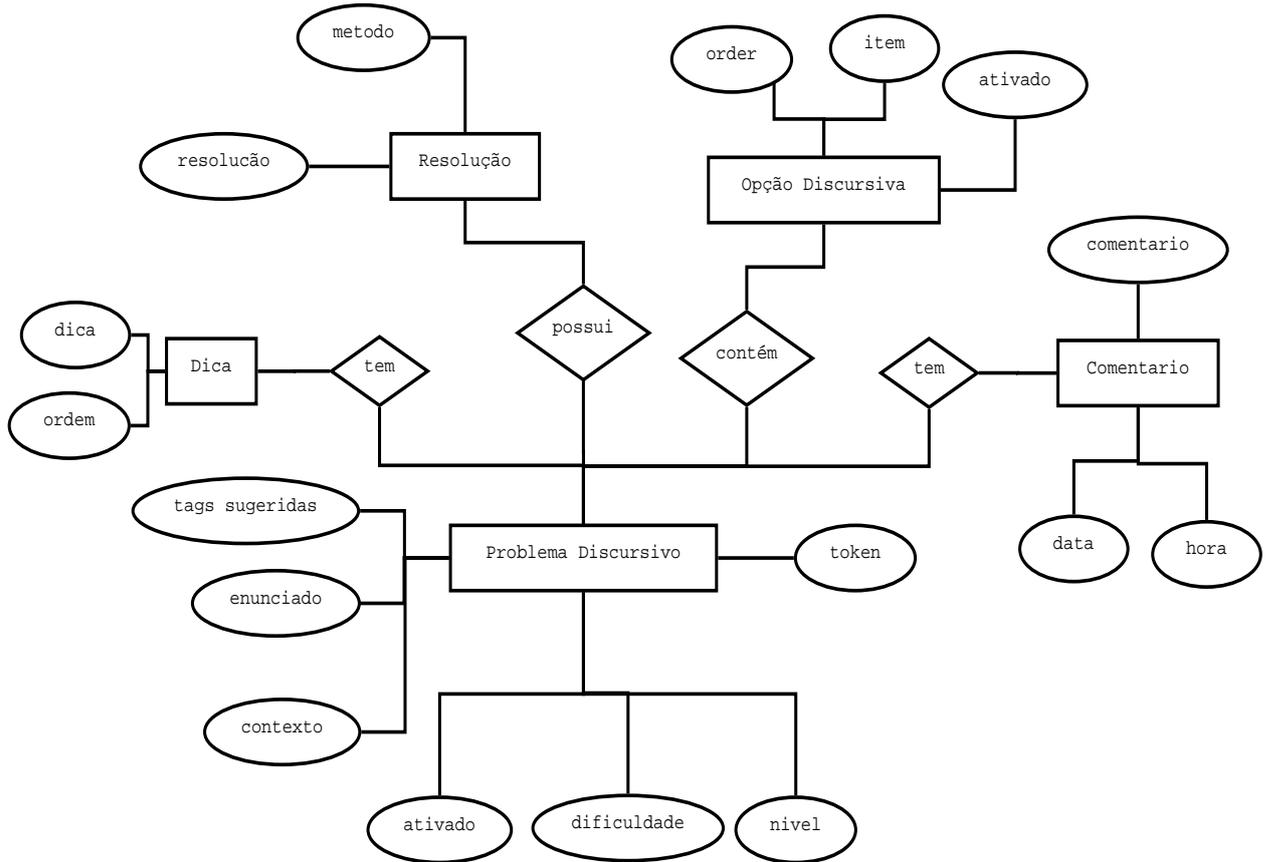


Figura 28 – Diagrama de entidade relacionamento para um problema discursivo

Fonte: elaborado pelo autor

não ativado e disponibilizado, um contexto, seu enunciado (a pergunta ou texto) e algumas tags sugeridas (subassuntos envolvidos). Além disso, um problema tem *Dicas* (que possua a dica e uma ordem de surgimento), possui *Resolução* (a resolução da questão propriamente dita e o método utilizado para responder), e tem *Comentário* (comentários feitos pelos alunos em uma data e hora específica sobre a questão). Por fim, uma questão discursiva possui várias *Opções Discursivas*, o que representa cada opção a ser respondida pelo estudante. Cada *Opção Discursiva* possui seu item (a pergunta em si), uma propriedade ativado que indica se aquela opção irá ou não aparecer ao estudante e a ordem no qual a opção irá aparecer.

Com a modelagem das questões, o próximo passo é modelar o formulário de avaliação que é criado pelo professor e disponibilizado aos alunos para que estes O respondam, julgando as respostas dos outros alunos. A figura 29 apresenta o diagrama de entidade relacionamento de um formulário de avaliação no sistema bem como das respostas do usuário.

Podemos observar na figura que um *Formulário de Avaliação* possui como atributos um título e uma data de criação. Além disso, o formulário é criado por um usuário (usuário nível professor). O formulário contém um conjunto de *Questões*, no qual cada

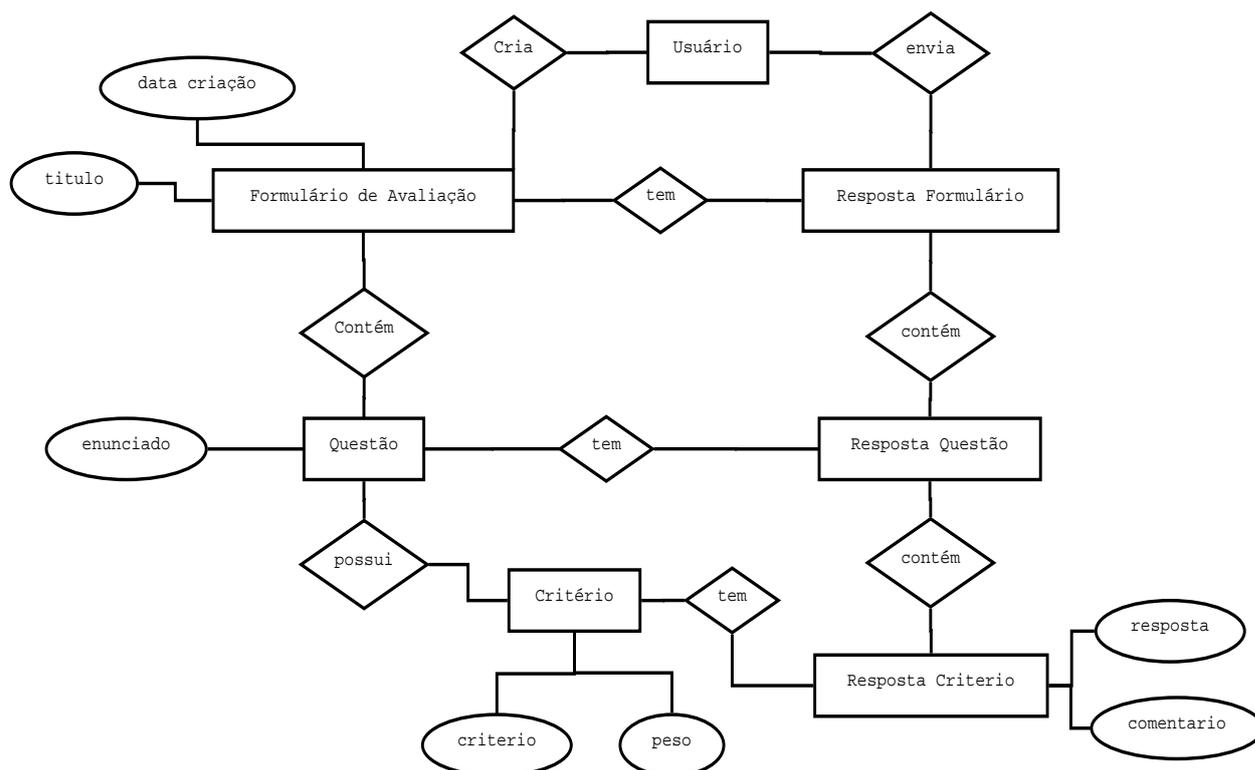


Figura 29 – Diagrama de entidade relacionamento para um formulário do sistema

Fonte: elaborado pelo autor

uma representa uma pergunta de avaliação para um usuário. Cada *Questão* possui o enunciado e um conjunto de *Critérios* que serão avaliados. Cada *Critério*, por sua vez, possui o texto do critério e seu peso na nota final.

Por outro lado, um *Usuário* (usuário com nível de aluno) poderá enviar as respostas desses formulários para uma determinada resolução do usuário (que será detalhada mais a frente, na figura 30). A *Resposta do Formulário* basicamente contém o *Formulário* envolvido e um conjunto de *Respostas das Questões*. Uma *Resposta de Questão*, por sua vez, contém a *Questão* envolvida e um conjunto de *Respostas dos Critérios*. Finalmente, uma *Resposta Critério* possui o *Critério* envolvido e as respostas do usuário e os comentários realizados.

Por fim, um usuário ao responder a questão enviará suas respostas ao sistema e este deverá armazená-las. Neste sentido o diagrama apresentado na Figura 30 mostra a modelagem da resposta do usuário bem como do relatório final a ser apresentado a ele quando sua resposta for avaliada.

Podemos observar na figura que a *Resolução Avaliação Escrita* feita pelo estudante possui um texto (sua resposta) e um status, que indicará se a correção já foi ou não concluída. Além disso, possui o *Problema Discursivo* envolvido (detalhado na figura 28) e, caso o status seja completo, a resolução possuirá um *Relatório* com o resultado final do aluno e um conjunto de *Resultado Critério*, indicando o resultado do aluno de forma

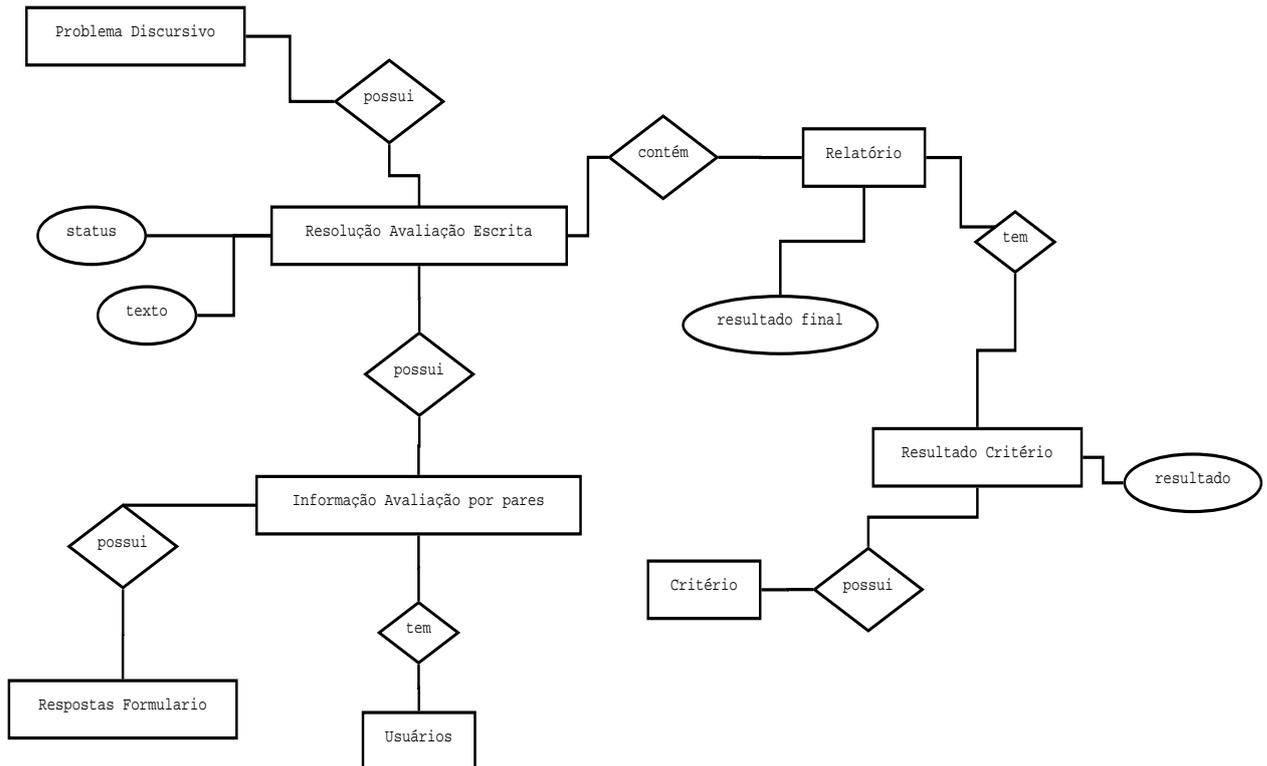


Figura 30 – Diagrama de entidade relacionamento para a resposta do usuário

Fonte: elaborado pelo autor

detalhada com as notas para cada critério envolvido na avaliação.

Além disso, a *Resolução Avaliação Escrita* também possui um conjunto de *Informação Avaliação por pares*. Essas informações são gravadas pelo sistema para realizar o gerenciamento das correções das avaliações. Entre as informações destacadas, destaca-se o conjunto de *Respostas Formulario*, que são as respostas dos formulários enviados por outros estudante referentes a *Resolução Avaliação Escrita* atual e o conjunto de *Usuário* que estão como responsáveis pela correção da respectiva resolução.

Ainda referente a modelagem, precisamos que algumas informações do usuário sejam guardadas a fim de que o modelo possa utilizá-las para efeitos de processamento. A Figura 31 apresenta a modelagem dessas informações para um determinado usuário.

Podemos notar na figura que um determinado usuário possui um relacionamento com a classe *Informação Objetivo*, que contém alguns de seus dados. As *Informações de Objetivos* do usuário corresponde ao número de atividades que este realizou, o número de atividades que foram corrigidas por ele, sua confiabilidade e sua reputação geral que servirá como parte do cálculo da nota final. Essas *Informações de Objetivos* também contém um conjunto de *Informações de contexto*, que são informações do usuário a respeito dos contextos envolvidos para avaliação. Cada *Informação de contexto* possui o contexto envolvido, a reputação do usuário naquele contexto e a data de atualização dos dados para ele.

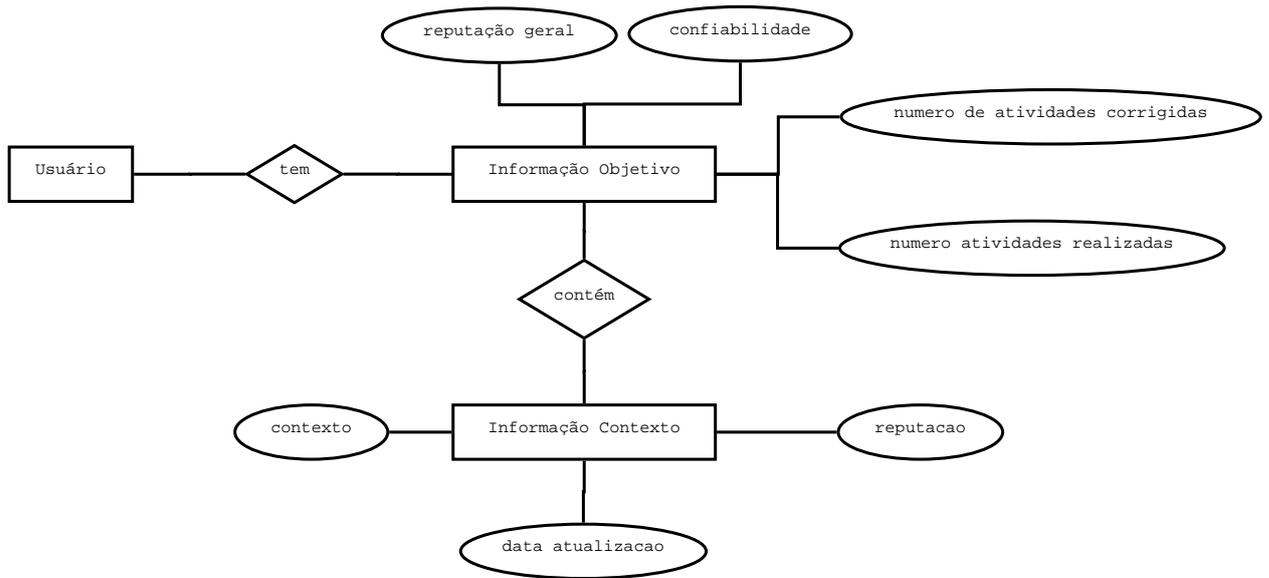


Figura 31 – Diagrama de entidade relacionamento para as informações do usuário

Fonte: elaborado pelo autor

Por último é necessário a modelagem referente a gamificação do modelo. A modelagem de gamificação é dividida em duas partes. A primeira é referente a modelagem dos elementos específicos de gamificação, isto é, a modelagem do framework específico do 6D (apresentado na Seção 2.2.3). A segunda parte é referente a modelagem do usuário interligado com a gamificação, isto é, os dados de gamificação do usuário. A Figura 32 apresenta a modelagem referente a primeira parte, apresentando os Loops de Atividade e suas especializações.

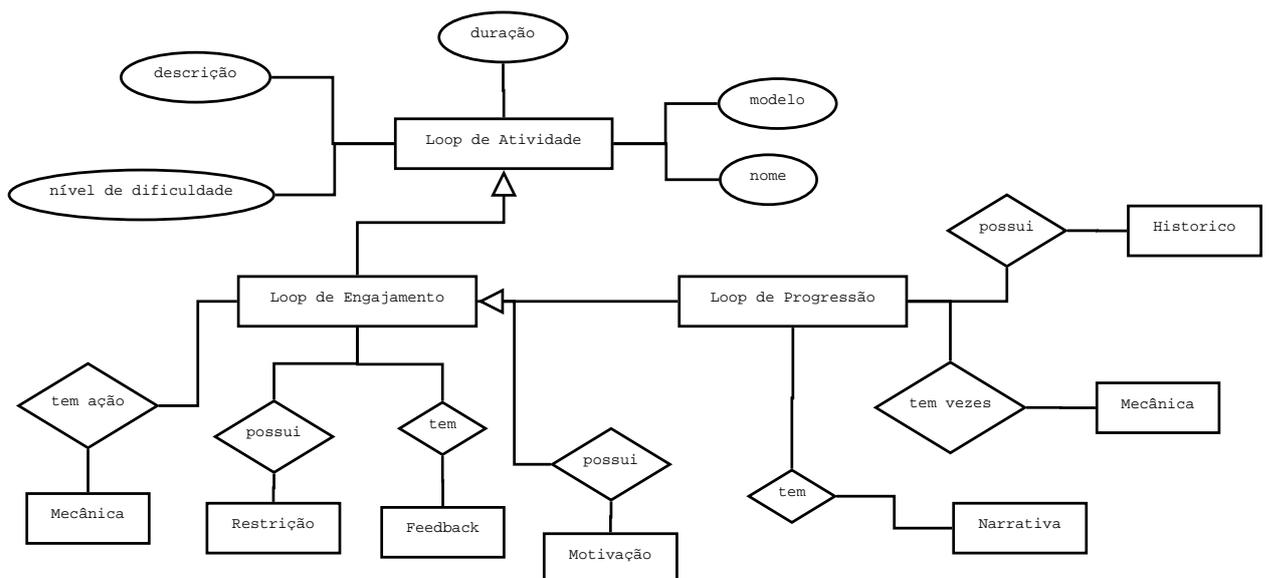


Figura 32 – Diagrama de entidade relacionamento para gamificação

Fonte: elaborado pelo autor

Podemos observar na figura que um *Loop de Atividade* é composto por um nome, uma descrição um nível de dificuldade, uma duração e um modelo, respectivamente. Um *Loop de Engajamento*, por sua vez, é uma especialização de um *Loop de Atividade*, herdando seus atributos, e além disso, possui uma ou mais ações *Mecânica*. Uma *Mecânica*, segundo o framework 6D, pode ser desafio, competição, cooperação, recompensa ou comentários.

Além disso, o *Loop de Engajamento* possui uma ou mais restrições (limitações daquele loop), como por exemplo, “apenas uma vez”. Também possui um conjunto de *Feedback*, que serão as recompensas que o usuário irá ganhar caso consiga realizar aquele loop. Um feedback pode ser um avatar novo, troféu, coleção, nível, pontos, desbloqueio de recursos, aumento ranking, entre outros. E finalmente, o loop possui uma *Motivação* para que este seja criado e realizado.

Um *Loop de Progressão*, por sua vez, herda dos *Loops de Engajamento*, apresentando todas as suas propriedades, porém, como neste tipo de loop, há uma ideia temporal, ele acrescenta algumas propriedades como a *Narrativa*, a quantidade de vezes que o usuário irá realizar as mecânicas e um *Histórico* de ações do usuário para que este consigo completar o loop.

E finalmente, a última modelagem refere-se a interligação do usuário e a gamificação do sistema. A figura 33 apresenta as informações de gamificação de um usuário pertencente ao modelo. Essas informações vão sendo atualizadas a medida que o usuário vai realizando os loops.

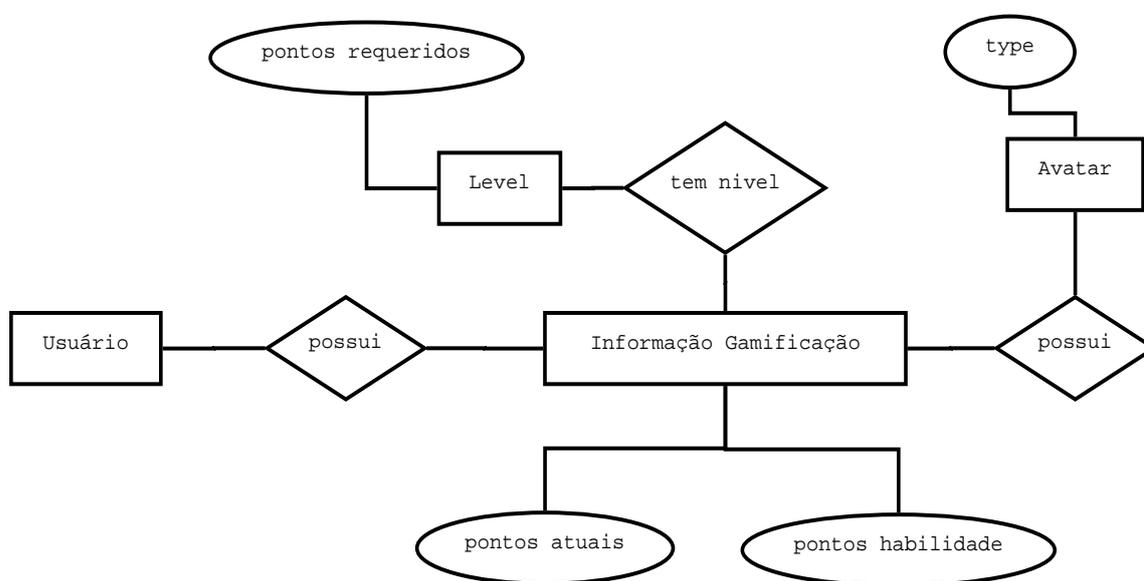


Figura 33 – Diagrama de entidade relacionamento para informações de gamificação do usuário

Fonte: elaborado pelo autor

Na figura, pode-se notar que o *Usuário* possui *Informações de Gamificação*. Estas, por sua vez, contêm o nível atual do usuário (*Level*), a quantidade de pontos de XP, a quantidade de pontos de habilidade e o *Avatar* do usuário. A medida que o usuário vai

ganhando pontos através dos feedbacks dos loops, seus pontos vão aumentando, a medida de que seu *Level* e seu *Avatar* acompanham com evoluções.

Toda a modelagem apresentada acima foi implementada utilizando ontologias OWL, tendo como base algumas das ontologias disponibilizadas pelo ambiente educacional MeuTutor. Neste sentido algumas das modelagens foram apenas modificações das ontologias do MeuTutor, sendo necessário a criação de apenas uma nova ontologia, para representar o conceito de avaliação por pares, que antes não existia no ambiente. A criação de ontologias e a utilização de OWL se deu exclusivamente porque estas tecnologias são utilizadas no MeuTutor.

Neste sentido, para a modelagem do usuário foram realizadas adaptações na ontologia *Learner.owl*, como a inclusão das informações de gamificação, as informações de resolução de atividades discursivas e o relatório da avaliação. Na ontologia de recursos (*Resource.owl*), foi necessário incluir a modelagem de questões discursivas. A ontologia de gamificação (*Gamificação.owl*) foi também adaptada para suportar os loops e as demais informações de gamificação. Por questões de limitação e espaço deste trabalho, alguns códigos e ilustrações destas ontologias serão apresentadas no Apêndice B.

Além das mudanças nas ontologias acima, foi necessário a criação de uma ontologia para suportar o modelo de avaliação por pares. Para isso foi criada a ontologia *PeerAssesment.owl*. Esta ontologia contém todas as informações referentes a criação de formulários de avaliação, bem como informações referentes as suas respostas pelos usuários, além de informações de controle do gerenciamento de distribuição das correções, atividade que é realizada pelo ambiente. A Figura 34 apresenta as classes da ontologia citada.

Na figura, podemos observar as classes que implementam a modelagem apresentada na Figura 29. Isto refere-se a criação dos formulários de avaliação bem como as suas respostas. Além disso, nessa ontologia, também estão incluídas algumas das entidades da Figura 30, como as informações de avaliação por pares.

Vale ressaltar que após todas as ontologias serão criadas com as modelagens dos diagramas de entidade relacionamento apresentados ao longo da dissertação, estas foram submetidas a um processo de compilação, onde, a partir das classes, atributos e relacionamentos descritos nas ontologias, foram geradas classes JAVA que os representam, tendo como resultado final um arquivo *.jar a ser incluído no projeto de implementação da proposta, que será apresentado a seguir.

Após a implementação da camada de persistência do modelo proposto, seguindo a modelagem descrita acima, o próximo passo foi sua implementação. A implementação, conforme já mencionado acima, utilizou a linguagem de programação JAVA e uma estrutura baseada em componentes, utilizando o padrão COSMOS. A implementação também utilizou o padrão arquitetural em camadas, respeitando as camadas de persistência, negócio (mgt), sistema e serviços (ws).

Basicamente, foi necessário a criação de três componentes principais: um componente

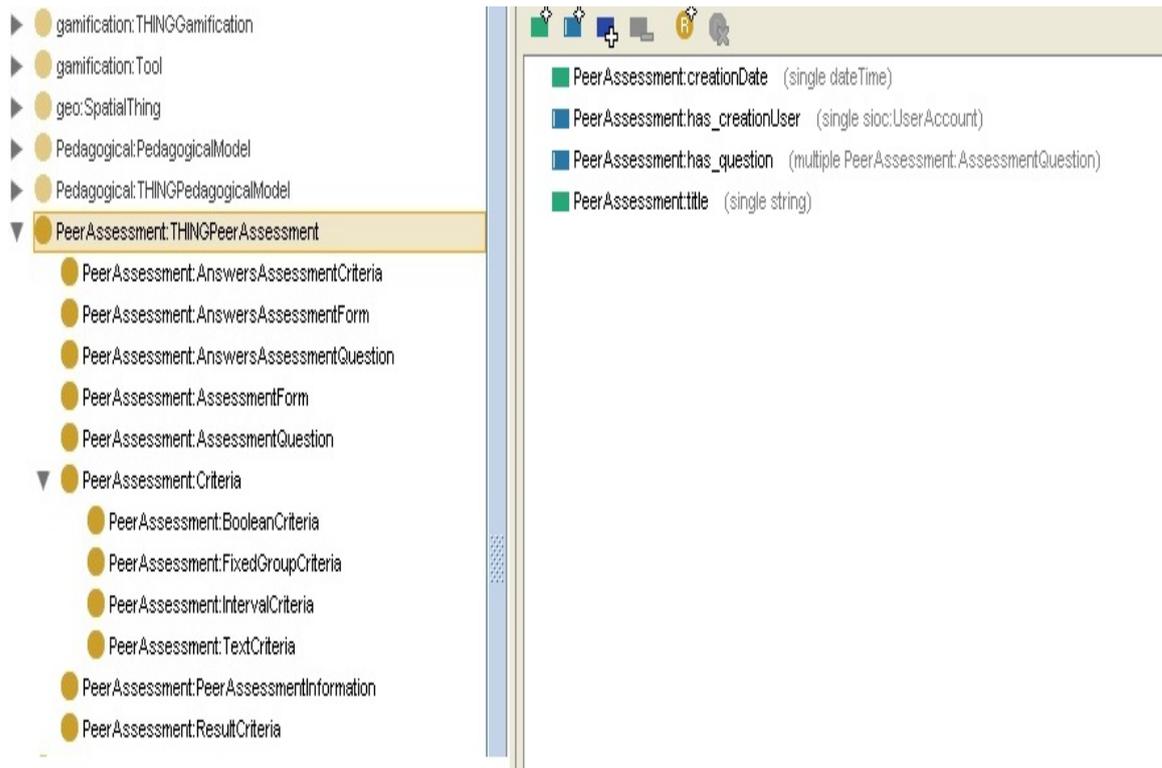


Figura 34 – Classes da Ontologia Peer Assessment.owl

Fonte: elaborado pelo autor

responsável pela avaliação por pares e o processo de gerenciamento desta (Peer Assessment); um componente responsável pelos problemas discursivos (problemas de avaliações escritas) (Writing Evaluation); um componente responsável pela gamificação (Gamification). Além disso, dois outros componentes foram criados: um para gerenciar a segurança (Security) e outro para converter os objetos vindo do modelo para objetos a serem transferidos via serviços web (ConvertDTO).

Além disso, com o objetivo de gerenciar a comunicação com o banco de dados foram necessários a criação de três componentes de negócio, um para cada componente principal do sistema descrito acima. Com isso foram criados mais três componentes: Peer Assessment Mgt, Writing Evaluation Mgt e Gamification Mgt. Todos eles fazendo a comunicação entre o banco de dados em RDF (Banco que armazena as ontologias) e os componentes do sistema.

A figura 35 mostra um diagrama de componentes com os respectivos componentes, suas interligações e suas interfaces requeridas e providas.

Podemos observar na figura a clara divisão entre as camadas implementadas no sistema. Na camada de negócio, os três componentes são responsáveis por manter a conexão com o banco de dados e assim oferecer consultas, criação, remoção e atualização de informações. Cada componente oferece uma interface própria, sendo o componente *Peer Assessment Mgt* oferecendo a interface *IPeerAssessmentMgt*, a qual possui métodos para

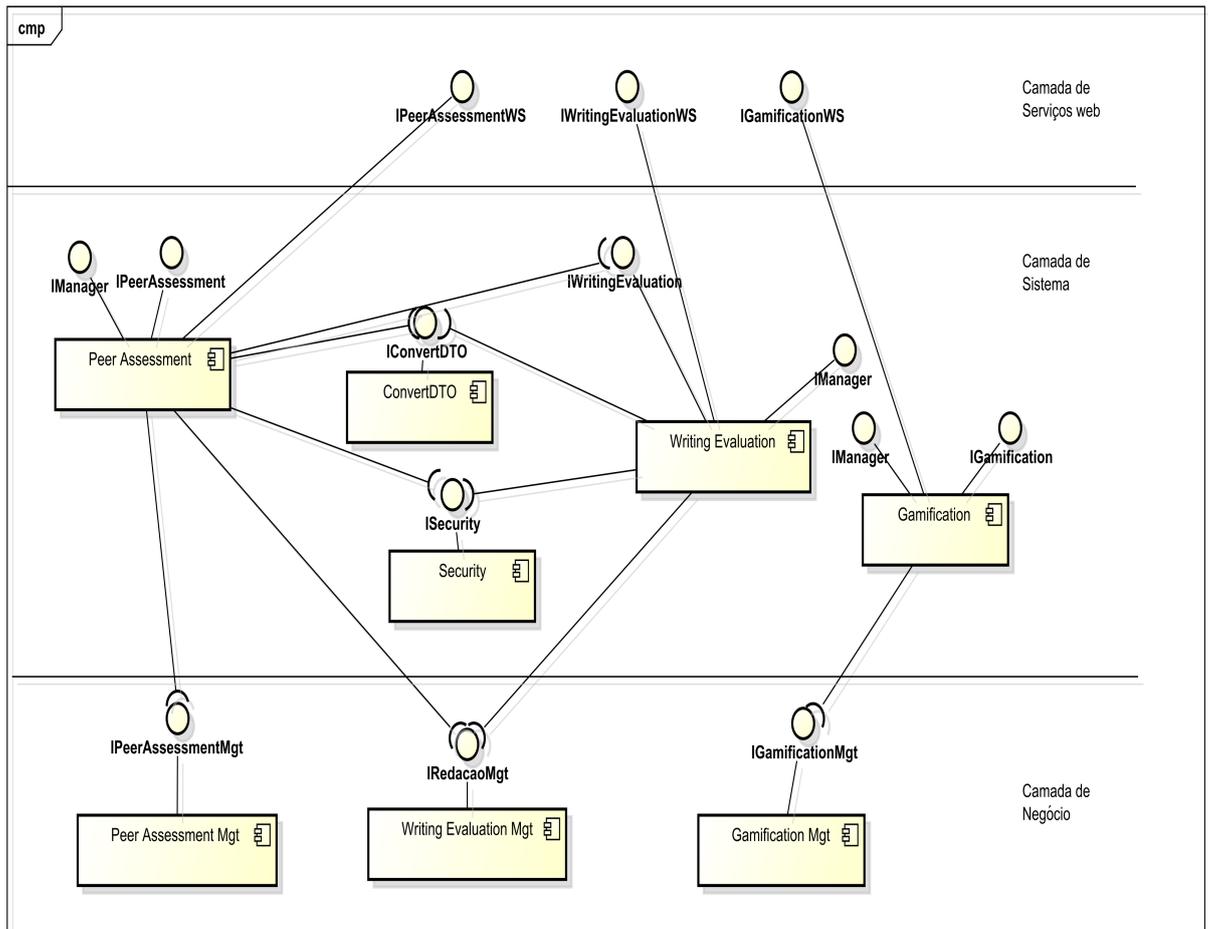


Figura 35 – Diagrama de componentes da implementação do modelo proposto

Fonte: elaborado pelo autor

criação de formulários, questões, critérios, bem como métodos para criar as respostas destes formulários e os relatórios dos resultados da avaliação.

Por outro lado, o componente *Writing Evaluation Mgt* oferece a interface `IWritingEvaluationMgt`, responsável pela criação de contextos, problemas discursivos, recuperação de problemas, criação das respostas do problema discursivo, entre outras responsabilidades. Finalmente, o componente *Gamificacao Mgt* oferece os métodos para criação e recuperação de loops, criação e recuperação das informações de gamificação dos usuários e criação de seus históricos de gamificação.

A segunda camada, camada de sistema, é responsável pela implementação da lógica do modelo. Os componentes são responsáveis por fornecer funcionalidades reais aos usuários finais bem como a outros componentes do sistema. Nesta camada temos os seguintes componentes e suas descrições:

- **Peer Assessment:** Componente responsável por toda a implementação vinculada a formulários de avaliação, incluindo tanto sua criação, recuperação, alteração bem

como tratamento das respostas dos usuários a estes formulários. Possui a interface provida *IPeerAssessment* fornecendo estas funcionalidades e requer a interface *IConvertDTO*, *ISecurity* e *IWritingEvaluation* dos outros componentes.

- **Writing Evaluation:** Este componente oferece através da interface *IWritingEvaluation* serviços para criação, recuperação e alteração de problemas discursivos, vinculação de problemas com formulários, recuperar resultados dos problemas, bem como todo o tratamento de responder e recuperar respostas dos problemas discursivos de um usuário. Requer as interfaces *IConvertDTO* e *ISecurity*.
- **Gamification:** Componente que implementa a gamificação do modelo. Responsável por todas as funcionalidades de criação de loops, bem como o tratamento deles para com o usuário, envolvendo o controle das ações realizadas pelo usuário, bem como o tratamento dos feedbacks e recompensas para o usuário. Todos os serviços são fornecidos através da interface *IGamification*.
- **Security:** Esse componente tem como responsabilidade garantir a segurança dos dados envolvidos bem como assegurar que as chamadas de serviços estão sendo feitas de maneira correta e por usuários reais.
- **ConvertDTO:** Esse componente surgiu da necessidade de enviar os dados vindos do banco de dados através dos serviços web de forma serializada. Este componente tem como objetivo converter os dados para modelos em DTO (Objeto de transferência de dados).

É importante ressaltar que todos os componentes possuem a interface provida *IManager*. Isso ocorre devido ao padrão COSMOS, e esta interface tem como função controlar o acesso ao componente, bem como o controle das interfaces providas/requeridas por ele. Além disso, cada componente principal também tem como interface requerida uma interface de acesso ao banco (pois a camada de sistema não pode acessar diretamente o banco). Esta interface requerida pelos componentes da camada de sistema é a mesma interface provida pelos componentes da camada de negócio.

Por fim, além desses componentes de negócio e sistema e suas interfaces já descritas, todas as funcionalidades do modelo proposto são disponibilizadas via serviços web na terceira camada. Neste sentido, é necessário incluir uma interface de serviço web em cada componente principal do sistema, fornecendo serviços que podem ser utilizados pelos ambientes educacionais. Com isso, três interfaces de serviços foram criadas e disponibilizadas.

Cada interface tem a responsabilidade de prover os serviços que seus componentes implementam através da WEB. Neste sentido, foi incrementada a nova interface (com sufixo WS, para diferenciá-la da original) que oferece os serviços necessários. Por questões de

espaço, visto que os códigos das interfaces ocupam mais de mil linhas, não é possível apresentar o código de todas as interfaces WS implementadas no sistema ao longo desse texto. Contudo, no Apêndice C é apresentado o código Java da interface *IWritingEvaluationWS*.

Vale ressaltar que sua implementação é baseada em serviços web descritos em WSDL, e utilizam o protocolo SOAP como forma de comunicação entre o cliente (ambiente que fará as requisições) e o servidor (modelo implementado). A implementação do modelo foi então compilada, virando um arquivo *.war e implantada em um servidor glassfish e está disponível online, juntamente com o MeuTutor ⁷.

Tanto o modelo apresentado na seção 4.1, bem como sua implementação apresentada na Seção 4.2 foram pensados e planejados para serem independentes do ambiente educacional. Nesse caso, para um ambiente educacional utilizar a implementação apresentada, basta realizar chamadas de serviços via SOAP ao modelo implantado. A descrição dos serviços (WSDL) estão online e podem ser acessadas.

4.2.3 Integração do modelo ao ambiente MeuTutor

Nas subseções anteriores foram apresentados o ambiente educacional MeuTutor bem como a implementação do modelo proposto nesse trabalho. O ambiente educacional MeuTutor se apresentou como adequado a implementação do modelo por possuir características favoráveis, conforme já citadas anteriormente. Como explanado, o MeuTutor tem como objetivo ensinar os conteúdos das disciplinas do ENEM, contudo um dos conteúdos cobrados neste exame é uma prova de redação.

Inicialmente o MeuTutor não tinha suporte a redações. Por esse motivo, os idealizadores do ambiente tinham como planejamento incluir um módulo de redação dentro do ambiente para que seus alunos pudessem praticar redações dentro do próprio sistema. Nesse caso, como solução promissora, foi realizado um planejamento para oferecer o suporte a redações utilizando o modelo proposto nesse trabalho.

Como nosso objetivo era implementar o modelo e implantar no ambiente educacional MeuTutor para suportar redações, foi preciso então implementar mais um passo na nossa arquitetura. Ou seja, precisamos implementar/incrementar as telas em PHP do MeuTutor (ver arquitetura do MeuTutor na figura 27), de tal maneira que suporte o modelo (fazendo chamadas aos serviços web do modelo quando necessário).

Basicamente houveram dois grupos de telas que precisaram ser criadas/adaptadas: as telas dos professores e as telas dos alunos. No lado do professor, as principais funcionalidades são a criação de temas de redação (tela ilustrada na Figura 36), a criação da redação propriamente dita (Figura 37) e a criação do formulário de avaliação.

Na figura 36, podemos visualizar três temas de redações criados (Natureza, Política e Atualidades). Nessa tela, é possível a criação de novos temas de redações (utilizando o

⁷ A lista navegável com os serviços disponibilizados está disponível através do endereço <http://redacao.meututor.com.br:8080/redacao/ws>



Figura 36 – Tela de listagem de temas de redação do MeuTutor

Fonte: elaborado pelo autor

botão azul “Cadastrar tema de redação”), bem como a listagem de redações por tema e a edição de um tema previamente cadastrado.

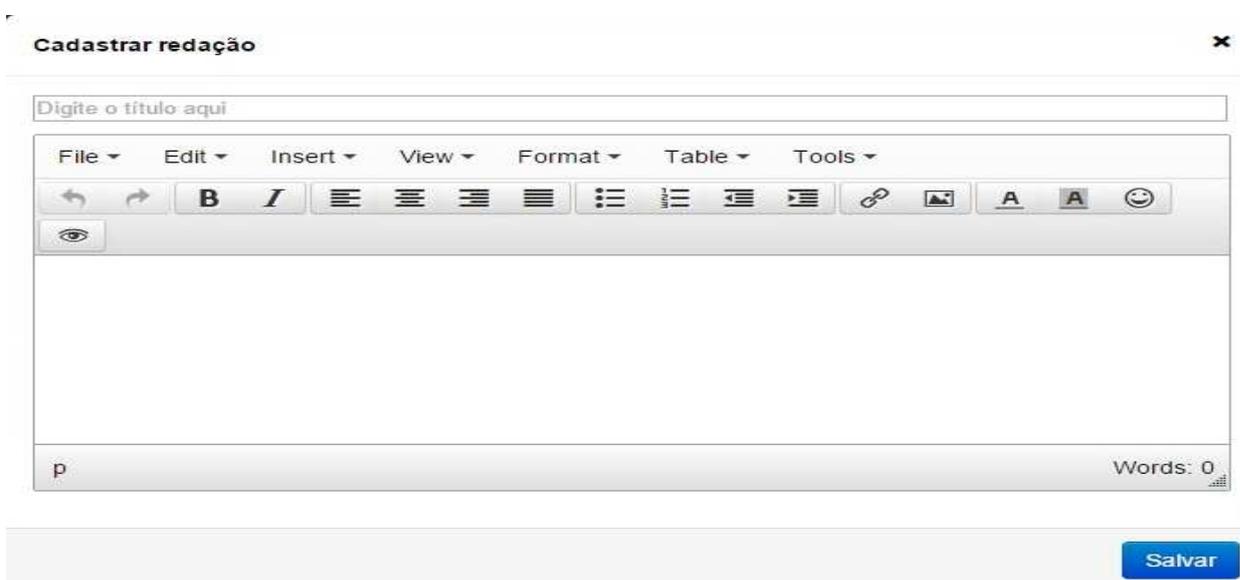


Figura 37 – Tela de cadastro de redações do MeuTutor

Fonte: elaborado pelo autor

Na tela da figura 37, o professor poderá cadastrar uma redação digitando apenas seu título e seu texto, uma vez que já selecionou previamente o tema no qual a redação está envolvida. É importante ressaltar que um texto de uma redação pode incluir formatações (negritos, itálico, sublinhado, alinhamento), assim como pode conter imagens e emoticons.

Do ponto de vista do aluno, este ao entrar no sistema e clicar no menu de redação na barra superior, irá visualizar se há alguma redação disponível no sistema para ser respondida por ele, conforme ilustrado na Figura 38.



Figura 38 – Tela inicial de redação para um aluno no MeuTutor

Fonte: elaborado pelo autor

Na Figura 38, o aluno verá que está na primeira etapa do processo de avaliação (a etapa de submissão), onde ele deverá realizar (responder) a atividade proposta. Logo abaixo, é possível visualizar o título da redação que estará disponível. Com isso, o aluno poderá iniciar a redação clicando no botão “Iniciar Redação”.

Ao clicar em iniciar redação o aluno visualizará toda a proposta da redação que está sendo aplicada e poderá escrever sua redação diretamente na tela. Além disso, serão apresentadas algumas instruções que o aluno deverá respeitar ao criar sua redação, conforme pode ser visto na Figura 39.

Entre as restrições estão que o aluno deverá responder no mínimo seis linhas e a impossibilidade de copiar textos presentes na proposta. Além disso, receberá nota zero todo aluno que fugir do tema ou não escrever texto argumentativo-dissertativo, contrariar os direitos humanos e/ou apresentar texto deliberadamente desconectado do tema apresentado.

Após enviar a sua resposta para a redação ao sistema, o aluno deverá aguardar o prazo para o início da segunda etapa - a correção. Este prazo é definido pelo professor ao cadastrar uma determinada redação. Enquanto o prazo não se inicia, aparecerá para o aluno um cronômetro indicando quanto tempo resta para o início das correções. Com isso, o aluno poderá planejar suas avaliações.

PROPOSTA DE REDAÇÃO

Título**proposta...**

Instruções:

- Mínimo de 6 linhas
- Na folha de redação abaixo o texto deve ser escrito com o mínimo de 7 e o máximo de 30 linhas.
- A redação que apresentar cópia dos textos da Proposta de Redação terá o número de linhas copiadas desconsiderado para efeito de correção.

Receberá nota zero, em qualquer das situações expressas a seguir, a redação que:

- fugir ao tema ou que não atender ao tipo dissertativo-argumentativo.
 - apresentar proposta de intervenção que desrespeite os direitos humanos.
 - apresentar parte do texto deliberadamente desconectada com o tema proposto.
-

1	
2	
3	
4	
5	
6	
7	
8	

Figura 39 – Tela proposta de uma redação para um aluno no MeuTutor

Fonte: elaborado pelo autor

Quando o prazo de submissão acaba, a segunda etapa de correção das redações se inicia. Nessa etapa, o aluno verá que existem outras redações para que ele corrija (geralmente duas ou três outras redações). Nesse caso, para cada redação que o aluno deverá avaliar, será apresentado a redação junto com o formulário de avaliação criado pelo professor, conforme ilustrado na Figura 40.

O aluno deverá então ler calmamente e de forma crítica a redação de seu par, julgando-a sob os critérios pré-definidos pelo professor e, em seguida, deverá preencher as respostas para o formulário definido pelo professor de acordo com o seu julgamento do trabalho apresentado, clicando, ao final, no botão “Enviar avaliação”.

1	Quando o Brasil foi selecionado para sediar o Mundial de
2	Futebol, ouviram-se vários discursos do governo
3	ressaltando a importância desse evento para o
4	desenvolvimento do país. A preparação para a realização dos
5	jogos mobilizou vários setores, tais como: o econômico e o
6	social. Entretanto a herança deixada para o país pela Copa
7	não será tão positiva quanto o imaginado inicialmente.
8	Os planejamentos feitos para sediar a Copa do Mundo
9	foram insuficientes e o desenvolvimento desses
10	projetos levanta suspeitas. Atrasos na entrega de
11	estádios-sede dos jogos e de obras que visam a reformular
12	a estrutura física das cidades-sede ainda em andamento, tais

Critérios de eliminação

<input type="checkbox"/>	Impropérios, desenhos e outras formas propositais de anulação ou parte do texto deliberadamente desconectada do tema proposto	<input type="checkbox"/>	Fuga total ao tema
<input type="checkbox"/>	Desrespeito aos direitos humanos	<input type="checkbox"/>	Não obediência à estrutura dissertativo-argumentativa

Notas por área

Competências	Notas	Descrição
Demonstrar domínio da modalidade escrita formal da Língua Portuguesa	 160	Bom.
Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa	 120	Tema adequadamente desenvolvido, abordagem superficial, argumentação previsível.
Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista	 80	Aspectos do tema pouco articulados ou contraditórios.

Figura 40 – Tela onde um aluno corrige a redação de outro

Fonte: elaborado pelo autor

Em seguida, o sistema apresentará um relatório breve contendo as notas que foram atribuídas pelo aluno, e, após conferir as suas respostas, este deverá confirmá-las e enviar ao sistema. Este processo deverá ser repetido pelo aluno até que ele tenha concluído a correção das redações que foram atribuídas a ele. No final, não haverá mais nada a ser feito na segunda etapa de correção.

Finalmente, após o prazo da segunda etapa ser finalizado, todas as redações estarão com as correções concluídas e será iniciada a terceira etapa - que é a apresentação dos resultados finais. Esta etapa consiste na apresentação dos resultados obtidos pelo processo de avaliação ao longo das etapas de submissão e correção.

O aluno então poderá visualizar o seu relatório final, com as notas de cada critério avaliado bem como o resultado final, conforme pode ser visto na Figura 41.

Podemos observar na Figura 41 o resultado final do aluno e ainda a redação no qual o relatório se refere. Observe que a nota final da redação do aluno foi 600.00 (seiscentos

Resultado redação



Nota total
600.00

Qual será o legado da Copa do Mundo para o Brasil?

Suas notas em cada diretriz

Competências	Notas	Comentários
Demonstrar domínio da modalidade escrita formal da Língua Portuguesa	 120.00	Adequado.
Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa	 120.00	Tema adequadamente desenvolvido, abordagem superficial, argumentação previsível.
Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista	 120.00	Propõe aspectos do tema de forma pouco consistente, pouca articulação entre os argumentos.
Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação	 120.00	Algumas inadequações, eventuais desvios, domínio regular da coesão.
Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos	 120.00	Proposta de intervenção relacionada ao tema, mas pouco articulada.

Figura 41 – Tela relatório final de um aluno

Fonte: elaborado pelo autor

pontos). Essa nota é a soma da nota atribuída em cada critério. Como a figura é apenas ilustrativa, as notas de cada critério foram 120 (cento e vinte).

Cada critério de avaliação é representado por uma linha na tabela com três colunas, sendo a primeira coluna o texto explicando qual é o critério que está sendo avaliado, a segunda coluna com sua nota final e a terceira com uma explicação/comentários com relação à nota atribuída.

5 VALIDAÇÃO - EXPERIMENTO PROCESSO AVALIAÇÃO PARES

Assim como toda pesquisa científica, é necessária uma maneira de validar o que está sendo proposto, a fim de apresentar seus resultados positivos bem como possíveis resultados negativos que vierem a surgir. Nesse sentido existem diversas técnicas para validação de trabalhos científicos. Nesse trabalho escolhemos a técnica de experimento.

A proposta desse trabalho, em resumo, é um modelo de avaliação por pares gamificado que poderá ser utilizado em ambientes educacionais online. No sentido de avaliar sua eficácia, foram projetados e realizados dois experimentos. O primeiro avalia a efetividade do processo de avaliação por pares dentro do modelo, o qual será descrito e apresentado neste capítulo. O segundo experimento avalia o aspecto motivacional dos estudantes a participarem do processo, analisando os benefícios da gamificação aplicada ao modelo proposto. Este experimento será apresentado no próximo capítulo (capítulo 6).

Por definição, um experimento é um processo ou estudo que resulta na coleta de dados. Os resultados e experiências não são conhecidos com antecedência. Geralmente, os experimentos estatísticos são realizados em situações em que os pesquisadores podem manipular as condições do experimento e pode controlar os fatores que são irrelevantes para os objetivos da pesquisa (YANG; EL-HAIK, 2003). O experimento foi conduzido seguindo alguns princípios da engenharia de software experimental (WOHLIN et al., 2012).

O projeto do experimento é o processo de planejamento de um experimento para atender determinados objetivos específicos. Planejar um experimento corretamente é muito importante, a fim de garantir que o tipo correto de dados e o tamanho e poder da amostra sejam suficientes e estejam disponíveis para responder às questões de pesquisa de interesse da forma mais clara e eficiente possível (YANG; EL-HAIK, 2003).

Neste sentido, em primeiro momento, foi realizado um projeto de experimento com a finalidade de validar o processo de avaliação por pares apresentado no modelo proposto. O projeto de experimento está apresentado na Seção 5.1. Após a conclusão do projeto do experimento, este seguiu para o processo de execução, conforme descrito na Seção 5.2. A seção 5.3 mostra como os dados foram coletados e o processo de análise deles. Por fim, a Seção 5.4 apresenta os principais resultados alcançados.

5.1 Projeto de experimento

Esta seção tem como objetivo apresentar o projeto de experimento desenvolvido para validar o modelo proposto. A proposta do experimento está envolvida no contexto do projeto MeuTutor ENEM, ambiente apresentado na Seção 4.2.1.

O MeuTutor ENEM, conforme já foi explanado, é ambiente educacional online para ensino a distância, com cerca de 10 mil estudantes, cujo principal objetivo é fazer um

acompanhamento da aprendizagem dos alunos de forma individualizada, focando na qualidade do ensino e no desempenho dos alunos utilizando uma gestão inteligente da aprendizagem.

O primeiro passo em uma proposta de experimento é delinear bem o problema que está sendo trabalhado, apresentando uma contextualização, bem como os problemas de negócio e problemas técnicos envolvidos. Essas informações de definição do problema serão apresentadas na Seção 5.1.1. Em seguida, os objetivos do experimento foram levantados e apresentados na Seção 5.1.2. Finalmente, a Seção 5.1.3 apresenta o planejamento do experimento, contendo as questões de pesquisa e hipóteses, fatores e variáveis de resposta, níveis dos fatores, definição formal das hipóteses, unidades experimentais, design de experimento, um plano de execução, instrumentação e finalmente, uma análise de ameaças à validade do experimento.

5.1.1 Definição do Problema

Essa seção tem como objetivo apresentar uma visão geral do problema que está sendo tratado nesse experimento. Conforme explanado, o experimento está no contexto do MeuTutor ENEM. Este tem como objetivo principal auxiliar os alunos a se prepararem para a prova do ENEM. A prova do ENEM, por sua vez, é dividida em 5 grandes áreas do conhecimento. As áreas e suas relações com as disciplinas do ensino médio são:

- **Ciências da Natureza e suas Tecnologias:** Física, Química e Biologia
- **Ciências Humanas e suas Tecnologias:** História e Geografia
- **Matemática e suas Tecnologias:** Matemática
- **Linguagens, Códigos e suas tecnologias:** Português, Literatura, Inglês e/ou Espanhol
- **Redação**

A prova contém 180 questões (45 para cada área) e uma redação e é realizada ao longo de dois dias (Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), 2013). O ENEM é usado como porta de entrada para as universidades federais e algumas particulares. A redação, nesse contexto, possui uma importância alta, pois corresponde a 20% da nota final. No início, o ambiente MeuTutor ENEM oferecia o suporte completo as 4 primeiras áreas citadas, disponibilizando vídeo aulas, questões e simulados aos alunos, onde as correções eram feitas de forma automática por se tratar de problemas de múltipla escolha. Contudo, houve a necessidade de incluir a disciplina de redação no ambiente devido à sua importância.

Para se ter uma ideia da importância de provas de redação e da inviabilidade de sua correção em ambientes online da forma tradicional, no Novo ENEM (Exame Nacional

do Ensino Médio), aplicado anualmente no Brasil como forma de seleção unificada nos processos seletivos das universidades públicas federais, onde uma de suas etapas é uma prova de redação, o número de inscrições superou a marca de 7 milhões (mais precisamente, 7.173.574, segundo o Instituto Nacional de Estudos e Pesquisas Educacionais - INEP(2013).

Diante do apresentado, como problema de negócio do ambiente MeuTutor tínhamos que **A inclusão de questões discursivas (redações) como forma de avaliação no MeuTutor ENEM era uma tarefa que apresentava um alto custo e uma sobrecarga no professor, se tornando cada vez mais inviável a correção da produção textual de forma personalizada devido à quantidade cada vez maior do número de estudantes presentes no ambiente.**

Dessa maneira, desejava-se aplicar o modelo de avaliação por pares juntamente com técnicas de gamificação proposto na tentativa de suprir os problemas. Com isso, incluímos o modelo proposto nesse trabalho dentro do ambiente MeuTutor, conforme já explanado na seção 4.2.3, com o objetivo principal de suportar redações entre os alunos.

Essa inclusão do modelo nos levou a questionar a qualidade das correções com a seguinte pergunta: **“Como nós podemos avaliar a qualidade das correções feitas pelos alunos?”** e, principalmente, **“essas avaliações podem ser comparadas com uma avaliação de um especialista (professor)?”**. Além disso, também nos questionamos sobre **“Será que o modelo aplicado nesse contexto realmente soluciona os problemas citados?”** (Redução do custo e sobrecarga do professor, entre outros).

Em seguida aos questionamentos, também devemos começar a destacar algumas métricas que poderiam nos levar as respostas de tais perguntas, como por exemplo: Semelhança entre a nota obtida com a correção pelos alunos em comparação com a nota obtida pela correção junto aos professores; Custo vinculado a correção das redações; Tempo gasto pelo professor em todo o processo de aplicação do modelo proposto; Grau de dependência entre o custo e o número de estudantes no ambiente; entre outras.

Essas perguntas bem como suas métricas e o problema de negócio apresentado nos motiva a levantar os problemas técnicos envolvidos. Visando a resolver o problema de negócio definido anteriormente (completamente ou parcialmente), deseja-se **validar se o modelo de avaliação por pares proposto é ideal às necessidades e ao perfil dos usuários do MeuTutor ENEM, de tal forma que não haja uma dependência entre o número de usuários e o custo associado, garanta eficiência na aprendizagem e ainda permita uma personalização, sem necessariamente aumentar o trabalho do professor.**

O problema técnico apresentado nos levou às seguintes perguntas:

- O modelo de avaliação por pares proposto é adequado a ser aplicado no MeuTutor ENEM?

- Os resultados do uso de técnicas de avaliação por pares junto com gamificação são semelhantes aos resultados da correção pelos professores?
- O modelo auxilia o trabalho dos professores (diminuindo sua sobrecarga de trabalho) no MeuTutor ENEM?
- A dependência entre o custo e o número de alunos é diminuída com a aplicação do modelo?

As respostas para os problemas de negócio bem como os problemas técnicos e suas respectivas perguntas serão apresentadas ao longo desse capítulo, especialmente nas Seções 5.3 e 5.4, onde os dados obtidos com o experimento serão levantados, analisados e algumas conclusões apresentadas.

5.1.2 Objetivos da Investigação

A pesquisa a ser realizada é do tipo experimental, e tem como objetivo avaliar o comportamento do modelo de avaliação por pares proposto quando aplicado no MeuTutor ENEM, com muitos usuários. Os dados serão coletados a partir de um intervalo de tempo e analisados.

Os resultados das execuções serão analisados sob possíveis métricas de custo, redução da sobrecarga do professor, grau de dependência aluno x custo e a eficácia das correções com a implantação do modelo. Para avaliar a eficácia será realizada uma comparação entre os resultados obtidos pelo modelo junto aos alunos e os resultados obtidos por especialistas.

Formalmente, o objetivo da investigação pode ser definido no formato GQM como **analisar** o modelo de avaliação por pares proposto **com a intenção de** avaliá-lo a respeito de seu custo, benefícios aos professores/alunos, grau de dependência e precisão **do ponto de vista dos** usuários de ambientes educacionais **no contexto** do ambientes educacionais MeuTutor ENEM.

Como isso, temos como objetivos específicos os seguintes itens:

- Analisar o modelo proposto, seguindo os atributos do GQM, obtidos em sua execução quando aplicado no ambiente educacional MeuTutor ENEM.
- Verificar qual dos fatores do experimento é mais significativo para explicar a variação nos valores obtidos para as variáveis de resposta após a execução do modelo de avaliação por pares proposto.

5.1.3 Planejamento do experimento

Nesta seção, será detalhado o planejamento do experimento que foi projetado para este trabalho. Dentro do planejamento, encontra-se a definição da questão de pesquisa e

derivação de hipóteses, a seleção das variáveis dependentes e independentes, a identificação da unidade experimental e a seleção do modelo experimental que será utilizado, entre outros detalhes envolvidos, que serão detalhados nas subseções a seguir.

O experimento adotado é do tipo comparativo, no qual será comparada uma execução sem o modelo aplicado e o uso do modelo proposto. O contexto do experimento é industrial, posto que será executado em um ambiente educacional da indústria, executando o modelo com um conjunto de dados de usuários reais existentes nesse ambiente. Nas próximas subseções, o planejamento do experimento será descrito.

5.1.3.1 Questões de pesquisa e hipóteses

Como mencionado no início deste capítulo, este experimento visa avaliar o modelo proposto com relação ao custo, redução de sobrecarga do professor e eficácia. Levando isto em consideração, algumas questões de pesquisa foram definidas e serão apresentadas.

A eficácia tenta avaliar se o modelo proposto no trabalho se apresenta tão ou até mais eficiente que as correções de redações da forma tradicional. O método tradicional para incluir redações nos ambientes a distância (sem usar nenhum modelo de avaliação por pares, onde as mesmas são corrigidas por professores especialistas) também precisa ser avaliado. Neste sentido a primeira questão de pesquisa visa responder se usar o modelo de avaliação por pares proposto realmente é equivalente ao método tradicional dentro do ambiente MeuTutor ENEM. Podemos notar a equivalência se eles apresentarem resultados semelhantes. Dessa forma surge nossa primeira questão de pesquisa.

P1 - O uso do modelo de avaliação por pares proposto apresenta a mesma eficiência, isto é, apresenta resultados semelhantes se comparado ao método tradicional de correções?

O que nos leva às seguintes hipóteses:

H1-0: O uso do modelo de avaliação por pares proposto é equivalente ao método tradicional.

H1-1: O uso do modelo de avaliação por pares proposto não é equivalente ao método tradicional

A primeira pergunta tenta responder se a qualidade das notas dada pelos pares pode ser comparada com as notas de especialistas em redações (professores). Caso a hipótese nula seja refutada - indicando que há uma equivalência nos resultados - precisamos identificar em qual/quais características o uso do modelo por pares proposto traz diferenças significativas vantajosas com relação ao método tradicional.

Nesse sentido, ao aplicar o modelo no MeuTutor ENEM, é necessário avaliar o custo necessário para manter o modelo no ambiente e o tempo gasto pelos professores e alunos no processo. Essas métricas são amplamente utilizadas para avaliar alguns dos benefícios do modelo de avaliação por pares. Sendo assim, como questões de pesquisa secundárias do projeto, temos a determinação e comparação do modelo com respeito à essas métricas.

P2 - O uso do modelo de avaliação por pares proposto apresenta diferenças nas métricas de tempo em relação a não usar esse modelo?

O que nos leva às seguintes hipóteses:

H2-0: O uso do modelo de avaliação por pares proposto não traz diferenças de tempo em relação ao método tradicional

H2-1: O uso do modelo de avaliação por pares proposto traz diferenças significativas de tempo em relação ao método tradicional

P3 - O uso do modelo de avaliação por pares proposto apresenta diferenças nas métricas de custo em relação a não usar esse modelo?

O que nos leva às seguintes hipóteses:

H3-0: O uso do modelo de avaliação por pares proposto não traz diferenças de custo em relação ao método tradicional

H3-1: O uso do modelo de avaliação por pares proposto traz diferenças significativas de custo em relação ao método tradicional

Por fim, para que os resultados finais da pesquisa sejam melhor compreendidos, seria interessante conhecer as interações entre os fatores e as variáveis respostas. Dessa forma, desejamos saber qual dos fatores do experimento é mais responsável pela variação nas saídas obtidas nas execuções dos tratamentos. Assim, a última pergunta de pesquisa do experimento é definida do seguinte modo:

P4 - Qual o fator que mais explica a variação nas variáveis-resposta da pesquisa?

- Para responder essa pergunta será utilizada estimação de parâmetros, com alocação de variação para os fatores.

Formalmente, as hipóteses descritas anteriormente podem ser definidas conforme a Tabela 6. As funções N, T e C, apresentadas na tabela, retornam respectivamente, o valor da nota final, o tempo gasto e o custo envolvido, com relação a utilização do modelo tradicional M1 ou a utilização do modelo proposto M2, quando aplicado no ambiente educacional MeuTutor E1 ou quando aplicado sem ambiente educacional E2 (para efeitos de comparação da pergunta P4).

Tabela 6 – Definição formal das hipóteses de pesquisa.

Hipótese	Hipótese Nula	Hipótese Alternativa
H1	H1-0: $N(M1,E1) = N(M2,E1)$	H1-1: $N(M1,E1) \neq N(M2,E1)$
H2	H2-0: $T(M1,E1) = T(M2,E1)$	H2-1: $T(M1,E1) \neq T(M2,E1)$
H3	H3-0: $C(M1,E1) = C(M2,E1)$	H3-1: $C(M1,E1) \neq C(M2,E1)$

Fonte: elaborado pelo autor.

5.1.3.2 Fatores e variáveis de resposta

Após a definição de nossas hipóteses de pesquisa, o próximo passo do planejamento do experimento é a definição das variáveis contidas nele. Basicamente, temos dois tipos de variáveis: Primeiro, as variáveis independentes, também chamadas de fatores, são aquelas no qual queremos avaliar; Segundo, as variáveis dependentes, são o caso das métricas de avaliação.

Como variável independente temos:

- **Modelo de avaliação:** o modelo de correções que será utilizado em determinado momento.
- **Ambiente educacional:** o ambiente educacional onde o modelo será executado.

Os níveis dos fatores apresentados acima são definidos na Tabela 7.

Tabela 7 – Níveis dos fatores.

Fator	Nível	Descrição
Modelo	M1	Modelo de avaliação tradicional
Modelo	M2	Modelo de avaliação proposto
Ambiente	E1	Ambiente educacional MeuTutor-ENEM
Ambiente	E2	Sem ambiente educacional

Fonte: elaborado pelo autor.

As variáveis dependentes (métricas) são definidas abaixo:

- **Resultado final (N):** Esta variável representa a nota final obtida pelo modelo em uma determinada execução.
- **Tempo de execução (T):** Esta variável representa o tempo gasto pelo modelo para obter um resultado final
- **Custo de execução (C):** Esta variável representa o custo total envolvido pelo modelo para obter um resultado final

5.1.3.3 Unidades de Experimento

O objetivo do nosso experimento é avaliar o comportamento do modelo tradicional e do modelo proposto executados no contexto de redação. Em nosso projeto de experimento, após definir as questões, hipóteses e fatores e variáveis de resposta, precisamos definir qual será a unidade de experimento utilizada.

No nosso cenário, a unidade experimental é bastante específica. A unidade experimental da pesquisa é uma determinada redação, pré-definida pelo professor e um determinado

formulário de avaliação. O conjunto redação + formulário é uma amostra que será utilizada sob a qual o “tratamento” do experimento (conjunto de fatores) é aplicado para se obter as variáveis resposta (dependentes) citadas anteriormente. É através desta unidade que será possível obter variação estatística na análise dos resultados da investigação.

Nesse sentido, foi criada uma unidade de experimento definida abaixo. A unidade consiste em uma prova de redação sobre o tema de inteligência artificial. Esse tema foi escolhido de forma aleatória e foi solicitado a criação de uma redação por um especialista. A proposta de redação criada está apresentada no Apêndice D.

Além da redação criada, também foi necessário criar um formulário de avaliação da redação, a ser utilizado na hora de sua correção. Esse formulário contém todos os critérios de avaliação que deverão ser analisados, bem como seus possíveis valores e condições. No caso, como o contexto do ambiente MeuTutor é ENEM, este já possui um formulário pré-definido onde suas redações são corrigidas. Nesse caso, optamos por utilizar esse formulário em nossa unidade de experimento. O formulário definido está apresentado no Apêndice E.

5.1.3.4 Seleção do Projeto Experimental

O experimento apresentando nesse trabalho é do tipo comparativo, portanto o projeto experimental escolhido deve pertencer à categoria de projetos comparativos. Os projetos dessa categoria são apropriados para pesquisas onde queremos tomar uma decisão a respeito de um fator importante (no caso deste experimento, o modelo de correção) com respeito a uma série de métricas, na presença ou não de outros fatores menos significativos (ambiente, por exemplo).

Os experimentos fatoriais são muito usados em experimentos envolvendo vários fatores para os quais é necessário estudar o efeito conjunto dos fatores sobre a resposta. No nosso caso, como temos fatores com apenas 2 níveis cada, podemos utilizar um experimento fatorial 2^k sem repetição, retirando uma combinação que é impossível de ser realizada (que é a utilização do modelo proposto M2 em nenhum ambiente educacional E2).

Cada execução do modelo tem um custo relativamente alto além de exigir bastante tempo para preparação do ambiente, alunos e demais esforços necessários para sua conclusão. Por este motivo, foi escolhido um projeto fatorial sem repetição. Nesse sentido temos apenas 3 ensaios possíveis, sendo executado cada um apenas uma vez, totalizando três tratamentos possíveis e conseqüentemente três execuções. A Tabela 8 descreve cada um dos tratamentos.

5.1.3.5 Plano de execução

Esta seção descreve como será a execução do experimento que foi planejado e descrito ao longo das seções anteriores. A execução do experimento envolve os seguintes passos:

Tabela 8 – Definição dos tratamentos.

Número do tratamento	Referência	Modelo utilizado	Ambiente
1	T1	M1	E1
2	T2	M2	E1
3	T3	M1	E2

Fonte: elaborado pelo autor.

1. Preparação do ambiente educacional Meututor ENEM:

Inicialmente, é necessário preparar o ambiente educacional Meututor com relação a unidade de experimento definida, isto é, cadastrar a redação junto com o formulário no ambiente. Após esse cadastro, o ambiente educacional será implantado em um servidor da amazon EC2 e disponibilizado na internet¹

2. Seleção dos usuários que irão participar do experimento:

O segundo passo consiste na seleção de alunos que irão participar do experimento. Nesse caso serão selecionados 30 (trinta) alunos que desejaram participar.

3. Agrupamento de usuários: Destes trinta usuários selecionados, deverão ser criados três grupos com 10 alunos, onde cada grupo executará um tratamento específico. Dois desses grupos utilizarão o MeuTutor para fazer as redações, enquanto que um grupo fará as redações da forma tradicional.

4. Seleção de especialistas: Em nosso experimento, precisamos que as correções das redações sejam feitas por especialistas em 2/3 dos casos (tratamento 1 e 3). Nesse caso, será necessário a seleção de 2 especialistas para corrigir as redações.

5. Seleção de mediador: É necessário uma pessoa que realize um processo de intermediação na execução do experimento, para garantir confiabilidade, fluidez e corretude do experimento proposto.

6. Execução do experimento: Os grupos de alunos serão instruídos e poderão executar seus tratamentos propostos. O mediador explicará a cada grupo o que deverá ser feito. Em seguida, parte das redações serão enviadas para os especialistas, para que estes as corrijam de acordo com seus tratamentos.

7. Coleta de dados: Os dados serão coletados após a execução do passo anterior do experimento. Nos tratamentos que não houver sistema envolvido, os dados serão obtidos via questionário. Nos tratamentos que houver sistema, os dados serão obtidos pelo ambiente. Porém, alguns dados, mesmo que seja um tratamento dentro do sistema, serão obtidos por questionário.

¹ Disponibilizado através do endereço <http://www.redacao.meututor.com.br>

8. **Análise dos dados:** Em seguida, deverá ser feita uma análise estatística dos dados. Os resultados de cada métrica obtidos na execução serão sumarizados na ferramenta R e organizados para exibição gráfica através de histogramas e/ou boxplot, com o propósito de facilitar a comparação. Quanto aos testes estatísticos, deseja-se realizar o teste de comparação paramétrico (ANOVA) caso o modelo matemático obtido seja adequado e cumpra os requisitos do teste. Se o modelo matemático não for adequado, será utilizado um teste não-paramétrico para testar as hipóteses. Em nosso caso, queremos um nível de confiança de pelo menos 95%.

Após a execução dos passos anteriores, os resultados poderão ser obtidos e comparados entre si. A análise dos dados bem como as principais conclusões obtidas estão apresentadas nas Seções 5.3 e 5.4, respectivamente.

5.1.3.6 Análise de ameaças à validade

Embora todo o experimento tenha sido planejado para minimizar possíveis ameaças que comprometam suas conclusões, existem algumas que devem ser mencionadas.

A primeira ameaça é com relação ao tamanho da amostra do experimento. O número escolhido de 30 alunos, resultando em 10 alunos para cada tratamento, é um número pequeno. O fato de não ter repetições de execução, devido aos custos, pode também influenciar em um baixo poder estatístico para a análise dos dados. Isso ocorreu devido às restrições de tempo e principalmente custos envolvidos para a execução desse experimento. Com isso, não foi possível a execução de mais repetições e conseqüentemente a análise desses dados.

É possível que os níveis dos fatores selecionados (modelo e ambiente educacional) não sejam suficientes para observar diferenças significativas de eficácia e/ou desempenho entre os modelos. Desse modo, temos uma ameaça à validade de construção causada pela confusão entre constructos e seus níveis. Essa limitação está melhor detalhada nas limitações dentro da seção de conclusão.

A unidade experimental da pesquisa foi selecionada a partir de uma única fonte, que pode ter características próprias que não valem para todos as demais questões discursivas. Logo, há uma ameaça à validade externa na interação de seleção e tratamento, o que pode dificultar a generalização dos resultados além do escopo estudado.

5.2 Narrativa de execução do experimento

Essa seção descreve o processo de execução do experimento. Em primeiro momento foi necessário configurar a plataforma MeuTutor ENEM com a redação criada pelo especialista junto com seu formulário de avaliação. Para configurar o MeuTutor ENEM para este experimento, é necessário configurar a camada de banco de dados, serviços e a interface.

Todas as camadas foram implantadas em um único servidor virtual da amazon EC2 (para reduzir os custos), com a seguinte configuração:

- Instância: m3.medium
- Processamento: vCPU: 1/ECU: 3
- Memória (GiB): 3.75
- Armazenamento da instância (GB): 1 x 4 SSD
- Sistema operacional: Ubuntu Server 14.04 LTS (HVM)

Além disso, foram necessários a instalação de alguns softwares nesse servidor para funcionamento correto do MeuTutor ENEM. Os softwares que foram instalados são:

1. Virtuoso RDF Store (opensource) - 06.01.3127
2. GlassFish Server Open Source Edition 4.1 (build 13)
3. Java(TM) SE Runtime Environment (build 1.8.0_40-b25)
4. Apache HTTP Server Project 2
5. PHP 5 + libapache2-mod-php5

O banco de dados da plataforma Meututor foi configurado através do software 1 e está disponibilizado online ². O virtuoso fornece uma interface de comunicação http na porta 1111, onde os serviços poderão se comunicar com o banco. A camada de serviços foi configurada usando o software 2. Também está disponível online para administração³ e visualização dos serviços⁴. Finalmente, a interface foi configurada usando os softwares 4 e 5 e está disponível na web⁵.

Após a configuração do ambiente MeuTutor, o próximo passo foi a escolha dos alunos que iriam participar do experimento. Para isso, foram selecionados 30 usuários do MeuTutor que gostariam de participar desse experimento. Os usuários são alunos recém ingressados em cursos de nível superior, sendo 20 alunos no curso de matemática vespertino e 10 alunos no curso de contabilidade noturno da ufal. Vale ressaltar que todos os usuários possuem o mesmo conhecimento na disciplina abordada neste experimento. Em seguida, foram convidados dois especialistas em correções de redações, indicados aqui por especialista 1 e especialista 2. Ambos possuem formação específica em nível superior, sendo um deles especialista em história.

² Disponibilizado através do link <http://redacao.meututor.com.br:8890>

³ Administração através do endereço <http://redacao.meututor.com.br:4848>

⁴ Visualização dos arquivos WSDL pelo endereço <http://redacao.meututor.com.br:8080/redacao/ws>

⁵ Para acessar a interface deve-se utilizar o link <http://redacao.meututor.com.br>

O grupo com alunos de contabilidade executaram o tratamento 3, isto é, fizeram a redação de forma manual (sem a utilização do ambiente E2) e a correção de suas redações foram feitas pelo especialista 1. Por outro lado, os alunos de matemática fizeram suas redações no ambiente Meututor (E1) através do link disponibilizado. As correções de suas redações foram feitas tanto utilizando o modelo proposto nesse trabalho (tratamento 2) quanto o modelo tradicional (tratamento 1), onde nesse caso, foi realizada pelo especialista 2.

No caso do tratamento 3, os alunos fizeram suas redações e entregaram ao mediador a folha de resposta. Este, ao recolher todas as folhas de respostas, entregou ao especialista 1 para sua avaliação. O especialista 1 teve como trabalho avaliar a redação e preencher um formulário online disponibilizado através do GoogleForms, uma vez que esse não tem acesso ao ambiente nesse tratamento. A Figura 42 apresenta parte do formulário disponibilizado ao especialista 1.

O primeiro campo representa um identificador (id) gerado para a redação. Os dados de avaliação contêm todos os critérios de avaliação do ENEM. Em seguida, também é perguntado ao especialista a duração da correção e o custo associado para a correção, conforme podemos ver na Figura 43.

Os dados são armazenados em planilhas online, fornecidas pelo próprio Google de depois exportados para planilhas em excel. No caso dos tratamentos 1 e 2, as redações foram feitas diretamente no ambiente MeuTutor. Isso facilitou o processo, uma vez que basta que o especialista entre no ambiente, evitando intermediários no processo. Dessa forma, em primeiro momento, os alunos criaram suas redações no ambiente ao longo de 3 dias. No total, foram criadas 18 redações (2 alunos desistiram no processo).

Em seguida, os próprios alunos, utilizando o ambiente com o modelo proposto, corrigiram as redações de seus pares ao longo de 3 dias também. No total foram realizadas 47 correções. O tempo dessas correções também foi armazenado. Finalmente, essas mesmas redações foram disponibilizadas ao especialista 2 no ambiente e este pode corrigi-las. Para cada correção, a nota final, o tempo e custo desse especialista foram obtidos.

5.3 Análise dos dados

O experimento foi executado seguindo todo o planejamento descrito anteriormente neste capítulo, conforme podemos ver a narração apresentada na Seção 5.2. Os dados necessários foram coletados e agora já podem ser analisados. Após a execução do experimento, os dados foram analisados por meio da ferramenta R, onde alguns scripts nessa linguagem tiveram de ser construídos para efetuar tal análise (esses scripts foram omitidos desse trabalho devido à grande quantidade de linhas de código).

Esta seção tem como objetivo apresentar, inicialmente, uma análise preliminar dos dados na Subseção 5.3.1. Em seguida, uma análise descritiva dos dados obtidos na execução do experimento. Os dados, nessa etapa, serão apresentados na Subseção 5.3.2 em

Formulário de avaliação da redação por especialista

Este formulário servirá para um especialista enviar sua avaliação de uma determinada redação, no modelo do ENEM

*Obrigatório

Dados da redação

Qual a redação você está corrigindo nesse momento? *

Digite apenas o que vem após "WritingEvaluationSolving_"

Dados da avaliação

Os dados da avaliação abaixo são definidos conforme o modelo do ENEM, divididos em 5 competências, variando entre 0 a 200 pontos.

1- Demonstrar domínio da modalidade escrita formal da Língua Portuguesa *

- 0
- 40
- 80
- 120
- 160
- 200

2- Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa *

- 0
- 40
- 80
- 120
- 160
- 200

Figura 42 – Formulário disponibilizado ao especialista para avaliação de redação fora do ambiente

Fonte: elaborado pelo autor

forma gráfica, através de histogramas para cada métrica em cada tratamento executado. Além disso, também serão apresentados diagramas de caixa comparativos (boxplots) e no final uma breve sumarização de algumas estatísticas analisadas sobre o conjunto de dados. Ao se realizar uma breve análise dos dados, é possível ter noção das respostas para as questões de pesquisa levantadas no experimento, além de permitir um melhor entendimento do comportamento das variáveis e detectar possíveis padrões e/ou inconsistências nos dados.

Dados da correção

Quanto tempo demorou para corrigir essa redação? *

Horas ▼ : Minutos ▼ : Segundos ▼

Quanto você cobraria para corrigir essa redação? *

Enviar

Nunca envie senhas em Formulários Google.

Figura 43 – Formulário disponibilizado ao especialista para avaliação de redação fora do ambiente - Parte 2

Fonte: elaborado pelo autor

Em seguida, após a apresentação e análise descritiva dos dados, serão apresentadas na Subseção 5.3.3 as verificações das hipóteses de pesquisa, que será feita com a análise dos intervalos de confiança e execução de testes de hipóteses, quando necessário.

5.3.1 Análise preliminar dos dados

Os dados foram organizados em planilhas de acordo com os tratamentos executados, conforma ilustra a Tabela 9.

Tabela 9 – Custo, nota e tempo dos tipos de avaliações (professor ou pares de alunos) e tipos de ambiente (MeuTutor ou Questionario on line).

Avaliação Feita pelo professor No ambiente Meu Tutor			Avaliação Feita pelos pares de alunos No ambiente Meu Tutor			Avaliação Feita pelo professor Questionario on line		
Nota	Custo	Tempo	Nota	Custo	Tempo	Nota	Custo	Tempo
680	8.97	1009	626	1.49	450	440	6.07	840
480	8.04	905	573	1.23	270	440	3.03	420
720	7.63	858	680	1.49	450	640	4.33	600
480	8.00	900	520	1.49	450	520	3.90	540
480	3.98	448	573	1.19	240	560	4.77	660
400	5.33	600	380	1.27	300	400	3.57	494
640	10.4	1175	600	1.78	645	680	6.62	917
680	16.2	1833	520	1.64	555	600	4.77	660
760	4.73	532	546	1.71	600			
⋮	⋮	⋮	⋮		⋮	⋮		

Fonte: elaborado pelo autor.

Como podemos ver (Tabela 9) variáveis quantitativas tipicamente assumem muito

valores. Sendo assim, inicialmente, precisamos sumarizar esses dados. Nesse sentido, podemos construir histogramas que representam a distribuição de frequência desses dados (BARBETTA; REIS; BORNIA, 2004), conforme apresentado na Figura 44.

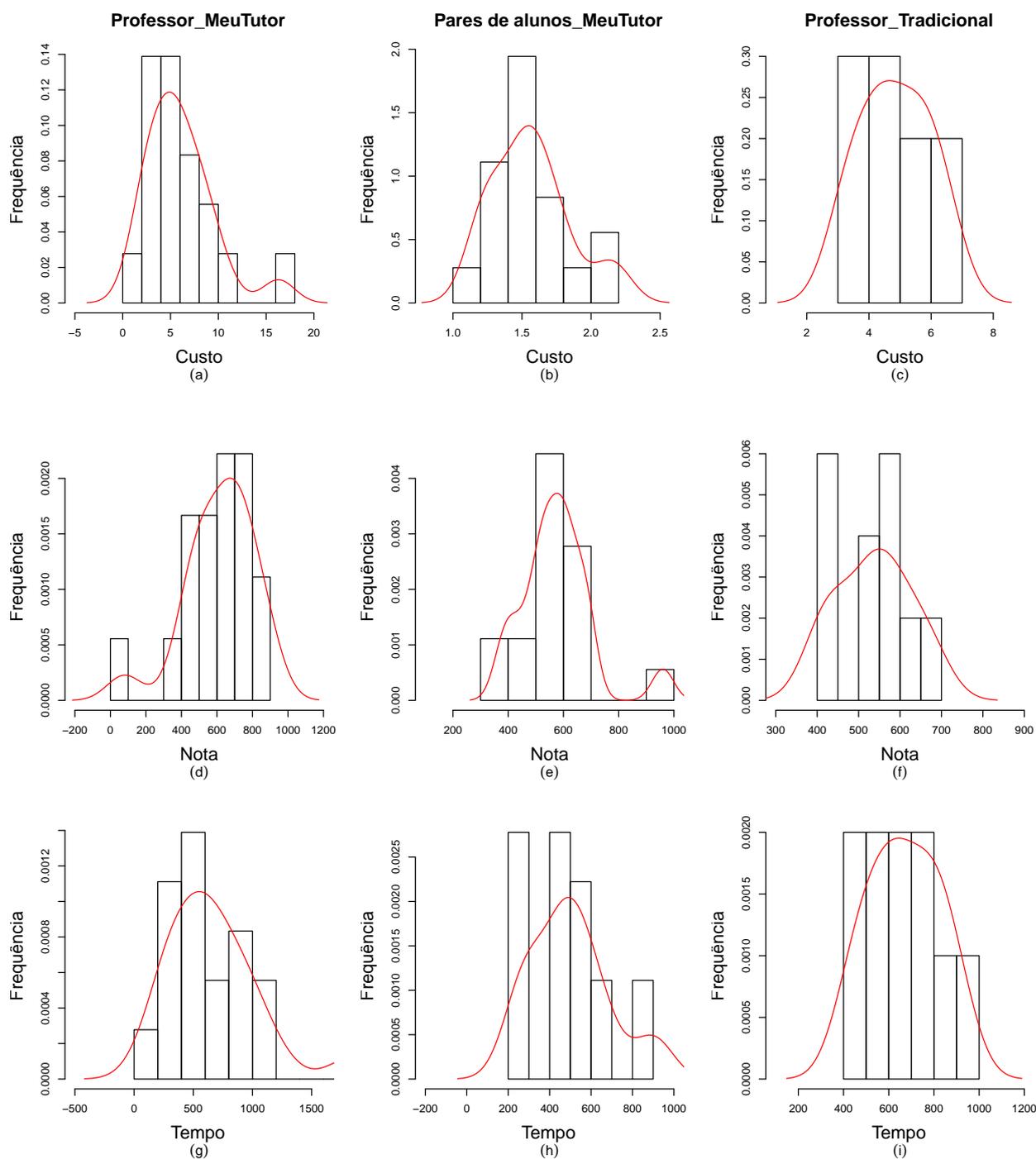


Figura 44 – Histograma com a representação das frequências em todas as métricas

Fonte: elaborado pelo autor

Com base nesses histogramas apresentados na Figura 44 percebemos por exemplo que os custos apresentam distribuições de frequências distintas entre si, em especial o custo das avaliações feita por pares de aluno no ambiente MeuTutor, inclusive na forma da

densidade estimada. As variáveis Notas e Tempos apresentam distribuições de frequência bem similares quanto aos três métodos para fazer as correções das questões discursivas. Ainda exploraremos mais estes histogramas na Seção 5.3.2.

Além de analisar as frequências de aparição dos dados também devemos observar algumas medidas de dispersão dos dados (Variabilidade). A variância, o desvio padrão e o coeficiente de variação são as medidas de dispersão mais utilizadas. A variância de uma variável aleatória é uma medida da sua dispersão, indicando quão longe em geral os seus valores se encontram do valor esperado (valor médio) e é calculada a partir de

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

onde \bar{Y} é a média da amostra (BARBETTA; REIS; BORNIA, 2004).

O desvio padrão define-se como a raiz quadrada da variância. O coeficiente de variação é uma medida de dispersão que se presta à comparação de distribuições diferentes. O desvio-padrão, uma medida de dispersão relativo à média e como duas distribuições podem ter médias diferentes, o desvio dessas duas distribuições não são comparáveis (BARBETTA; REIS; BORNIA, 2004). A solução é usar o coeficiente de variação, que é igual ao desvio-padrão dividido pela média: $CV_{\text{amostral}} = \frac{S}{\bar{Y}}$. O coeficiente de variação é ainda multiplicado por 100, passando a ser expressado como percentagem.

Representamos os resultados sobre a dispersão dos dados na Tabela 10. Com base nesta tabela podemos notar como o coeficiente de variação é útil para comparar a **dispersão** (variabilidade) de uma amostra de uma variável aleatória. Podemos perceber que o grupo com maior variabilidade é o associado ao método que utiliza o professor para a correção das redações associado ao MeuTutor. Se observássemos apenas o desvio padrão deste grupo para a variável custo, teríamos a impressão errônea que a dispersão do grupo é pequena.

No entanto, o coeficiente de variação igual a 58.9% revela que essa é a amostra com maior variabilidade entre as nove amostras estudadas. Não existe um limite para definir o quão dispersa é uma amostra, mas quanto menor seu coeficiente de variação (porém diferente de zero) mais informação a amostra contém sobre o problema estudado.

Tabela 10 – Medidas de dispersão das amostras do experimento com pares de alunos.

Grupos	Custo			Nota			Tempo		
	Var	Desvio	Coef. var.	Var	Desvio	Coef. var.	Var	Desvio	Coef. var.
T1	12.6	3.6	58.9%	37673.2	194.1	31.5%	159959.3	399.9	58.8%
T2	0.1	0.2	18.1%	17227.8	131.3	22.7%	37435.3	193.5	38.6%
T3	1.3	1.2	23.9%	8248.9	90.8	16.9%	25508.5	159.7	23.9%

Fonte: elaborado pelo autor.

5.3.2 Análise descritiva dos dados

Como mencionado anteriormente, as variáveis e níveis definidos no planejamento do experimento foram:

- **Nota final da avaliação (N):** Nota da correção variando entre 0 e 1000
- **Tempo por correção(T):** Tempo de cada correção em minutos
- **Custo por correção(C):** Custo de cada correção

A seguir, serão analisados os dados de cada uma das variáveis estudadas. A primeira variável analisada é a nota final (N). A Figura 45 apresenta o diagrama de caixa para a métrica da nota final N com relação aos tratamentos T1, T2 e T3.

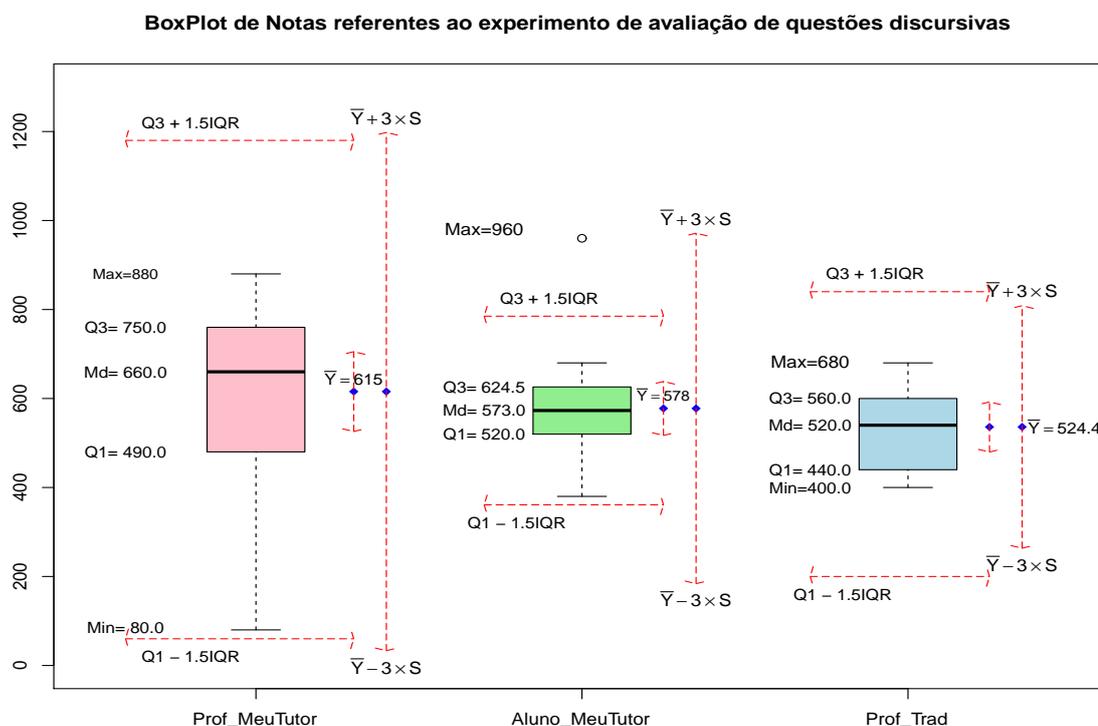


Figura 45 – Diagrama de caixa com comparativo da métrica nota para os tratamentos T1, T2 e T3

Fonte: elaborado pelo autor

Podemos ver na figura que a média (média (T1)=615 contra média(T2)=578) e mediana (mediana(T1) = 660 contra mediana(T2)=573), nos dois casos, estão bem próximas. Esta figura sugere que as notas obtidas com a aplicação desses tratamentos possuem variações estatísticas semelhantes, o que nos indica uma certa semelhança nos tratamentos. No

entanto, ainda não foram geradas evidências estatísticas para afirmar isso. Tais afirmações só poderão ser feitas quando os testes estatísticos forem realizados (especificamente na subseção 5.3.3).

Podemos observar também que a média do tratamento T3 ($média(T3)=524.4$) e a mediana ($mediana(T3)=520$) ainda continua próxima dos resultados obtidos no tratamento T2, mesmo tendo um número menor de redações corrigidas nesse tratamento se comparado com os outros dois. Essa proximidade da média e mediana dos outros tratamentos, indicando bastante semelhança, contudo nada se pode afirmar no momento, conforme dito anteriormente.

Como uma de nossas questões de pesquisa é avaliar as notas obtidas pelo modelo proposto comparadas com as notas obtidas pelo modelo tradicional, então podemos criar um diagrama de caixa com os dados da diferença entre as notas obtidas por ambos os modelos. A Figura 46 apresenta o diagrama de caixa para a diferença entre as notas obtidas pelos modelos diferentes quando executados dentro do ambiente (T1 x T2).

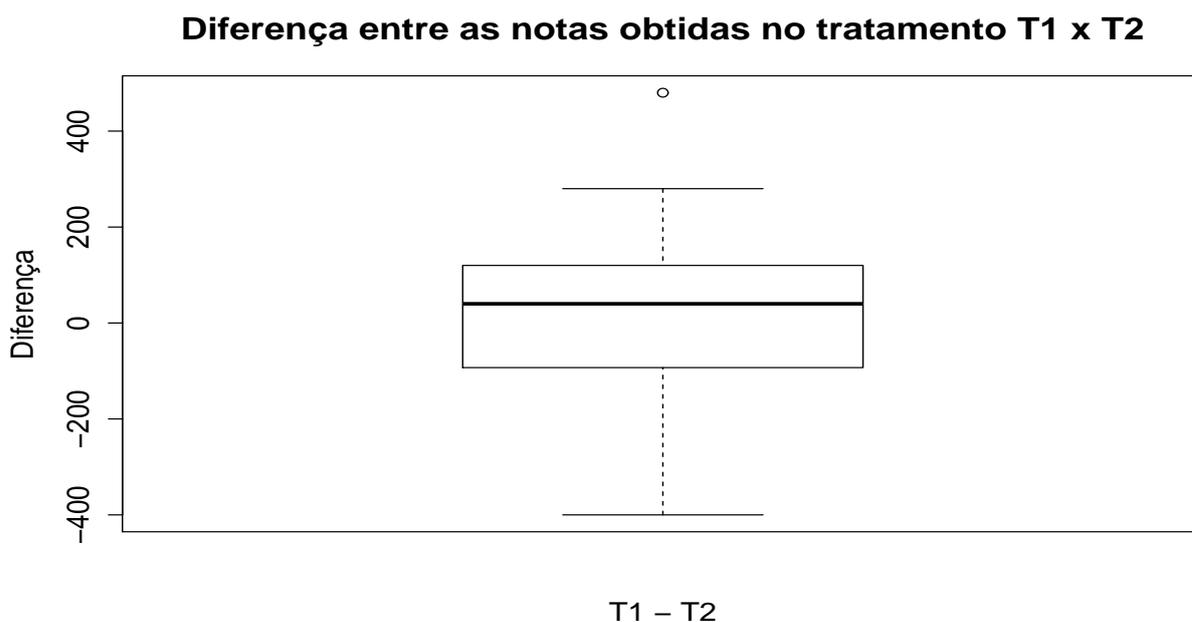


Figura 46 – Diagramas de caixa com a diferença da métrica nota nos tratamentos T1 x T2

Fonte: elaborado pelo autor

Podemos observar na figura que a diferença das notas do tratamento T1 e T2, na média ($média(T1 - T2)= 37,8$) e mediana ($mediana(T1 - T2)= 40$), estão bem próximas de 0 (zero), indicando que são bem equivalentes. Contudo há algumas diferenças um pouco mais significativas, incluindo um outlier. Essa diferença pode ter sido ocasionada pela escala de 40 pontos utilizada para atribuição das notas no formulário de avaliação. Isso pode justificar o fato de que a média e mediana ficaram perto desse valor. Podemos

também considerar a hipótese de que dois corretores, mesmo sendo especialistas, podem atribuir notas diferentes, porém próximas, devido à alta subjetividade das avaliações.

No Geral, a diferença entre as notas T1 e T2 variou entre -79,75 (o resultado negativo indica que a nota do especialista no tratamento T1 foi menor que a nota do modelo no tratamento T2) a 110 (resultado positivo indica que a nota do modelo foi menor que a nota do especialista). Se considerarmos a escala de notas entre 0 e 1000, temos que as diferenças entre as notas variam entre 7.9% a 11%. O comportamento dessa diferença será analisado posteriormente.

Além da análise do diagrama de caixa, podemos analisar os dados do ponto de vista de suas frequências de aparição através da criação de histogramas. As Figuras 47, 48 e 49 apresentam, respectivamente, os histogramas dos dados nos tratamentos T1, T2 e T3, para a métrica nota.

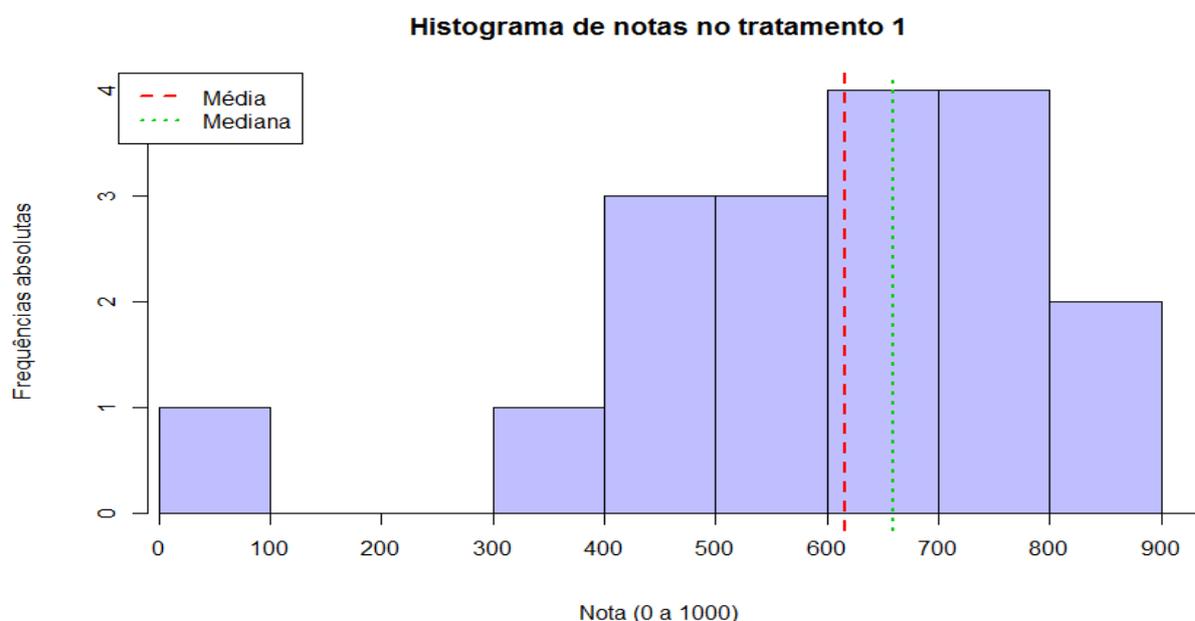


Figura 47 – Histograma para a métrica nota no tratamento T1

Fonte: elaborado pelo autor

Podemos observar nestes histogramas que a distribuição dos dados para o tratamento T1 e T2, desconsiderando o outlier, tem características próximas a uma distribuição normal. No caso do tratamento T3 não aparenta ser normal e nem possui uma distribuição conhecida. Contudo, só com métodos estatísticos será possível afirmar se a distribuição desses dados é normal ou não. Vale ressaltar que a verificação de normalidade dos dados é pré-requisito para escolha de qual teste de hipótese estatístico será utilizado para analisar as hipóteses.

Da mesma forma que fizemos com os boxplots, podemos também analisar os histogramas de forma combinada, na tentativa de encontrar semelhanças entre os dados. A

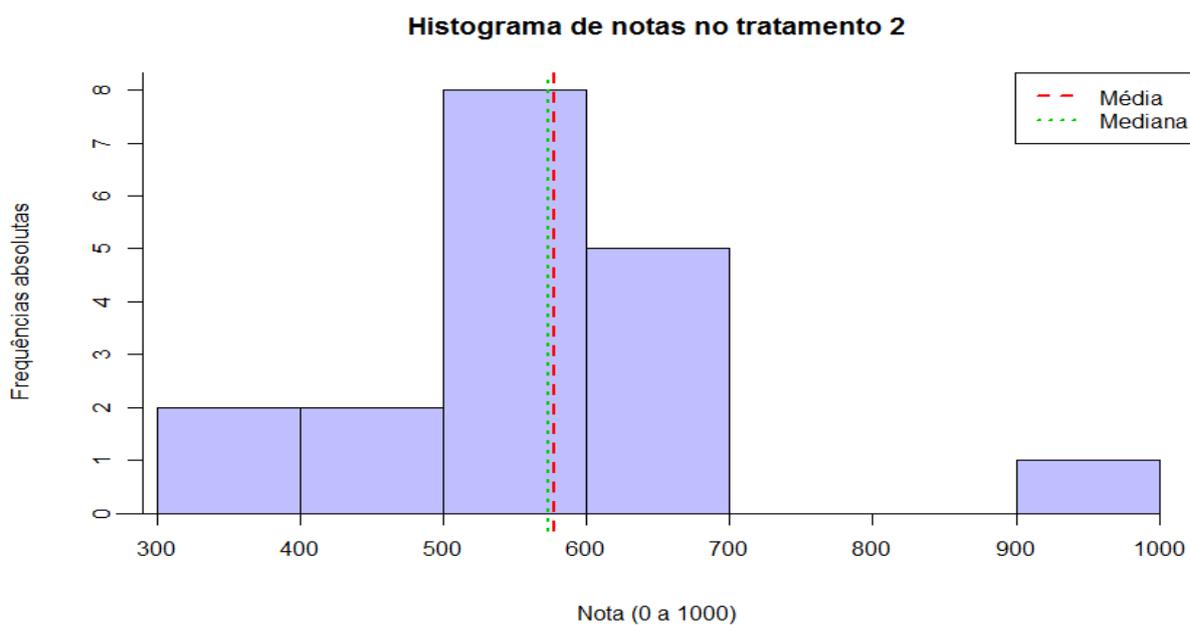


Figura 48 – Histograma para a métrica nota no tratamento T2

Fonte: elaborado pelo autor

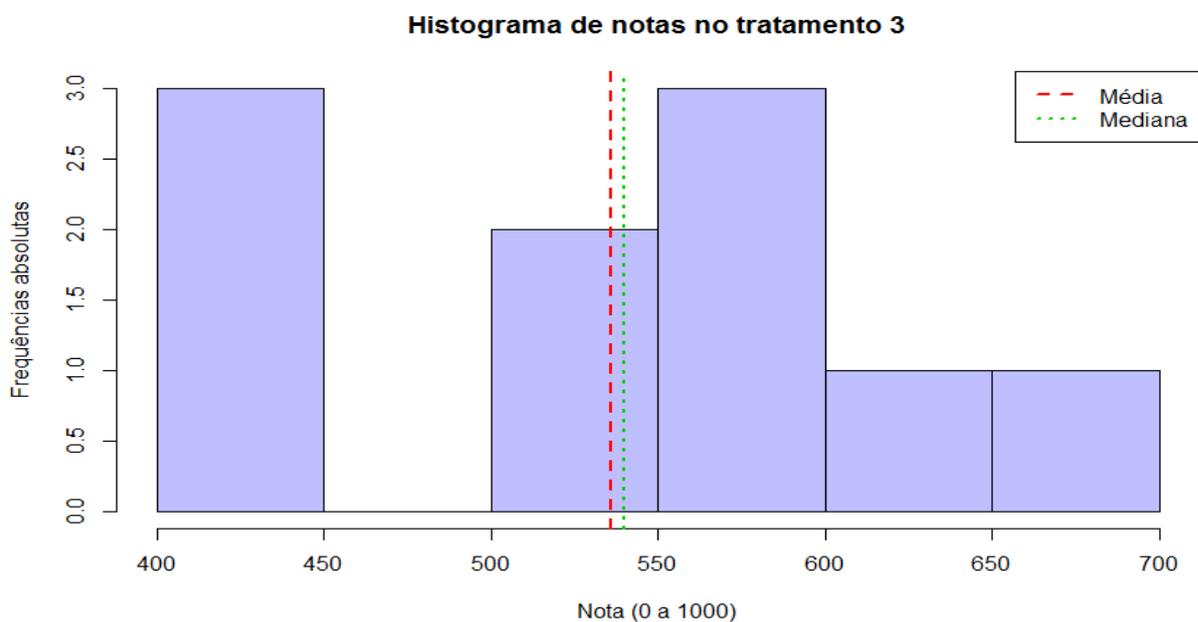


Figura 49 – Histograma para a métrica nota no tratamento T3

Fonte: elaborado pelo autor

Figura 50 apresenta os histogramas dos tratamentos T1 (apresentado na cor azul) e T2 (apresentado na cor vermelha). A combinação apresentada na figura, entre o tratamento T1 e T2, foi escolhida de forma proposital para analisarmos os resultados da aplicação do

modelo no ambiente e a aplicação do método tradicional.

Através da figura, podemos observar que as tendências dos dados são muito semelhantes, com frequências muito parecidas, indicando uma possível equivalência entre os tratamentos. Há a existência de dois outliers, tanto no tratamento T1 quanto no tratamento T2. No geral, a média e a mediana ficaram próximas dos 600 pontos, sendo a média de T2 um pouco menor que a média de T1.

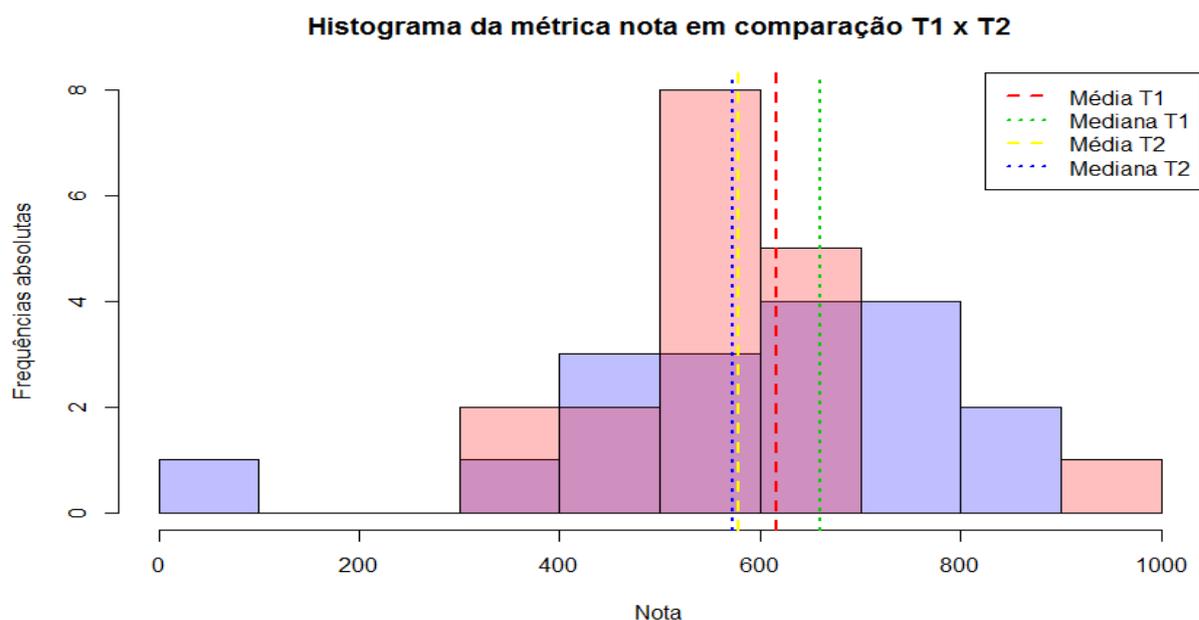


Figura 50 – Histograma para a métrica nota na combinação T1 (azul) x T2 (vermelho)

Fonte: elaborado pelo autor

Em resumo, os dados apresentados nos gráficos para a métrica nota (N) obtidos com a execução dos tratamentos (T1, T2 e T3) estão sumarizados na Tabela 11, para uma melhor visualização.

Tabela 11 – Sumarização dos dados relativos a variável Nota (N).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	80.0	490.0	660.0	615.6	750.0	880.0
T2	380.0	520.0	573.0	577.7	624.5	960.0
T3	400.0	460.0	540.0	536.0	590.0	680.0
T1 - T2	-400.00	-79.75	40.00	37.89	110.00	480.00

Fonte: elaborado pelo autor.

A próxima métrica a ser analisada é o tempo gasto para correção (T). A análise será feita de forma semelhante a análise já apresentada na métrica de custo. Em primeiro momento será apresentado o diagrama de caixa comparativo com os dados obtidos em

relação ao tempo e, em seguida, alguns histogramas com os dados de cada tratamento em específico.

Vale ressaltar que o tempo para correção inclui todo o tempo decorrido desde que a redação foi entregue ao corretor/avaliador até quando sua nota foi devolvida (seja ao sistema [tratamento T1] ou ao formulário de avaliação apresentado [caso do tratamento T3]). No caso em que há mais de um corretor (como exemplo do tratamento T2), os dados de tempo que foram utilizados na análise foram os dados do corretor mais lento, uma vez que, supondo que os corretores estejam corrigindo a mesma redação ao mesmo tempo (em paralelo), a nota final só poderá ser obtida quando o último corretor terminar sua avaliação, sendo necessário, portanto, esperar seu tempo para obtenção da nota final.

A Figura 51 apresenta um conjunto de diagramas de caixa para a métrica do tempo (T) com relação aos tratamentos T1, T2 e T3.

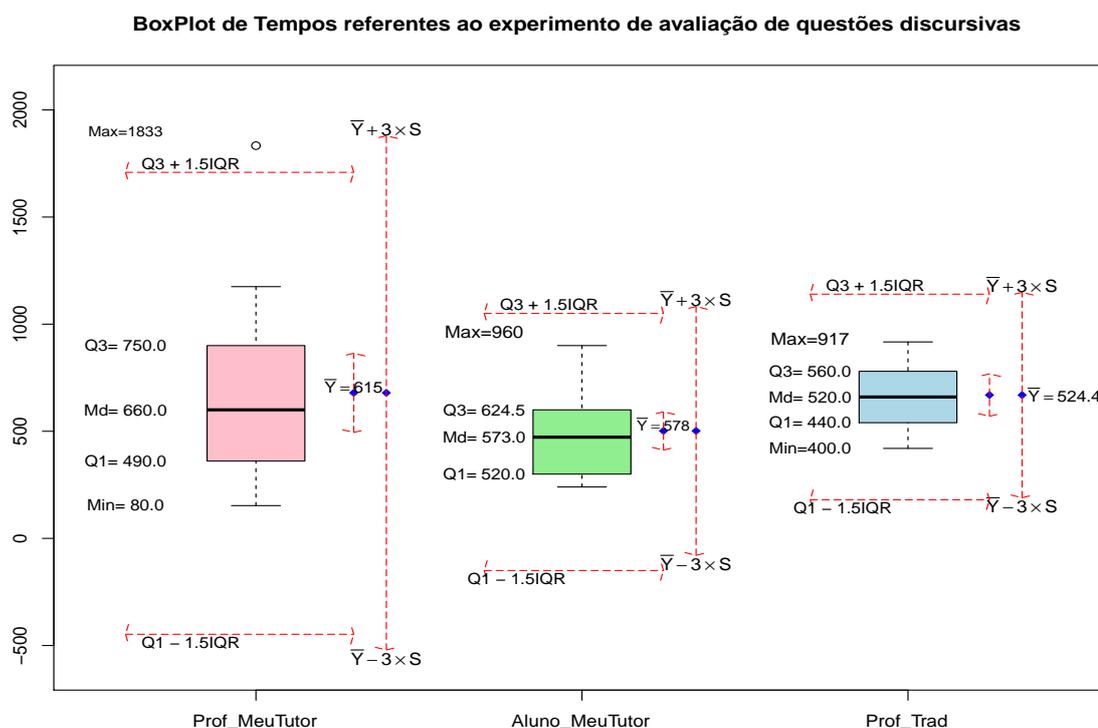


Figura 51 – Diagrama de caixa com comparativo da métrica tempo para os tratamentos T1 , T2 e T3

Fonte: elaborado pelo autor

Ao se analisar os dados de média e mediana dos tratamentos em que há um professor especialista envolvido (média(T1)=679,3 / média(T3)=669,1 e mediana(T1)=600 / mediana(T3)=660), podemos verificar que os dados são bem semelhantes, contudo apre-

sentando uma variação/dispersão maior nos dados no tratamento T1 em relação aos dados do tratamento T2.

Se compararmos esses dados com os dados do tratamento T2 (média(T2)=501,6 / mediana(T2)=472,5), notamos que o tempo no tratamento T2 é significativamente menor que o tempo nos outros dois tratamentos. Contudo, como os diagramas se interceptam, é necessário verificar o intervalo de confiança dos dados e a execução dos testes de hipóteses para ter uma conclusão estatisticamente válida.

Também foram criados histogramas com os dados referentes à métrica Tempo, para uma análise mais detalhada deles. As figuras 52, 53 e 54 apresentam, respectivamente, os histogramas da métrica tempo referentes aos dados obtidos na execução dos tratamentos T1, T2 e T3.

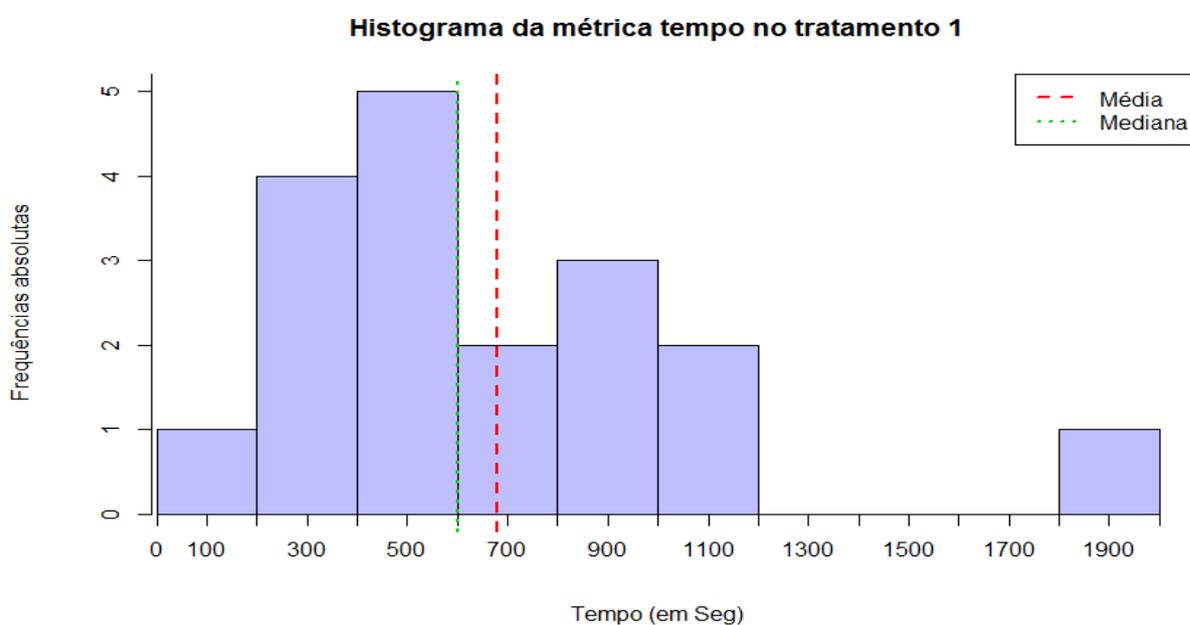


Figura 52 – Histograma para a métrica tempo no tratamento T1

Fonte: elaborado pelo autor

Podemos observar, nas figuras de histograma da métrica Tempo apresentadas, que os dados do tempo no tratamento T1 tem características semelhantes a uma distribuição normal, se desconsiderarmos o outlier apresentado. Da mesma forma, ao analisarmos os dados do tratamento T2, verificamos que ele possui uma mistura entre uma distribuição normal com uma distribuição com cauda longa a esquerda. Finalmente, os dados do tratamento T3 indicam que ele possui uma distribuição uniforme dos dados. Contudo, em todos os casos são necessárias a execução de testes de normalidade nos dados antes da execução dos testes estatísticos, para confirmar se os dados apresentam ou não uma distribuição normal.

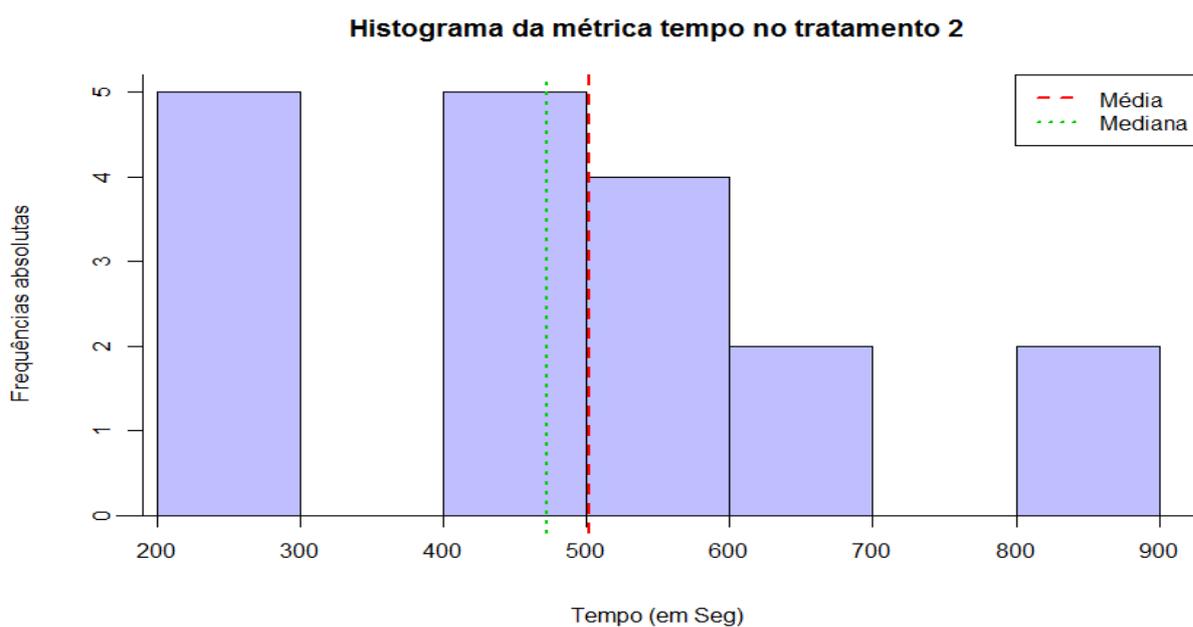


Figura 53 – Histograma para a métrica tempo no tratamento T2

Fonte: elaborado pelo autor

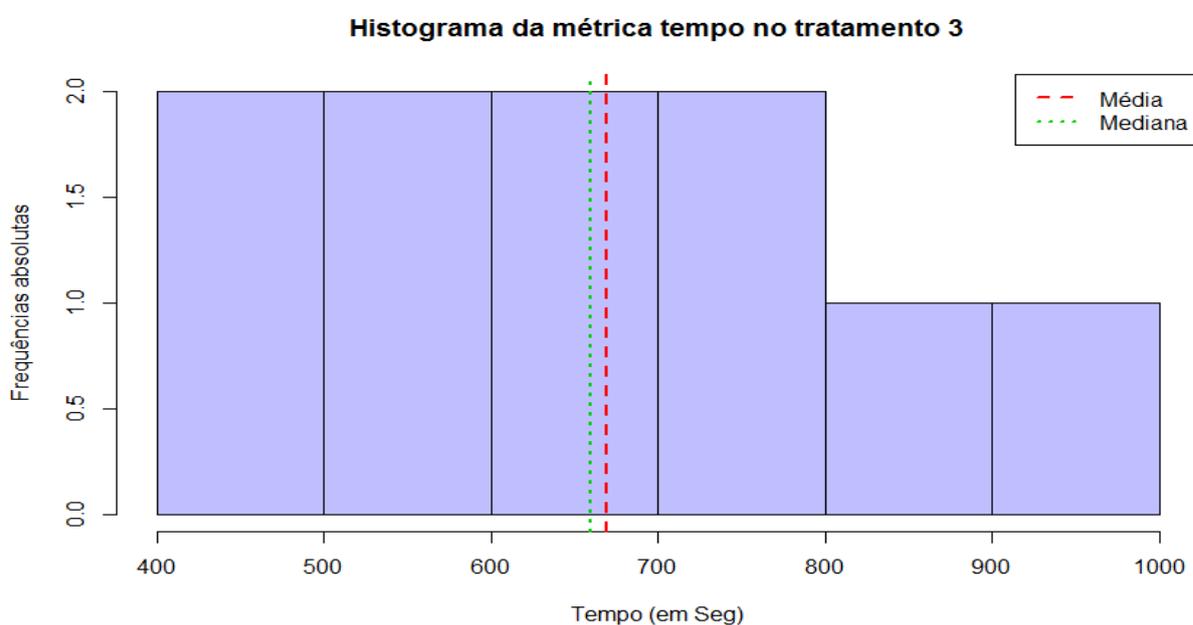


Figura 54 – Histograma para a métrica tempo no tratamento T3

Fonte: elaborado pelo autor

De forma resumida, os dados apresentados nos gráficos para a métrica Tempo (T) obtidos com a execução dos tratamentos (T1, T2 e T3) estão sumarizados na Tabela 12, para uma melhor visualização.

Tabela 12 – Sumarização dos dados relativos a variável Tempo (T).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	153.0	382.8	600.0	679.3	889.5	1833.0
T2	240.0	337.5	472.5	501.7	588.8	900.0
T3	420.0	555.0	660.0	669.1	780.0	917.0

Fonte: elaborado pelo autor.

Finalmente, a última métrica a ser analisada é o custo da correção (C). O custo envolvido na correção está fortemente ligado a utilização de métodos tradicionais, uma vez que a utilização do modelo não gera custos adicionais. Para efeitos de comparação do modelo, vamos usar o custo da solução implantada dividido pelo número de usuários que efetivamente fizeram a correção, como o custo por usuário do tratamento T2.

Nesse sentido, o custo para a implantação do sistema em um servidor da amazon é R\$0,21/hora. Como o processo foi realizado ao longo de seis dias (3 dias para escrever a redação somados com mais 3 três dias para o processo de correção), temos que o custo envolvido nos 6 dias, ou nas 144 horas, foi equivalente a R\$ 30,24. Dividindo esse valor por 18 usuários que efetivamente fizeram a redação e as correções, temos que o custo por usuário foi R\$ 1,68.

Por outro lado, o custo de um especialista avaliando as redações (no caso dos tratamentos T1 e T3) se torna bastante elevado. Os especialistas geralmente cobram por hora de trabalho de acordo com sua formação. Nesse caso, eles indicaram quanto seria cobrado para cada redação corrigida baseado no tempo que levou para sua correção. Os dados foram então armazenados. A figura 55 apresenta o diagrama de caixa para a métrica de custo envolvido (C) com relação aos tratamentos executados.

Podemos observar que as médias e medianas nos tratamentos T1 e T3 (média(T1)=R\$6.03 / média(T3)=R\$4,83) são parecidas, uma vez que ambos os tratamentos utilizam especialistas nas correções. A variação existente entre elas se dá pelo fato de que os especialistas cobraram preços diferentes. Há um espalhamento (variação) maior nos dados do tratamento T1, enquanto que esse comportamento não ocorre no tratamento T3, indicando que o custo por usuário no tratamento T3 é mais uniforme que o custo por usuário no tratamento T1. Geralmente ocorre quando o avaliador mantém sempre um tempo fixo médio em suas correções. Isso foi ocasionado pela variação nos tempos (conforme podemos ver nos dados de tempo dos tratamentos T1 e T3 da Figura 51).

Por outro lado, o custo do tratamento T2 se apresenta fixo, e como podemos observar na Figura 55, se encontra significativamente abaixo dos outros dois tratamentos, o que pode nos indicar que o custo ao utilizar o modelo proposto no ambiente (T2) é menor que o custo da correção por especialistas usando ou não o ambiente (T1 e T3). Contudo, como já discutido anteriormente, somente com os testes estatísticos podemos concluir tais

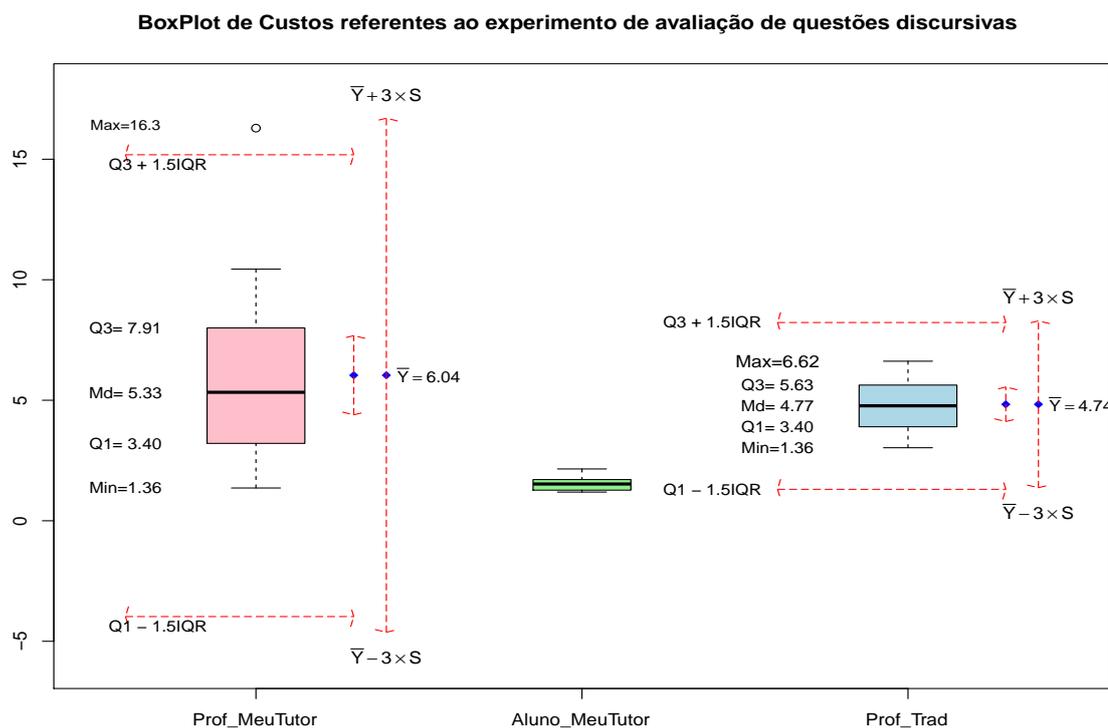


Figura 55 – Diagrama de caixa com comparativo da métrica custo entre os tratamentos T1, T2 e T3

Fonte: elaborado pelo autor

afirmações.

Além da análise do diagrama de caixa para a métrica Custo (C), podemos analisar estes dados do ponto de vista de suas frequências de aparição com os histogramas. As Figuras 56 e 57, mostram, respectivamente os histogramas da métrica custo correspondentes aos dados da execução dos tratamentos T1 e T3.

Podemos observar na Figura 56, referente ao histograma de custo com os dados do tratamento T1, que os dados apresentados possuem uma distribuição próxima a uma distribuição normal, se desconsiderarmos o outlier a direita. Por outro lado, observamos na Figura 57, referente ao histograma com os dados de custo do tratamento T3, que a distribuição apresentada nesses dados se assemelha como uma mistura da distribuição uniforme e a distribuição com cauda longa a esquerda. Porém, mais uma vez, testes de análise da amostra devem ser feitos para validar a normalidade ou não das distribuições apresentadas nos dados antes de qualquer teste estatístico ser realizado.

De forma resumida, os dados apresentados nos gráficos para a métrica custo (C) obtidos com a execução dos tratamentos (T1, T2 e T3) estão sumarizados na Tabela 13, para uma melhor visualização. Vale ressaltar que os dados foram arredondados para duas

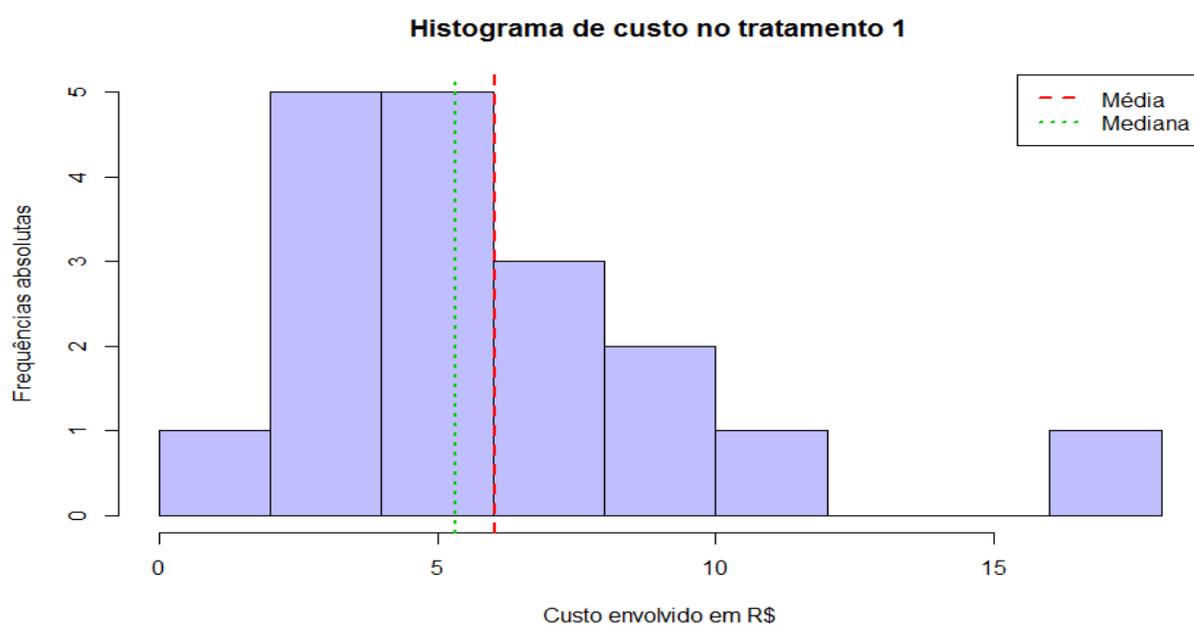


Figura 56 – Histograma para a métrica custo no tratamento T1

Fonte: elaborado pelo autor

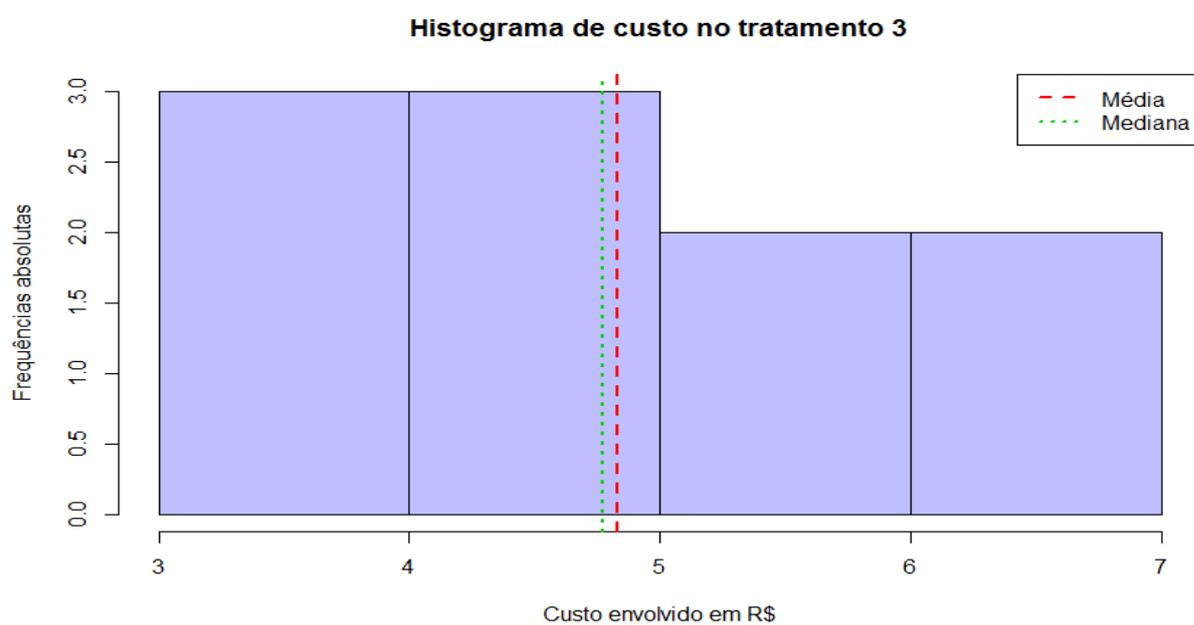


Figura 57 – Histograma para a métrica custo no tratamento 3

Fonte: elaborado pelo autor

casas decimais.

Tabela 13 – Sumarização dos dados relativos a variável Custo (C).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	1.36	3.40	5.33	6.03	7.90	16.29
T2	1.68	1.68	1.68	1.68	1.68	1.68
T3	3.03	4.01	4.77	4.83	5.63	6.62

Fonte: elaborado pelo autor.

5.3.3 Verificação das hipóteses

A seção anterior apresentou uma análise descritiva usando o conjunto de dados que foram obtidos com a execução dos tratamentos definidos no experimento proposto. Após essa análise preliminar dos dados, é necessário a realização de uma análise mais aprofundada com o objetivo de verificar a validade das hipóteses de pesquisa apresentadas na Seção 5.1.3.1.

Podemos ressaltar que para cada hipótese levantada na seção, precisamos indicar qual tratamento foi melhor. Para realizar tal verificação serão necessários alguns testes estatísticos e uma análise sobre os intervalos de confiança em cada métrica de avaliação definida.

Um intervalo de confiança (IC) é um intervalo estimado de um parâmetro de interesse de uma população. Em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis. O quanto estas estimativas são prováveis será determinado pelo coeficiente de confiança $(1 - \alpha)$, para $\alpha \in (0, 1)$. Este intervalo estimado é calculado a partir de um determinado conjunto de dados de amostra. Se essas amostras independentes são obtidas repetidamente da mesma população, então, uma determinada percentagem (nível de confiança) dos intervalos irá incluir o parâmetro populacional desconhecido (JAIN, 2008). Vale ressaltar que o nível de confiança escolhido em nossas análises deste experimento, conforme seu planejamento, é de 95%, ou seja, considerando um α (nível de significância) de 5% ou 0,05. Esses valores são os mais utilizados em análise estatística.f

A primeira tarefa para se realizar a verificação de uma determinada hipótese de pesquisa é analisar a normalidade da distribuição dos dados que estão envolvidos em sua resposta. A normalidade dos dados é importante pois determina qual teste estatístico deve ser utilizado na análise. Em caso afirmativo, isto é, os dados são normais, deve-se ser utilizado um teste paramétrico. Caso contrário (onde os dados não são normais), o teste deve ser não-paramétrico.

Existem alguns testes estatísticos para verificar a normalidade dos dados, porém o mais recomendado pelos estatísticos é o teste de Shapiro-Wilk (SHAPIRO; FRANCA, 1972). Por esse motivo, será utilizado nos cálculos. Basicamente, esse teste consiste na refutação ou não das seguintes hipóteses:

- H_0 : A amostra provem de uma população normal
- H_1 : A amostra não provém de uma população normal

A decisão é baseada em Rejeitar H_0 ao nível de significância α se $W_{calculado} < W_\alpha$ (os valores críticos da estatística W_α de Shapiro-Wilk são disponibilizados em uma tabela) e $P_{valor} < \alpha$. Caso $W_{calculado} > W_\alpha$ e $P_{valor} > \alpha$, não conseguimos rejeitar H_0 e, portanto, a distribuição dos dados é normal.

Em seguida, precisamos escolher um teste paramétrico e um não paramétrico a ser utilizado após o resultado do teste de Shapiro. Para casos de distribuições normais, utilizaremos o T-Test (WELCH, 1938). Para distribuições não normais, será utilizado o teste de Wilcoxon (WILCOXON, 1945). Estas escolhas foram motivadas pela simplicidade desses testes e pela implementação dos mesmos pela ferramenta R, a qual está sendo utilizada em toda a análise dos dados.

A primeira verificação de hipótese será feita com relação a métrica de nota. Conforme a Seção 5.1.3.1, as hipóteses nula e alternativa são, respectivamente:

H1-0: O uso do modelo de avaliação por pares proposto não é equivalente ao método tradicional.

H1-1: O uso do modelo de avaliação por pares proposto é equivalente ao método tradicional

A primeira análise a ser feita é observando os intervalos de confiança dos tratamentos executados. A Figura 58 apresenta esses intervalos de confiança para os tratamentos T1, T2 e T3 com 5% de significância.

A Figura 58 apresenta os intervalos de confiança de cada tratamento comparando sob o efeito da métrica nota com 5% de significância. Podemos observar que o intervalo de confiança do tratamento T2 se sobrepõe quase que totalmente se compararmos com os intervalos de confiança dos outros dois tratamentos T1 e T3, que usam a forma tradicional de correção por especialista. Dada esta sobreposição, ainda não há como fazer qualquer afirmação sobre a equivalência dos modelos aplicados nos tratamentos. Dessa forma, há uma necessidade real de verificar as hipóteses por meio de um teste de hipótese comparativo.

Nesse caso, em primeiro momento, precisamos aplicar o teste de Shapiro-Wilk para descobrir a normalidade ou não dos dados de cada tratamento. Os resultados da aplicação do teste de Shapiro-Wilk quando executado sobre os dados da métrica nota nos tratamentos estão apresentados na tabela 14.

Ao analisar esses dados, pode-se notar que os valores de $W_{calculado}(T1) = 0.92223 > W_\alpha(T1) = 0.897$ e $P_{valor}(T1) = 0.1416 > \alpha = 0.05$. Portanto, não é possível refutar a hipótese nula do teste de Shapiro e, conseqüentemente, podemos afirmar com nível de significância de 5% que a amostra T1 provém de uma população normal.

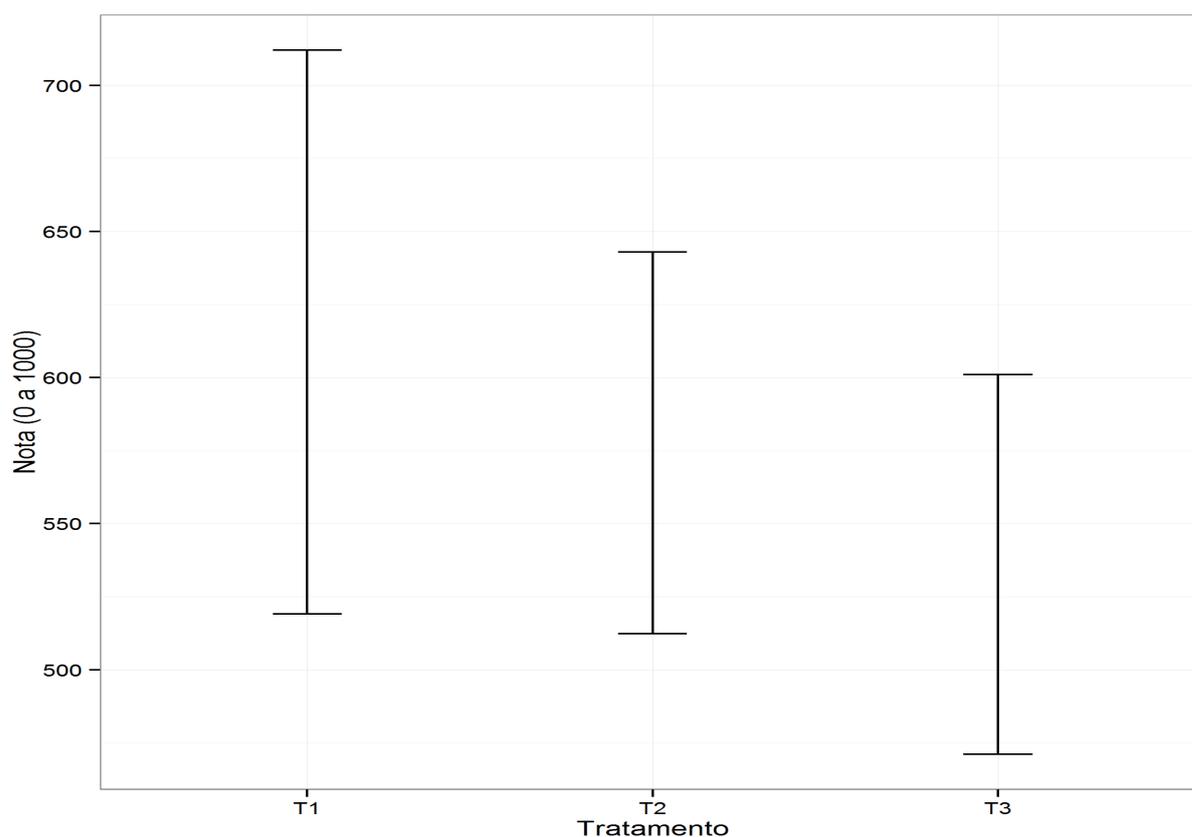


Figura 58 – Intervalo de confiança da métrica nota para os tratamentos T1, T2 e T3

Fonte: elaborado pelo autor

Tabela 14 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Nota

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.92223	0.897	0.1416	0.05
T2	0.90344	0.897	0.06599	0.05
T3	0.96201	0.842	0.8086	0.05

Fonte: elaborado pelo autor.

Da mesma forma, se analisarmos os dados do tratamento T2, temos que $W_{calculado}(T2) = 0.90344 > W_{\alpha}(T2) = 0.897$ e $P_{valor}(T2) = 0.06599 > \alpha = 0.05$. Assim como acontece no tratamento T1, a hipótese nula do teste de Shapiro não pode ser refutada e a distribuição dos dados de T2 também é normal. Por fim, o mesmo comportamento acontece com os dados da amostra T3, uma vez que $W_{calculado}(T3) = 0.96201 > W_{\alpha}(T3) = 0.842$ e $P_{valor}(T3) = 0.8086 > \alpha = 0.05$. Portanto, conclui-se que os dados da amostra T3 também se apresenta como uma distribuição normal.

Dessa forma, como as distribuições dos dados de todos os tratamentos são distribuições normais, podemos aplicar o T-Test, comparando os tratamentos. Os dados da execução do T-test estão apresentados na Tabela 15.

Tabela 15 – Resultado da aplicação do T-test com os dados da métrica Nota

Tratamento	t	df	p-value
T1 x T2	0.85383	17	0.4051
T1 x T3	1.4728	25.546	0.153
T2 x T3	0.98704	24.523	0.3333

Fonte: elaborado pelo autor.

Podemos observar que os valores de $p\text{-value}(T1 \times T2) = 0.4051$, $p\text{-value}(T1 \times T3) = 0.153$ e $p\text{-value}(T2 \times T3) = 0.3333$, obtidos com a execução do T-test, são todos maiores que $\alpha = 0.05$. Portanto, com 95% de confiança, estatisticamente, não pode se afirmar que os valores das notas obtidos entre os tratamentos possuem diferenças significativas entre si, isto é, não há evidências estatísticas que mostrem a não equivalência das notas.

O fato desses coeficientes serem maiores que o α indicam que os resultados das notas são equivalentes (iguais), não sendo possível gerar evidência estatística suficiente para refutar a hipótese nula $H1-0$, implicando que **O uso do modelo de avaliação por pares proposto é equivalente ao método tradicional.**

A próxima verificação de hipótese será feita com relação a métrica Tempo (T). Conforme a Seção 5.1.3.1, as hipóteses nula e alternativa são, respectivamente:

H2-0: O uso do modelo de avaliação por pares proposto não traz diferenças de tempo em relação ao método tradicional

H2-1: O uso do modelo de avaliação por pares proposto traz diferenças significativas de tempo em relação ao método tradicional

Da mesma forma como fizemos com a métrica nota, em primeiro momento será apresentado um gráfico com os intervalos de confiança dos dados relacionados ao tempo obtidos com a execução dos tratamentos indicados. A Figura 59 ilustra os intervalos de confiança dos tratamentos T1, T2 e T3 referentes a métrica tempo, com um valor de significância igual a 5%.

Da mesma forma como aconteceu com a métrica nota, os intervalos de confiança da métrica tempo nos tratamentos se sobrepõem. Contudo, se compararmos o intervalo de confiança do tratamento T2 (no caso do modelo proposto) com o intervalo de confiança do tratamento T1 (modelo tradicional), vemos que apesar da sobreposição, há uma ligeira vantagem para o tratamento T2. O mesmo ocorre se compararmos os dados do tratamento T2 com os dados do tratamento T3, porém, graficamente, a vantagem do T2 ainda é maior nesse caso. Isso sugere que o tempo no tratamento T2 tende a ser menor que nos outros dois tratamentos. Se analisarmos T1 comparado com T3, temos uma sobreposição de T3 completamente em T1, e, portanto, sugere que são semelhantes, isto é, não há evidência estatística que mostre diferenças entre T1 e T3.

Contudo, seguindo a mesma análise dos intervalos de confiança da métrica nota, de-

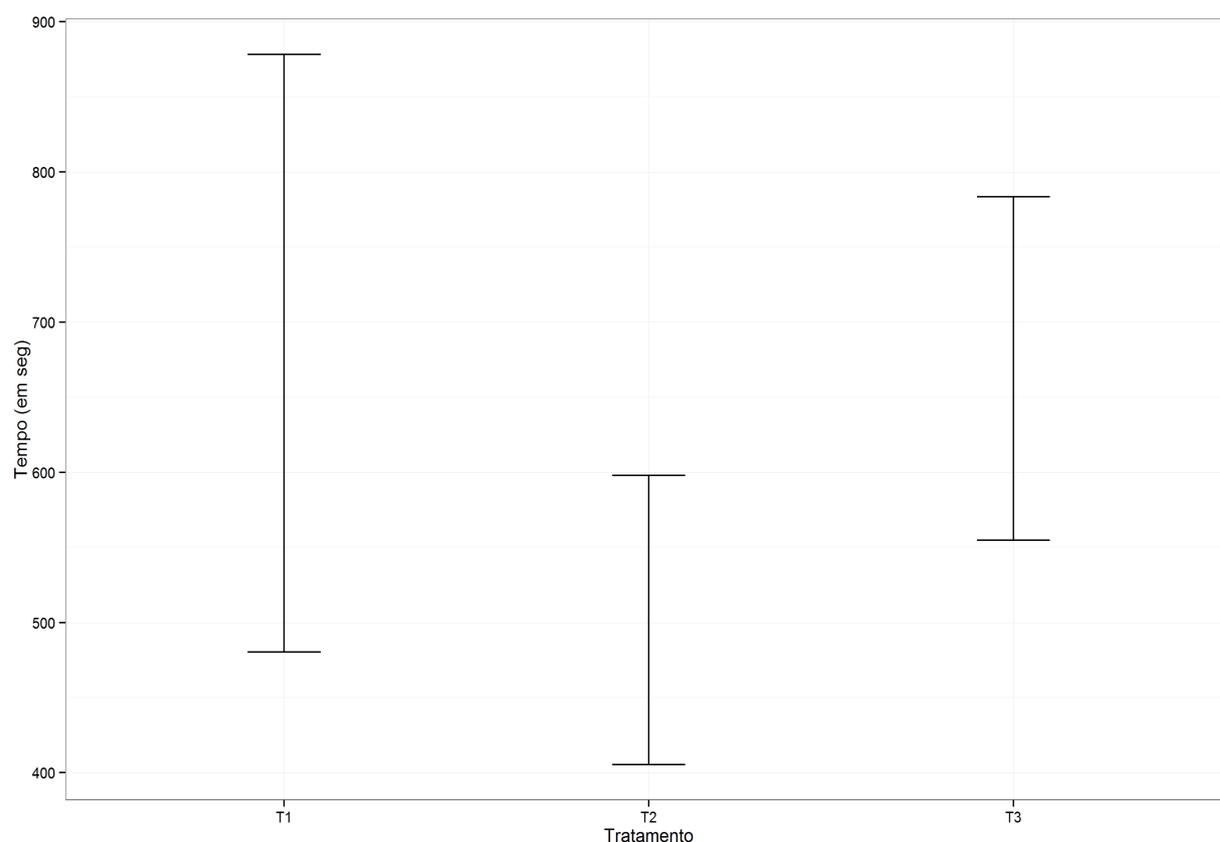


Figura 59 – Intervalo de confiança da métrica tempo para os tratamentos T1, T2 e T3

Fonte: elaborado pelo autor

vido as sobreposições, não tem como se fazer qualquer afirmação apenas com esta análise gráfica, fazendo-se necessário executar um teste de hipótese comparativo. Nesse caso, aplicou-se o teste de Shapiro-Wilk com os dados desses tratamentos para analisar a normalidade dos dados e o resultado pode ser visto na Tabela 16.

Tabela 16 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Tempo

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.89637	0.897	0.04969	0.05
T2	0.91724	0.897	0.1155	0.05
T3	0.97249	0.842	0.9129	0.05

Fonte: elaborado pelo autor.

Ao se analisar os dados da tabela, temos um caso especial no tratamento T1, uma vez que $W_{calculado}(T1) = 0.89637 < W_{\alpha}(T1) = 0.897$ e $P_{valor}(T1) = 0.04969 < \alpha = 0.05$. Com isso, podemos refutar a hipótese nula do teste de Shapiro-Wilk e consequentemente, podemos afirmar com nível de significância de 5% que a amostra não provém de uma população normal.

Se analisarmos os dados do tratamento T2, temos que $W_{calculado}(T2) = 0.91724 > W_{\alpha}(T2) = 0.897$ e $P_{valor}(T2) = 0.1155 > \alpha = 0.05$. Nesse caso, devido aos dois valores serem maiores, não conseguimos refutar a hipótese nula do teste de Shapiro e, portanto, os dados desse tratamento, com significância de 5%, provém de uma população normal. Por fim, os dados do tratamento T3 seguem o mesmo comportamento de T2, uma vez que $W_{calculado}(T3) = 0.97249 > W_{\alpha}(T3) = 0.842$ e $P_{valor}(T3) = 0.9129 > \alpha = 0.05$. Portanto, pode-se concluir que a amostra T3 também se apresenta como uma distribuição normal.

Nesse caso, como os dados do tratamento T1 possui uma distribuição não-normal, então quando os dados de comparação envolverem T1 deveremos usar um teste não-paramétrico, como o Teste de Wilcoxon. Nesse caso, apenas a combinação T2 x T3 envolverá o T-test, uma vez que ambas são distribuições normais. Os dados da execução desses testes estão apresentados na Tabela 17.

Tabela 17 – Resultado da aplicação dos testes estatísticos com os dados da métrica Tempo

Tratamento	Teste Utilizado	v/w	t	df	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	122	-	-	0.1169
T1 x T3	Wilcoxon rank sum test with continuity correction	81.5	-	-	0.7007
T2 x T3	Welch Two Sample t-test	-	-2.4605	21.938	0.02222

Fonte: elaborado pelo autor.

Podemos observar, de acordo com a Tabela 17, que como $p\text{-value}(T1 \times T2) = 0.1169 > \alpha = 0.05$, então não é possível concluir que há diferenças estatísticas entre esses tratamentos. Contudo, ao analisarmos o p-value dos tratamentos T2 e T3, $p\text{-value}(T2 \times T3) = 0.02222 < \alpha = 0.05$, podemos concluir que há diferenças estatísticas significativas e podemos concluir, com 95% de confiança, que o tempo em T2 é menor que o tempo em T3. Se analisarmos a combinação do tratamento T1 e T3 (onde ambos usam especialistas), temos um p-value $(T1 \times T3) = 0.7007 > \alpha = 0.05$ e, portanto, não há diferença estatística com a confiança desejada.

Como a nossa avaliação se baseia no comportamento dos dados do tratamento T2 (onde o modelo é aplicado), temos estatisticamente que o tempo envolvido nele é menor que o tempo do tratamento T3 (especialista sem ambiente) com 95% de confiança, porém não temos evidências estatísticas suficientes que mostrem que o tempo de T2 seja diferente do tempo em T1 (especialista no ambiente). De qualquer forma, há evidências estatísticas suficientes de que o uso do modelo proposto traz diferenças de tempo em relação ao método tradicional e com isso, pode-se refutar a hipótese nula H2-0 (que não haveria diferença) e aceitar a hipótese alternativa H2-1, com 95% de confiança, implicando que **o uso do modelo de avaliação por pares proposto traz diferenças significativas de tempo em relação ao método tradicional.**

Finalizando a verificação de hipóteses, serão analisadas as hipóteses que envolvem a

métrica de custo (C). De acordo com a Seção 5.1.3.1, as hipóteses nula e alternativa são, respectivamente:

H3-0: O uso do modelo de avaliação por pares proposto não traz diferenças de custo em relação ao método tradicional

H3-1: O uso do modelo de avaliação por pares proposto traz diferenças significativas de custo em relação ao método tradicional

Conforme o procedimento realizado nas duas métricas anteriores, em primeiro momento, é necessário analisar os intervalos de confiança da métrica custo com os dados obtidos nos tratamentos executados. A Figura 60 apresenta tais intervalos de confiança com significância de 5%.

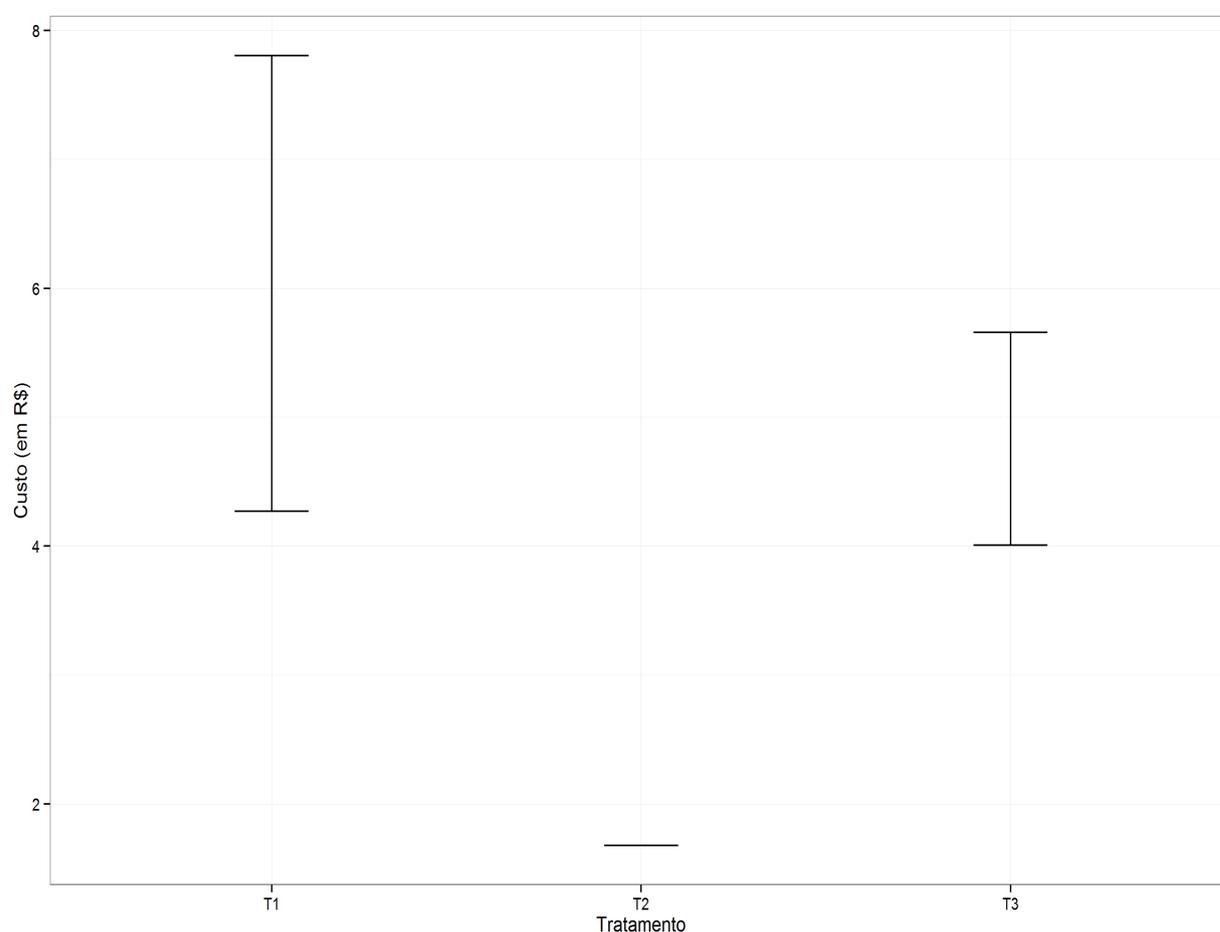


Figura 60 – Intervalo de confiança da métrica custo para os tratamentos T1, T2 e T3

Fonte: elaborado pelo autor

Como se pode observar, e diferentemente dos outros dois gráficos de intervalo de confiança, nesse gráfico há evidência estatística visualmente. Podemos observar que, no caso do tratamento T2, não há uma sobreposição em seu intervalo se comparado com os intervalos dos tratamentos T1 e T3. O fato de que não há sobreposição entre os intervalos

de confiança e que os dados do tratamento T2 estão claramente inferiores aos outros dois, já nos indica que o custo envolvido nesse tramito é menor que o custo dos outros dois.

Dessa forma, pode-se afirmar apenas visualmente com 95% de confiança que o custo do tratamento T2 é menor e conseqüentemente diferente do custo em T1 e T3, refutando, assim, a hipótese nula H3-0, implicando na aceitação da hipótese alternativa H3-1 com 95% de confiança.

Mesmo a análise gráfica dos intervalos de confiança sendo suficiente para refutar a hipótese nula H3-0 e a conseqüente aceitação da hipótese alternativa H3-1, verificaremos tal resultado aplicando os testes de hipótese comparativos, seguindo o mesmo procedimento adotado na avaliação das outras duas métricas.

O primeiro passo então é aplicar o teste de Shapiro-Wilk para analisar se as distribuições apresentadas nos dados são ou não normais, indicando qual teste estatístico utilizar. Os resultados da aplicação do teste de Shapiro-Wilk com os dados dos tratamentos T1, T2 e T3 estão ilustrados na Tabela 18.

Tabela 18 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Custo

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.89635	0.897	0.04965	0.05
T2	0.25269	0.897	1.057e-08	0.05
T3	0.97276	0.842	0.9152	0.05

Fonte: elaborado pelo autor.

Analisando os dados da tabela, observamos agora dois casos de anormalidade dos dados. No caso do tratamento T1, temos que $W_{calculado}(T1) = 0.89635 < W_{\alpha}(T1) = 0.897$ e $P_{valor}(T1) = 0.04965 < \alpha = 0.05$. Com isso, é possível refutar a hipótese nula do teste de Shapiro-Wilk e conseqüentemente, os dados da amostra T1 apresentam uma distribuição não-normal (anormal). O mesmo comportamento ocorre com os dados do tratamento T2, uma vez que $W_{calculado}(T2) = 0.25269 < W_{\alpha}(T2) = 0.897$ e $P_{valor}(T2) = 1.057e-08 < \alpha = 0.05$, logo a amostra T2 também é uma distribuição não-normal.

Por outro lado, ao analisarmos os dados do tratamento T3, vemos que $W_{calculado}(T3) = 0.97276 > W_{\alpha}(T3) = 0.842$ e $P_{valor}(T3) = 0.9152 > \alpha = 0.05$. Dessa forma, não conseguimos evidência estatística para refutar a hipótese nula de Shapiro-Wilk e, portanto, os dados da amostra T3 apresentam uma distribuição normal.

Contudo, devido ao fato de que os dados dos tratamentos T1 e T2 não serem normais, então devemos usar o teste de Wilcoxon, mesmo que a distribuição dos dados de T3 seja normal. Isso acontece porque o T-test só deve ser utilizado se as duas distribuições comparadas forem normais. Os resultados da aplicação do Teste de Wilcoxon com os dados desses tratamentos estão apresentados na Tabela 19.

Tabela 19 – Resultado da aplicação do teste de Wilcoxon com os dados da métrica Custo

Tratamento	Teste Utilizado	v/w	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	170	0.0002522
T1 x T3	Wilcoxon rank sum test with continuity correction	103	0.5486
T2 x T3	Wilcoxon rank sum test with continuity correction	0	1.111e-06

Fonte: elaborado pelo autor.

Com os dados da tabela, podemos observar que o $p\text{-value}(T1 \times T2) = 0.0002522 < \alpha = 0.05$, implicando que há diferenças estatísticas significativas entre o custo dos tratamentos T1 e T2. Com isso, podemos concluir que o custo do tratamento T2 é menor que o custo do tratamento T1 com uma confiança de 95%.

Da mesma forma, se analisarmos o valor de $p\text{-value}(T2 \times T3) = 1.111e-06 < \alpha = 0.05$, concluímos que o custo também são diferentes, isto é, com 95% de confiança temos que o custo de T2 é menor que o custo de T3. Por outro lado, o valor de $p\text{-value}(T1 \times T3) = 0.5486$ é maior do que $\alpha = 0.05$, o que implica que não há diferença significativa entre os tempos de T1 e T3.

Como nós queremos avaliar é o comportamento dos dados do tratamento T2 (onde o modelo é aplicado), temos comprovado estatisticamente que o custo envolvido em T2 é menor que os custos envolvidos tanto no tratamento T1 quanto no tratamento T3 com uma confiança de 95%. Com isso, temos evidências estatísticas suficientes de que o custo são diferentes e portanto, conseguimos refutar a hipótese nula $H3-0$ (onde não haveria diferença) e aceitar a hipótese alternativa $H3-1$, com uma confiança de 95%, implicando que **o uso do modelo de avaliação por pares proposto traz diferenças significativas de custo em relação ao método tradicional.**

Com a conclusão acima sabemos que o custo de aplicar o modelo é menor do que utilizar métodos tradicionais, contudo há uma necessidade de se precisar o quão menor é esse custo. Para isso criaremos um modelo de regressão (ESPINHEIRA; SILVA; SILVA, 2015) (FERRARI; CRIBARI-NETO, 2004). Considere o seguinte modelo linear:

$$y = X\beta + \epsilon = \tag{5.1}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}}_{\begin{pmatrix} x_1 & x_2 & \dots & x_k \end{pmatrix}} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Vemos que y é um vetor de n observações da variável aleatória dependente, ou ainda, y_1, \dots, y_n é uma amostra da população de interesse. X é uma matriz $n \times k$ formada

pelas covariadas. Note que cada coluna X é um conjunto de n observações da covariada x_t , $t = 1, \dots, k$, assim temos k covariadas. X não é variável aleatória, ela é observada e fixa. Ainda temos que, β é um vetor de k parâmetros também fixos e desconhecidos (não são variáveis aleatórias). Finalmente, ϵ é um vetor de n erros aleatórios com média zero $(E)(\epsilon_i) = 0$ e variância constante ao longo das observações, isto é, $\text{var}(\epsilon_i) = \sigma^2$ para todo $i = 1, \dots, n$.

Uma das principais suposições de modelos lineares de regressão é que

$$E(\epsilon) = \mu_\epsilon = 0 \quad (5.2)$$

Assim,

$$E(y) = E(X\beta) + E(\epsilon) \Leftrightarrow E(y) = X\beta \Leftrightarrow \mu = X\beta. \quad (5.3)$$

Então, **na prática** o nosso modelo final é

$$\mu = X\beta \quad (5.4)$$

$$\text{var}(y) = \underbrace{\text{var}(X\beta)}_0 + \underbrace{\text{var}(\epsilon)}_{\sigma^2} \Leftrightarrow \text{var}(y) = \sigma^2 \quad (5.5)$$

Podemos representar nosso modelo considerando a i -ésima observação como

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Para conhecermos de fato o modelo acima precisamos estimar $\beta_1, \beta_2, \dots, \beta_k$. Tipicamente fazemos isto usando o método de máxima verossimilhança. Assim, obtemos $\hat{m}u_i$ quando obtemos $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, tal que

$$\hat{\mu}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, \dots, n.$$

Agora podemos obter estimativas para μ_i .

Neste ponto, precisamos generalizar nosso modelo linear com o objetivo de permitir o uso de diversas distribuições além da normal. Fazemos isso considerando a expressão abaixo.

$$g(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}$$

▷ Aqui $g(\mu_i)$ é uma função de ligação, que conecta a média da variável resposta e o modelo envolvendo as covariadas e os β 's.

Por que usamos essa $g(\mu_i)$? Quando $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, temos que $g(\mu_i) = \mu_i$, ou seja g é o que chamamos de função identidade. Isto acontece no modelo normal porque assim como a resposta que pertence a todos os reais, isto é, $y \in (-\infty, +\infty)$, o mesmo ocorre com sua média $\mu \in (-\infty, +\infty) = \mathbb{R}$. Quando obtemos $\hat{\mu}_i$ a mesma coisa deve acontecer,

ou seja, $\hat{\mu}_i \in (-\infty, +\infty) = IR$. Daí, note que $\hat{\mu}_i$ pode assumir QUALQUER valor Real. Assim, $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ estão LIVRES para também assumir qualquer valor.

Isto não acontece por exemplo se a variável resposta segue uma distribuição gama. Porque, assim como a resposta que pertence só assume valores reais positivos, isto é, $y \in (0, +\infty)$, o mesmo ocorre com sua média $\mu \in (0, +\infty) = IR^+$ e DEVE acontecer com $\hat{\mu}$, ou seja, tem que acontecer: $\hat{\mu} \in (0, +\infty) = IR^+$. Com esta restrição os $\hat{\beta}$'s não estão livres, pois temos que garantir que $\hat{\beta}X$ só assumam valores reais positivos. Em X não podemos mexer (fixa e conhecida) então teríamos que realizar um processo de estimação do β 's com restrição, para garantir que $\hat{\beta}X \in (0, +\infty) = IR^+$, Esse processo pode ser bastante complicado.

ENTÃO a alternativa é aplicar uma função g em μ_i de forma que $g(\mu_i) \in (-\infty, +\infty) = IR$. Daí então os $\hat{\beta}$'s estão **LIBERADOS**. Vamos ao exemplo da gama, seja y uma variável aleatória com distribuição gama, aqui denotada por $Y \sim G(\mu, \phi)$, tal que

$$f_{(\mu, \phi)}(y) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu} \right)^{\phi} \exp\left(-\frac{\phi y}{\mu}\right) \frac{1}{y}, \quad y \geq 0, \mu > 0, \phi > 0, \Gamma(\phi) = \int_0^{\infty} t^{(\phi-1)} e^{-t} dt.$$

Se nossa variável resposta pertence aos reais positivos e decidimos usar a distribuição gamma, um modelo de regressão adequado seria

$$\log(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Isto porque, o logaritmo só pode ser calculado para valores reais positivos e o seu resultado assume valores em todos os reais, o que libera os $\hat{\beta}$'s. Note que $\log(u) \in (-\infty, +\infty)$ para todo $u \in (0, +\infty) = IR^+$.

No caso do custo no experimento, testamos o modelo

$$\log(\mu_i) = \beta_1 + \beta_2 \text{IndicadoraCustoT2}$$

em que μ_i é a média do Custo considerando todos os três grupos e IndicadoraCustoT2 é uma variável que assume o valor um quando o grupo é o de pares de alunos usando Meu Tutor e zero para os outros dois grupos.

Comandos no R

```
>fit <- glm(t(Custo)~t(Ind_T2), family=Gamma(link=log))# função para glm - generalized linear model -
modelos lineares generalizados - supondo distribuição gama para a resposta e usando a função de ligação log.
> summary(fit)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.7240      0.0809   21.31 < 2e-16 ***
t(Ind_T2)    -1.2751      0.1293   -9.86 1.03e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Gamma family taken to be 0.1832411)
Null deviance: 22.5610 on 45 degrees of freedom
Residual deviance: 7.1323 on 44 degrees of freedom
AIC: 157.56
```

Então, com base nos resultados acima temos que a covariável indicadora do grupo pares de aluno no MeuTutor é altamente significativa, a um nível maior que 0.00001. Ou seja, de fato método proposto de correções por pares tem um custo diferente dos custo dos outros dois grupos. Até já sabíamos isso, também sabíamos que era um custo menor, o que é confirmado pelo sinal negativo da estimativa. $\widehat{\beta}_2 = -1.2751$. Mas, agora temos uma informação adicional com base no modelo de regressão. Temos que $\{1 - \exp(\widehat{\beta}_2)\} \times 100$ representa **a redução percentual ao implantarmos o sistema de correção por pares de aluno com um valor igual a 72.4%**.

5.4 Principais conclusões

O experimento planejado, executado e descrito nesse capítulo teve como principal objetivo avaliar o modelo de avaliação por pares proposto no trabalho se comparado com as avaliações tradicionais (realizadas por especialistas), do ponto de vista de três métricas (nota (N), tempo (T) e custo (C)) importantes. Os dados de cada variável foram coletados ao longo da realização do experimento e permitiram que essas variáveis pudessem ser avaliadas separadamente.

De acordo com as análises desses dados (tanto a análise gráfica utilizada nos intervalos de confiança quanto a análise pelos testes de hipóteses comparativos - Shapiro, T-Test e Wilcoxon), foi possível verificar que:

- Não houve variações estatísticas significativas entre as notas obtidas tanto no modelo proposto quanto pelas correções usando método tradicional (professor corrigindo as atividades). Com isso, é possível utilizar o modelo proposto em substituição ao modelo tradicional nos ambientes e obter seus benefícios, sem comprometer o resultado final das avaliações.
- O tempo de correção usando o modelo proposto se mostrou mais rápido que o tempo de correção por especialista fora do ambiente (caso do tratamento T3), contudo não houveram evidências significativas que o tempo era menor que o especialista no ambiente (T1). Isso evidencia que alguns dos alunos submetidos no modelo proposto avaliaram a redação tão bem quanto um especialista levando, em alguns casos, até mesmo um tempo menor. Contudo, outros alunos levaram até mais tempo para corrigir.
- Houveram evidências estatísticas suficiente que indicaram que o custo envolvido do modelo proposto se mostrou menor que o custo do modelo tradicional. Esse resultado era esperado, uma vez que não eram necessários especialista para as correções das avaliações. Através da criação de um modelo de regressão apresentado, mostrou-se que a redução no custo é de cerca de 72.4%

6 VALIDAÇÃO - EXPERIMENTO GAMIFICAÇÃO

Um dos grandes diferenciais do trabalho apresentado é a aplicação de técnicas de gamificação no modelo de avaliação por pares na tentativa de trazer uma maior motivação para os alunos que precisam avaliar as atividades de seus pares. O experimento anterior tinha apenas o objetivo de verificar os benefícios trazidos pelo uso do modelo sob o ponto de vista do processo de avaliação por pares, não avaliando o aspecto motivacional dos estudantes envolvidos.

Este capítulo tem como principal objetivo apresentar o projeto e execução de um segundo experimento que tem como objetivo avaliar o impacto motivacional causado pela inclusão da gamificação no modelo apresentado no trabalho, visto que o uso desta técnica é um dos principais pontos diferenciais entre o modelo proposto e os trabalhos apresentados na literatura.

Diante do exposto, será apresentado em primeiro momento o projeto de experimento, o qual será detalhado na Seção 6.1. Em seguida, todo o processo de execução é discutido na Seção 6.2. Finalmente a Seção 6.3 mostra como os dados foram coletados e o processo de análise deles. Por fim, a Seção 6.4 apresenta as principais discussões acerca dos resultados obtidos com este novo experimento.

6.1 Projeto de experimento

Todo o experimento aqui apresentado será feito no ambiente MeuTutor-ENEM (descrito na Seção 4.2.1), em condições de configuração bem semelhantes às já apresentadas no capítulo anterior. Nesse caso, consideramos que o modelo apresentado já está incluído no ambiente educacional MeuTutor com o objetivo de incluir a disciplina de redação no ambiente devido à sua importância para o Exame Nacional do Ensino Médio (ENEM).

Os primeiros passos em uma proposta de experimento são delinear bem o problema que está sendo trabalhado e definir os objetivos da investigação. Através do estudo dos artigos na revisão sistemática, vimos que **um dos grandes problemas encontrados ao se utilizar uma abordagem de avaliação por pares é o aspecto motivacional dos alunos**, que pode influenciar de forma positiva ou negativa em suas atitudes e comportamentos, o que pode ser benéfico ou prejudicial ao modelo de avaliação por pares. Assim, esse aspecto motivacional deve ser considerado na hora da criação do modelo. Com isso, é preciso também **avaliar o impacto que a gamificação traz no processo de avaliação por pares** dentro do modelo proposto.

A pesquisa a ser realizada também é do tipo experimental, e **tem como objetivo avaliar o impacto da gamificação no modelo de avaliação por pares proposto** quando aplicado no MeuTutor-ENEM, com muitos usuários. Os dados serão coletados a

partir de um intervalo de tempo e analisados. Para avaliar motivação, é preciso analisar o comportamento dos estudantes dentro do sistema. Nesse sentido, os resultados das execuções serão analisados sob possíveis métricas de quantidade e frequência de acesso dos estudantes (logins), quantidade de cadastros, quantidade de redações realizadas, quantidade de redações corrigidas, quantidade de pontos e troféus ganhos e uso efetivo na plataforma.

Para avaliar tais métricas será realizada uma comparação entre os resultados obtidos pelo modelo junto aos alunos quando influenciados pela gamificação e quando não influenciados por ela. Após as definições do problema e objetivos do experimento, inicia-se a etapa de seu planejamento, a qual será apresentada em detalhes na Seção 6.1.1.

6.1.1 Planejamento Experimento

O planejamento do experimento é sua parte principal, tendo em vista que todos os aspectos essenciais que o compõe devem ser especificados. O modelo de experimento adotado é do tipo A/B. Como o nome já diz, duas versões são comparadas (A e B), as quais são idênticas exceto por uma variante que pode impactar o comportamento do utilizador. A versão A pode ser a versão utilizada atualmente (controle), enquanto a Versão B é a modificada (tratamento) (PURHONEN, 2012).

No caso do experimento aqui proposto, a variante A não apresentará nenhum elemento de gamificação descrito na Tabela 5 e será considerada a variante controle, enquanto que a variante B apresentará os elementos de gamificação, sendo considerada a variante modificada. O contexto do experimento continua sendo industrial, posto que será executado em um ambiente educacional da indústria, executando o modelo com um conjunto de dados de usuários reais existentes no ambiente. A seguir, todos os aspectos do planejamento do experimento serão apresentados.

Questões de pesquisa e hipóteses: O foco principal desse experimento é analisar o aspecto motivacional trazido pela inclusão de elementos de gamificação no modelo de avaliação por pares. Levando isto em consideração, definimos nossa questão de pesquisa, que visa responder se usar o modelo de avaliação por pares com gamificação realmente se diferencia de usar o modelo sem a gamificação, do ponto de vista motivacional. Assim, nossa questão de pesquisa é:

QP - O uso de elementos de gamificação no modelo de avaliação por pares traz diferenças significativas do ponto de vista motivacional aos alunos envolvidos?

Responder a essa pergunta exige um certo conhecimento de como avaliar a motivação dos alunos. Tendo em vista que motivação não é algo que possa ser mensurado facilmente, nossa avaliação será baseada nos comportamentos esperados pelos alunos dentro do sistema, avaliando se os elementos de gamificação incentivaram os alunos a participarem mais das atividades impostas (se cadastrar, acessar a plataforma, realizar e corrigir redações, por exemplo).

Nesse sentido a pergunta principal pode ser quebrada em perguntas secundárias. Tendo em vista à quantidade de acessos dos alunos no sistema, podemos formular a seguinte hipótese de pesquisa:

P1 - O uso de elementos de gamificação no modelo de avaliação por pares apresenta diferenças significativas na quantidade de acessos dos alunos?

O que nos leva às seguintes hipóteses:

H1-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de acessos dos estudantes na plataforma.

H1-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de acessos dos estudantes na plataforma.

Tendo em vista à realização de redações dos alunos no sistema, podemos formular a seguinte hipótese de pesquisa:

P2 - O uso de elementos de gamificação no modelo de avaliação por pares apresenta diferenças significativas na quantidade de redações feitas pelos alunos?

O que nos leva às seguintes hipóteses:

H2-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações realizadas na plataforma.

H2-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de redações realizadas na plataforma.

Finalmente, analisando às correções de redações feitas pelos alunos no sistema, podemos formular a seguinte hipótese de pesquisa:

P3 - O uso de elementos de gamificação no modelo de avaliação por pares apresenta diferenças significativas na quantidade de redações corrigidas pelos alunos?

O que nos leva às seguintes hipóteses:

H3-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações corrigidas na plataforma.

H3-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de redações corrigidas na plataforma.

Por fim, para que os resultados finais da pesquisa sejam melhor compreendidos, seria interessante conhecer as interações entre os fatores e as variáveis respostas. Dessa forma, desejamos saber qual dos fatores do experimento é mais responsável pela variação nas saídas obtidas nas execuções dos tratamentos. Assim, a última pergunta de pesquisa do experimento é definida do seguinte modo:

P4 - Qual o fator que mais explica a variação nas variáveis-resposta da pesquisa?

- Para responder essa pergunta será utilizada estimação de parâmetros, com alocação de variação para os fatores.

Fatores e variáveis de resposta: Como variável independente temos:

- **Modelo de avaliação** - o modelo de avaliação que será utilizado em determinado momento.

Esse fator apresenta dois níveis, conforma a Tabela 20.

Tabela 20 – Níveis dos fatores do experimento de gamificação

Fator	Nível	Descrição
Modelo	M1	Modelo sem elementos de gamificação
Modelo	M2	Modelo com elementos de gamificação

Fonte: elaborado pelo autor.

As variáveis dependentes (métricas) são definidas abaixo:

- **Acesso (A):** Esta variável representa a quantidade de logins do usuário na plataforma.
- **Redações feitas(RF):** Esta variável representa a quantidade de redações realizadas pelo usuário.
- **Redações corrigidas(RC):** Esta variável representa a quantidade de redações corrigidas pelo usuário.

Formalmente, as hipóteses descritas anteriormente podem ser definidas conforme a Tabela 21. As funções A, RF e RC, apresentadas na tabela, retornam respectivamente, o valor dos acessos, quantidade de redações feitas e quantidade de redações corrigidas, com relação a utilização do modelo sem elementos de gamificação (M1) ou a utilização do modelo com elementos de gamificação (M2).

Tabela 21 – Definição formal das hipóteses de pesquisa do experimento de gamificação

Hipótese	Hipótese Nula	Hipótese Alternativa
H1	H1-0: $A(M1) = A(M2)$	H1-1: $A(M1) \neq A(M2)$
H2	H2-0: $RF(M1) = RF(M2)$	H2-1: $RF(M1) \neq RF(M2)$
H3	H3-0: $RC(M1) = RC(M2)$	H3-1: $RC(M1) \neq RC(M2)$

Fonte: elaborado pelo autor.

Unidades de Experimento: Uma unidade de experimento é uma amostra que será utilizada sob a qual o “tratamento” do experimento (conjunto de fatores) é aplicado para se obter as variáveis resposta (dependentes) citadas anteriormente. É através desta unidade que será possível obter variação estatística na análise dos resultados da investigação.

No nosso cenário, a unidade experimental é bastante específica. Assim como no experimento anterior, a unidade experimental é composta por redações + formulário de avaliação. No caso, foram criadas mais três redações por um professor especialista e inseridas no ambiente. A unidade experimental da pesquisa agora são quatro redações

(apresentadas no Apêndice D), pré-definidas pelo professor e um determinado formulário de avaliação (o mesmo criado anteriormente, definido no Apêndice E).

O ambiente MeuTutor ENEM com essas redações cadastradas será apresentado aos alunos de ensino médio para que estes possam participar do experimento. Foram selecionados os alunos do ensino médio (das turmas do primeiro, segundo e terceiro anos) da escola “Atual Cursos”, localizada no Tabuleiro dos Martins, na Rua Fontam de Melo, número 3, a qual se disponibilizou a participar. A escola contém em torno de 100 alunos nas turmas citadas.

Projeto Experimental: Assim como no experimento anterior, o experimento apresentado é do tipo comparativo e, portanto, o projeto experimental escolhido deve pertencer à categoria de projetos comparativos. No caso desse experimento em específico, como temos apenas um único fator variante (modelo utilizado) com apenas 2 níveis, podemos utilizar o mesmo design do experimento anterior (fatorial 2^k sem repetição), tendo em vista que cada repetição do experimento ainda tem um custo relativamente alto além de exigir bastante tempo para preparação do ambiente, alunos e demais esforços necessários para sua conclusão.

Com um projeto fatorial sem repetição e um único fator com 2 níveis, teremos então apenas 2 ensaios possíveis, sendo executados cada um apenas uma vez, totalizando dois tratamentos possíveis e conseqüentemente duas execuções. A Tabela 22 descreve cada um dos dois tratamentos.

Tabela 22 – Definição dos tratamentos do experimento de gamificação

Número do tratamento	Referência	Modelo utilizado	Descrição
1	T1	M1	Sem gamificação
2	T2	M2	Com gamificação

Fonte: elaborado pelo autor.

Plano de execução: Esta seção descreve como será a execução do experimento que foi planejado e descrito ao longo dessa seção. A execução do experimento envolve os seguintes passos:

1. **Preparação do ambiente educacional Meututor ENEM:** Da mesma forma que o experimento anterior, inicialmente, é necessário preparar o ambiente educacional Meututor tanto com relação a nova unidade de experimento definida, isto é, cadastrando as novas redações quanto em suas funcionalidades. Por possuir agora dois níveis de fatores que influenciam na execução do ambiente (com e sem gamificação), o ambiente deve ser configurado para funcionar das duas formas. Nesse

sentido, duas versões do ambiente deverão ser implantadas em um servidor da amazon EC2 e disponibilizadas na internet¹.

2. **Capacitação:** Uma conversa inicial com os alunos será feita para apresentar o ambiente, explicando as funcionalidades e o processo de avaliação. Na metade do prazo definido para o experimento, uma nova conversa deverá ser feita para saber as dificuldades e retirar dúvidas dos participantes.
3. **Seleção dos usuários que irão participar do experimento:** Este passo consiste na seleção de alunos que irão participar do experimento. Nesse caso serão selecionados todos os alunos que desejaram participar, após a apresentação anterior.
4. **Agrupamento dos usuários:** Como temos apenas duas versões do sistema em avaliação, precisamos de um grupo de controle para utilizar a versão sem gamificação (GC) e um grupo de teste (GT) para utilizar a versão com gamificação. O agrupamento dos usuários será da seguinte forma: dos usuários que desejaram participar na etapa anterior, estes serão agrupados de acordo com sua turma (primeiro, segundo ou terceiro ano). De cada grupo desse e de forma aleatória, metade deles pertencerá ao grupo GC e a outra metade ao grupo GT.
5. **Execução do experimento:** Os grupos de alunos serão instruídos e poderão executar seus tratamentos propostos ao longo de um prazo de duas semanas (cerca de 15 dias). Os alunos deverão fazer e corrigir as redações a livre escolha. Não haverá prazo limite para realização de cada redação, porém elas apresentam uma ordem específica, isto é, alunos não podem fazer a última redação sem ter feito a anterior, contudo podem, a qualquer momento fazer a primeira redação.
6. **Coleta de dados:** Os dados serão coletados após a execução do passo anterior do experimento. Os dados essenciais devem ser obtidos diretamente do ambiente, sem a interferência dos alunos, a fim de evitar problemas.
7. **Análise dos dados:** No final, assim como no experimento anterior, deverá ser feita uma análise estatística dos dados. Os resultados de cada métrica obtidos na execução serão sumarizados na ferramenta R e organizados para exibição gráfica através de histogramas e/ou boxplot, com o propósito de facilitar a comparação. Os mesmos testes estatísticos serão utilizados com o mesmo nível de confiança (95%).

Após a execução dos passos anteriores, os resultados poderão ser obtidos e comparados entre si. A análise dos dados bem como as principais conclusões obtidas com a execução desse experimento estão apresentadas nas Seções 6.3 e 6.4, respectivamente.

¹ Disponibilizadas através do endereço <http://grupo1.atualcursos.meututor.com.br> e <http://grupo2.atualcursos.meututor.com.br>

6.2 Execução do experimento

Essa seção tem como objetivo descrever, de forma resumida, o processo de execução do experimento, o qual deve seguir as etapas definidas no **Plano de Execução** da Seção 6.1.1. A plataforma foi configurada, nas suas duas versões (o que exigiu habilidades de programação para a retirada das funcionalidades de gamificação do modelo), em um único servidor virtual da amazon EC2(para reduzir os custos), com a seguinte configuração:

- Instância: m3.large
- Processamento: vCPU: 2/ECU: 6.5
- Memória (GiB): 7.5
- Armazenamento da instância (GB): 1 x 32 SSD
- Sistema operacional: Ubuntu Server 14.04 LTS (HVM)

Podemos notar que essa configuração é muito maior do que a configuração escolhida para o experimento anterior. Isso deve-se que agora precisaremos executar duas versões diferentes (com e sem gamificação) do sistema, quanto que no experimento anterior apenas uma versão do sistema foi executada. Os software utilizados foram os mesmos do experimento anterior, seguindo as mesmas configurações de utilização.

Vale ressaltar que duas versões precisam serem configuradas e portanto mais portas serão necessárias para o banco, os serviços e uma configuração de redirecionamento deverá ser incluída no apache para que suporte dois endereços de acesso (grupo1 e grupo2).

Após a configuração do ambiente MeuTutor, uma capacitação foi realizada no dia 20/07/2015 no colégio atual cursos, onde o ambiente foi apresentado às diretoras, coordenadores, professores e a todos os alunos das turmas do ensino médio (em separado) que estavam presentes no momento. Os quase 100 alunos que estavam presentes se interessaram e se disponibilizaram a participar. De forma aleatória e para cada turma, foi distribuído a cada aluno um grupo (1 ou 2), representando GC e GT.

Os alunos do grupo de controle (GC) executaram o tratamento T1 (sem gamificação) enquanto que os alunos do grupo de teste (GT) executaram o tratamento T2 (com gamificação). Vale ressaltar que ambos utilizavam o modelo de avaliação por pares proposto e se diferenciavam apenas na presença ou não dos elementos de gamificação.

O acesso ao ambiente varia de acordo com o grupo escolhido. Todas as intruções de acesso, configuração e os processos envolvidos foram dadas aos alunos. Tendo como base a data dessa capacitação, a data final do experimento foi definida em 03/08/2015 (15 dias depois, contando com o próprio dia inicial). Os alunos tiveram a opção de cadastrar na plataforma e fazer as atividades a livre escolha. Todos as suas interações foram armazenadas no banco de dados, obtidas diretamente pelo sistema. A análise desses dados será apresentada na seção a seguir.

6.3 Análise dos dados

Após a execução do experimento seguindo o planejamento apresentado, ocorrido como descrito de forma resumida na Seção 6.2, é preciso realizar a análise dos dados obtidos, uma vez que eles já foram coletados. Essa análise, assim como no experimento anterior, é feita utilizando a ferramenta de estatística R com os mesmos scripts que já foram criados.

Seguindo o mesmo modelo anterior, inicialmente será apresentada uma rápida análise preliminar dos dados na Seção 6.3.1. Em seguida, a Seção 6.3.2 apresenta uma análise descritiva dos dados obtidos na execução do experimento de forma gráfica, através de histogramas para cada métrica em cada tratamento executado, diagramas de caixa comparativos (boxplots) e no final uma breve sumarização de algumas estatísticas analisadas sobre o conjunto de dados. Em seguida, após essa análise, serão apresentadas na Seção 6.3.3 as verificações das hipóteses de pesquisa, que será feita com a análise dos intervalos de confiança e execução de testes de hipóteses, quando necessário.

6.3.1 Análise preliminar dos dados

Esta seção tem como objetivo apresentar de forma resumida alguns dos principais resultados obtidos no experimento, em uma comparação direta entre a aplicação do modelo sem gamificação (no caso do tratamento T1) com a aplicação do modelo com gamificação (no caso do tratamento T2). Os dados considerados importantes e observados estão apresentados na Tabela 23.

Tabela 23 – Sumarização dos dados relativos a variável Acesso (A).

Dado observado	T1	T2	Diferença	% Dif.
Número de cadastros	17	19	+2	+11.76%
Total Redações Aberto	4	4	0	0%
Média notas	669.33	563.92	-105.41	-15.75%
Número logins	42	69	+27	+64.28%
Total Redações Feitas	19	21	+2	+10.53%
Total Redações Corrigidas	45	54	+9	+20.0%

Fonte: elaborado pelo autor.

Podemos observar na tabela que houve 2 cadastros a mais no grupo com tratamento T2 (11.76%). O número de redações em aberto foi 4, independente do tratamento, isso pode ter sido ocasionado pelo curto prazo para realização e correção das redações, implicando que alguns alunos deixaram de corrigir o suficiente para que todas as redações pudesse ter notas. Por possuir mais redações feitas e devido a essa quantidade ser igual, podemos concluir que a proporção de diferença entre notas é maior na gamificação. Isso pode explicar a queda das notas, que atingiu cerca de 15.75% menor no tratamento com gamificação.

Por outro lado analisando as métricas definidas no experimento, podemos observar que a quantidade de acesso (logins) do tratamento com gamificação (T2) foi 64.28% maior do que sem gamificação, o que nos indica uma forte influência da gamificação no acesso dos alunos na plataforma. O número de redações feitas foi maior também no tratamento T2, com duas redações realizadas a mais que o tratamento T1. Apesar do número pequeno, isso representa cerca de 10.53% de redações a mais. O mesmo ocorre com relação a métrica de redações corrigidas. No tratamento T2 temos 9 correções a mais do que no tratamento T1, o que representa um aumento de cerca de 20%.

A Tabela 24, por sua vez, apresenta a quantidade de alunos que fizeram e/ou corrigiram as redações, detalhando os dados de redações feitas e corrigidas da tabela anterior.

Tabela 24 – Detalhamento da quantidade de alunos que fizeram/corrigiram x redações

Quantidade	T1		T2	
	Redações Feitas	Redações Corrigidas	Redações Feitas	Redações Corrigidas
0	5	7	5	4
1	7	1	10	1
2	3	2	1	5
3	2	3	3	2
4	-	0	-	3
5	-	0	-	2
6	-	0	-	0
7	-	2	-	1
8	-	1	-	1
9	-	1	-	0
TOTAL	17	17	19	19

Fonte: elaborado pelo autor.

Podemos observar que a quantidade de dados em cada tratamento é igual a quantidade de cadastros em cada um deles, tendo em vista que a tabela mostra a quantidade de alunos que fizeram/corrigiram uma certa quantidade de redações, incluindo aqueles que não fizeram nenhuma. Em ambos os tratamentos tivemos 5 alunos que não fizeram nenhuma redação.

Podemos observar na tabela que o número de alunos que fizeram somente a primeira redação é maior no tratamento T2 do que no tratamento T1. Outro ponto interessante que podemos observar na tabela é que dos 4 alunos que fizeram a segunda redação no tratamento T2, 3 deles prosseguiram e fizeram a terceira redação, o que corresponde a 75%. Isso pode ter sido ocasionado pelo fato de que existe um troféu que exige que os participantes façam três redações, o que deve ter motivado os alunos a continuarem no processo. Esse mesmo número cai para 40% nos dados do tratamento T1, onde tínhamos 5 alunos que fizeram a segunda redação, mas apenas 2 continuaram para a terceira.

6.3.2 Análise descritiva dos dados

O experimento apresenta três variáveis resposta sendo:

- **Acesso (A):** Representando a quantidade de logins dos alunos.
- **Redações feitas(RF):** Corresponde a quantidade de redações feitas.
- **Redações corrigidas(RC):** Indicando a quantidade de redações corrigidas.

Esta seção apresentará uma análise de cada uma dessas variáveis, iniciando pela variável de acesso (A). A Figura 61 apresenta o diagrama de caixa para essa métrica com relação ao tratamento T1 (sem gamificação) e T2 (com gamificação).

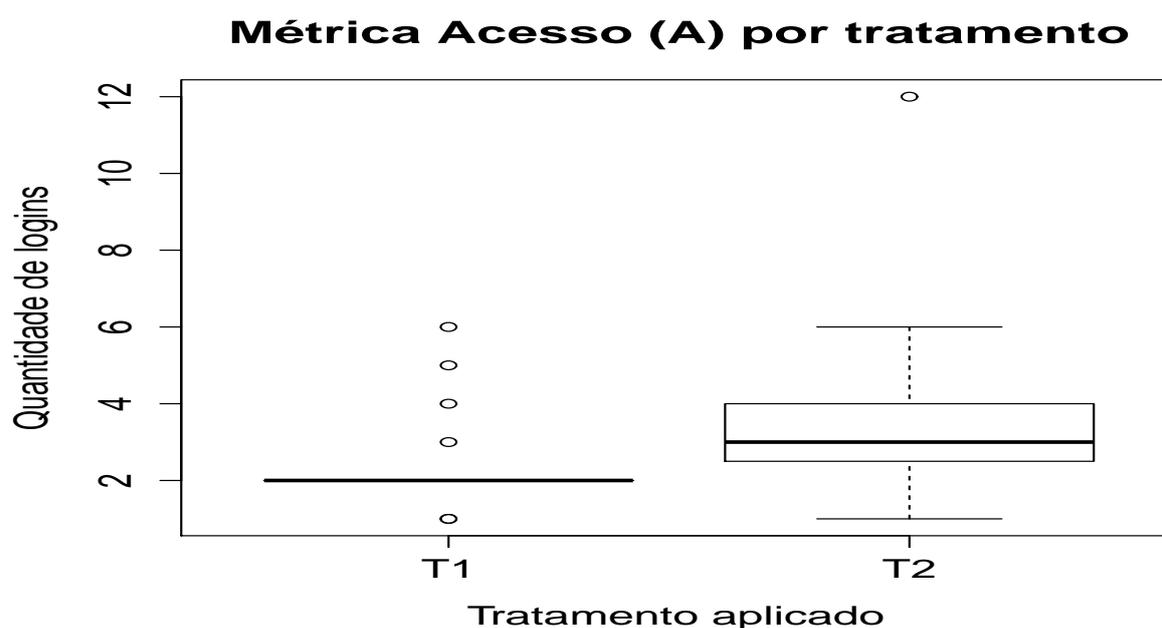


Figura 61 – Diagrama de caixa com comparativo da métrica Acesso para os tratamentos T1 e T2

Fonte: elaborado pelo autor

Podemos ver na figura que a média (média (T1)=2.471 contra média(T2)=3.632) e mediana (mediana(T1) = 2 contra mediana(T2)=3), apresentam certas diferenças de comportamento. Analisando apenas essas medidas, pode-se notar que há uma tendência de que a quantidade de acesso (logins) do tratamento T2 seja maior que a quantidade de logins do tratamento T1. A figura sugere que os dados dessa métrica obtidas com a aplicação desses tratamentos possuem variações estatísticas significativas, o que nos indica uma certa disparidade nos tratamentos.

Também podemos observar que a quantidade de *outliers* (pontos fora do padrão) é muito maior no tratamento T1 do que no tratamento T2, o que nos indica que poucos estudantes acessaram mais do que duas vezes (valor mediana) no tratamento T1, enquanto que isso ocorreu bem mais vezes no tratamento T2 (tendo em vista à dispersão apresentada e a mediana(T2) = 3 > mediana(T1) = 2).

No entanto, ainda não foram geradas evidências estatísticas para afirmar essas conclusões. Tais afirmações só poderão ser feitas quando os testes estatísticos forem realizados (especificamente na subseção 6.3.3).

Além da análise do diagrama de caixa, podemos analisar os dados do ponto de vista de suas frequências de aparição através da criação de histogramas. As Figuras 62 e 63 apresentam, respectivamente, os histogramas dos dados nos tratamentos T1 e T2 para a métrica acesso.

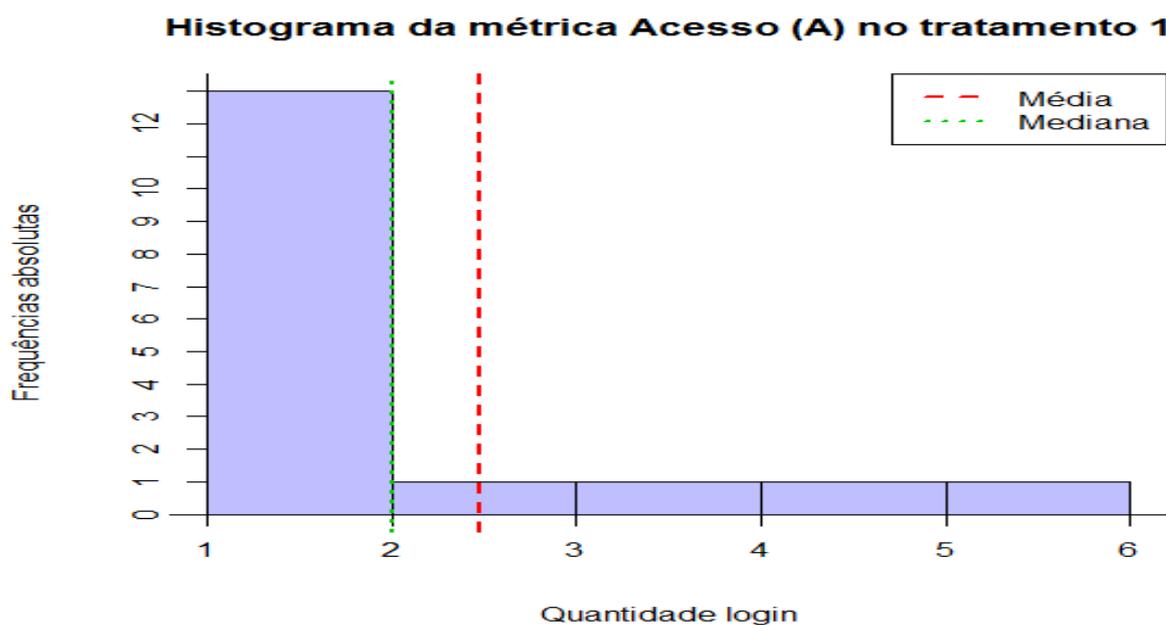


Figura 62 – Histograma para a métrica acesso (A) no tratamento T1

Fonte: elaborado pelo autor

Podemos observar nestes dois histogramas que a distribuição dos dados para o tratamento T1 (sem gamificação) não apresenta características de uma distribuição normal, devido à grande variação existente entre média e mediana e por possuir uma cauda muito longa a esquerda. Analisando os dados para o tratamento T2, se desconsideramos o outlier, a distribuição apresenta aspectos de normalidade, apesar da variação entre a média e mediana ainda ser alta.

Contudo, assim como explanado no experimento anterior, só com a utilização de métodos estatísticos será possível afirmar se as distribuições desses dados são normais ou não. Vale ressaltar que a verificação de normalidade dos dados é pré-requisito para escolha de qual teste de hipótese estatístico será utilizado para analisar as hipóteses.

Podemos ainda observar no histograma do tratamento T1 que a maior frequência absoluta ocorreu com valores de até dois acessos, restando apenas uma única frequência para valores de acesso maiores do que dois. Por outro lado, no histograma do tratamento

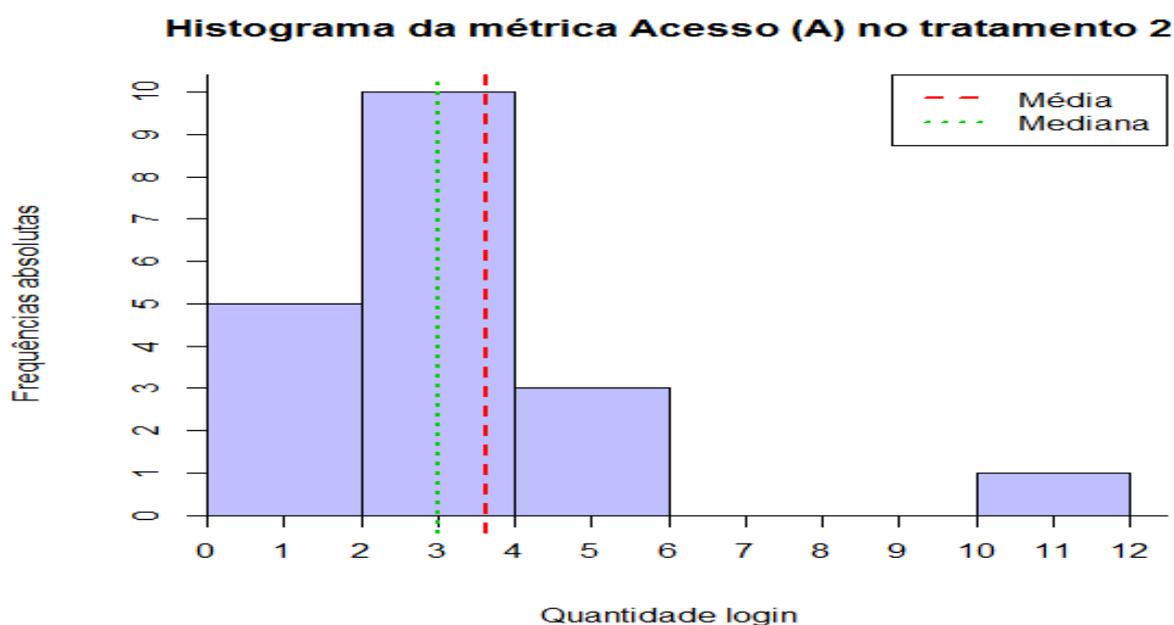


Figura 63 – Histograma para a métrica acesso (A) no tratamento T2

Fonte: elaborado pelo autor

T2, essa maior frequência ocorreu com valores a partir de dois até quatro acessos, o que a priori nos indica que a frequência de acessos em T2 é maior que em T1.

Também podemos analisar esses histogramas de forma combinada, na tentativa de encontrar possíveis semelhanças entre os dados. A Figura 64 apresenta os histogramas dos tratamentos T1 (apresentado na cor azul) e T2 (apresentado na cor vermelha).

Através da figura, podemos observar que as tendências dos dados sejam maiores no tratamento T2 se comparados ao tratamento T1. Podemos também observar que as linhas de média e mediana também são maiores no tratamento T2. Em resumo, os dados apresentados nos gráficos para a métrica acesso (A) obtidos com a execução dos tratamentos (T1 e T2) estão sumarizados na Tabela 25, para uma melhor visualização.

Tabela 25 – Sumarização dos dados relativos a variável Acesso (A).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	1.000	2.000	2.000	2.471	2.000	6.000
T2	1.000	2.500	3.000	3.632	4.000	12.000

Fonte: elaborado pelo autor.

A próxima métrica a ser analisada é a quantidade de redações feitas (RF). A análise será feita de forma semelhante a análise já apresentada nas métricas anteriores. Em primeiro momento será apresentado o diagrama de caixa comparativo com os dados obtidos

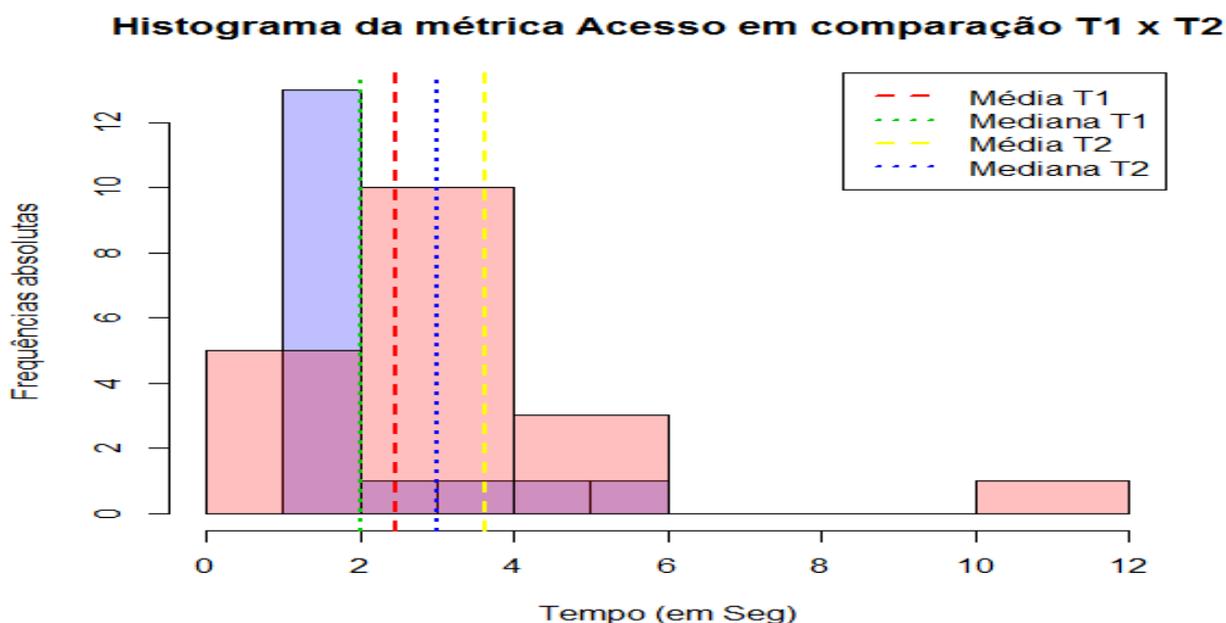


Figura 64 – Histograma para a métrica acesso(A) na combinação T1 (azul) x T2 (vermelho)

Fonte: elaborado pelo autor

em relação à quantidade de redações feitas (RF) e, em seguida, alguns histogramas com os dados de cada tratamento em específico.

A Figura 65 apresenta um conjunto de diagramas de caixa para a métrica de redações feitas (RF) com relação aos tratamentos T1 e T2.

Ao se analisar os dados de média e mediana dos tratamentos ($média(T1) = 1.118$ contra $média(T2) = 1.105$ e $mediana(T1) = mediana(T2) = 1.000$), observamos que os dados apresentam características praticamente idênticas, o que nos indica que não há uma variação estatística entre essas observações. Contudo podemos observar uma maior variação/dispersão nos dados do tratamento T1 enquanto que os dados do tratamento T2 apresentam alguns *outliers*. Devido à presença desses comportamentos, e como os diagramas se interceptam, não podemos concluir tais semelhanças sem a utilização dos devidos testes estatísticos.

Também foram criados histogramas com os dados referentes à métrica Redações Feitas, para uma análise mais detalhada deles. As Figuras 66 e 67 apresentam, respectivamente, os histogramas dessa métrica referentes aos dados obtidos na execução dos tratamentos T1 e T2.

Podemos observar nas figuras de histograma da métrica Redações Feitas (RF) que ambas as distribuições apresentam comportamentos praticamente idênticos, variando apenas em algumas frequências absolutas. Esse tipo de comportamento indica ainda mais uma forte semelhança entre os dados de ambas as distribuições, indicando que não houve variação estatística significativa com relação à essa métrica.

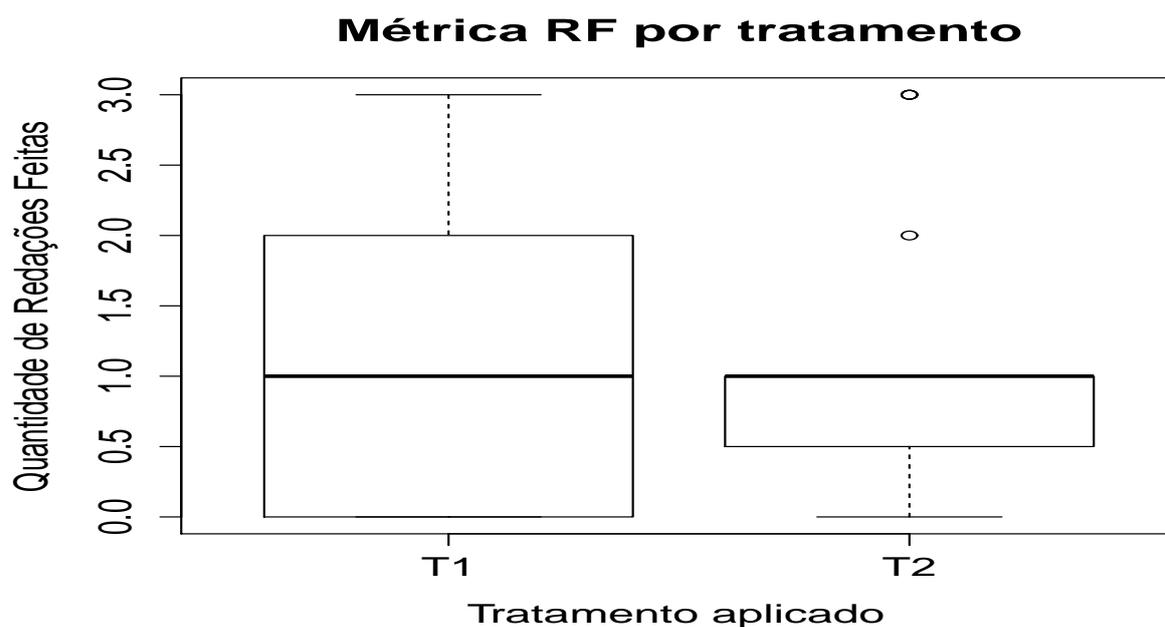


Figura 65 – Diagrama de caixa com comparativo da métrica RF para os tratamentos T1 e T2

Fonte: elaborado pelo autor

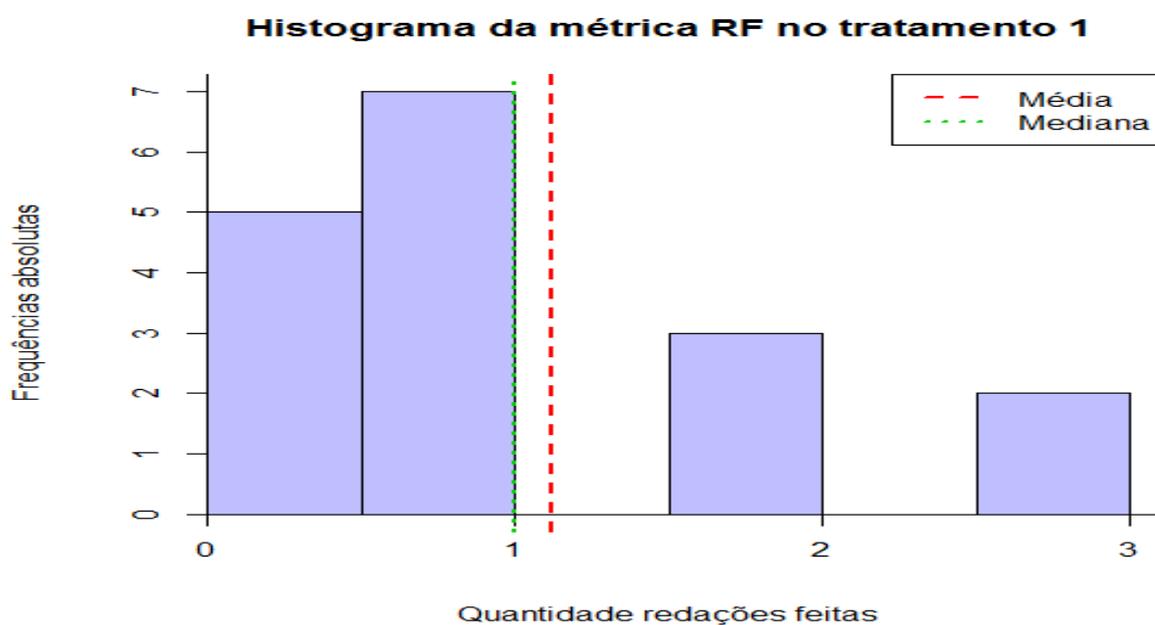


Figura 66 – Histograma para a métrica Redação Feita (RF) no tratamento T1

Fonte: elaborado pelo autor

Em ambos os casos, não há como especificar visualmente a semelhança com nenhuma distribuição conhecida, o que pode nos indicar que ambas não são distribuições normais. Contudo, em ambos os casos são necessárias a execução de testes de normalidade nos

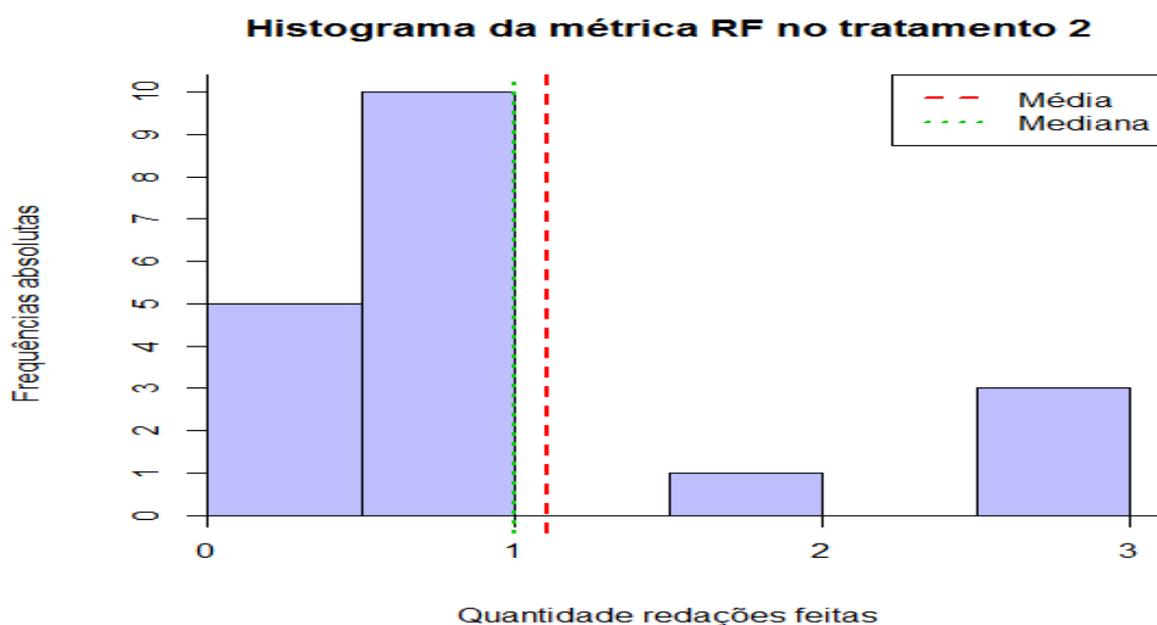


Figura 67 – Histograma para a métrica Redação Feita (RF) no tratamento T2

Fonte: elaborado pelo autor

dados antes da execução dos testes estatísticos, para confirmar se os dados apresentam ou não uma distribuição normal.

Para visualizar melhor a semelhança entre as distribuições, a Figura 68 apresenta os dois histogramas combinados, onde T1 é apresentado na cor azul e T2 na cor vermelha.

Podemos observar que os dois praticamente se sobrepõem e as linhas de média e mediana estão praticamente sobreescritas. Em resumo, os dados apresentados nos gráficos para a métrica Redações Feitas (RF) obtidos com a execução dos tratamentos (T1 e T2) estão sumarizados na Tabela 26, para uma melhor visualização.

Tabela 26 – Sumarização dos dados relativos a variável Redações Feitas (RF).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	0.000	0.000	1.000	1.118	2.000	3.000
T2	0.000	0.500	1.000	1.105	1.000	3.000

Fonte: elaborado pelo autor.

A próxima métrica a ser analisada é a quantidade de redações corrigidas (RC). A Figura 69 apresenta o diagrama de caixa para a métrica de redações corrigidas (RC) com relação aos tratamentos executados.

Podemos observar que as médias e medianas nos tratamentos (média(T1) = 2.647 contra média(T2) = 2.842 e mediana(T1) = mediana(T2) = 2.000), assim como na métrica

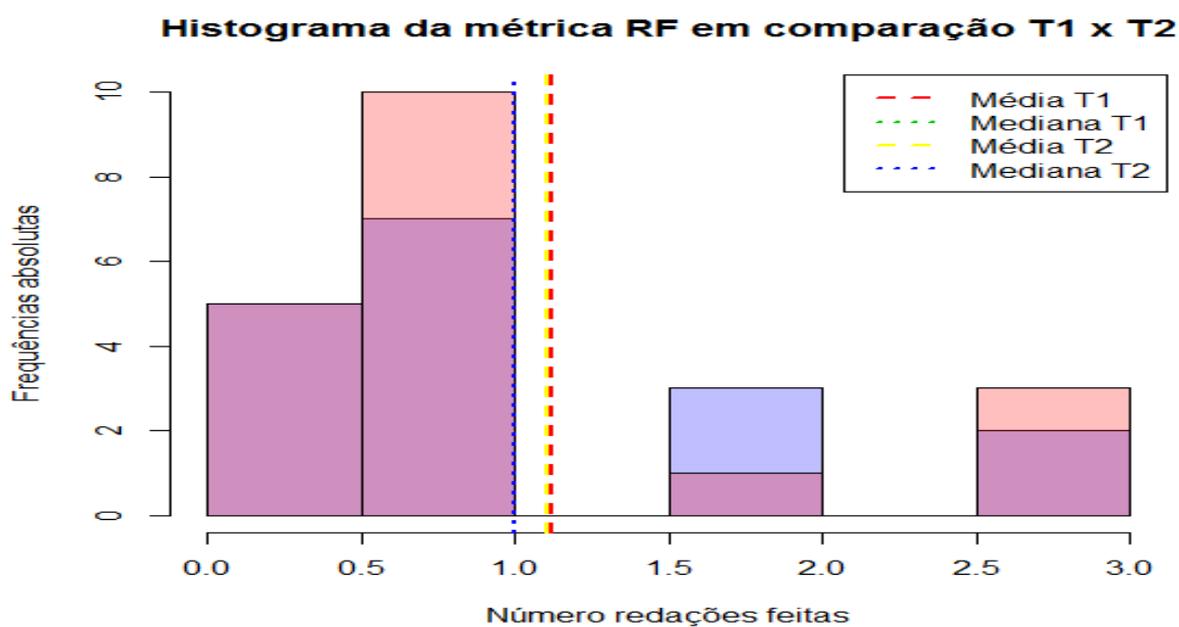


Figura 68 – Histograma para a métrica Redação Feita (RF) na combinação T1 (azul) x T2 (vermelho)

Fonte: elaborado pelo autor

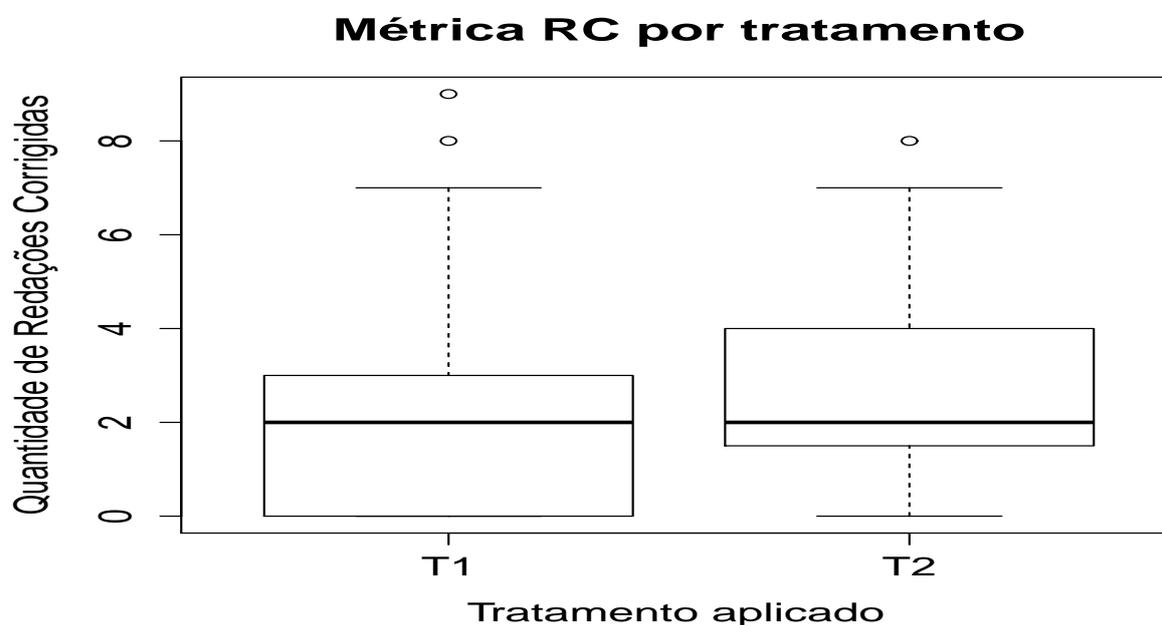


Figura 69 – Diagrama de caixa com comparativo da métrica Redação Corrigida (RC) para os tratamentos T1 e T2

Fonte: elaborado pelo autor

anterior de redações feitas (RF) são praticamente idênticas, indicando que não há uma

variação estatística entre essas observações.

Observando a dispersão dos dados no tratamento T1 para abaixo da mediana e nos dados T2 para acima da mediana, podemos ver uma tendência de um maior número de correções no tratamento T2 se comparado ao tratamento T1, contudo, devido à sobreposição, não podemos concluir nada sem a execução dos testes estatísticos.

Além da análise do diagrama de caixa para a métrica Redações Corrigidas (RC), podemos analisar estes dados do ponto de vista de suas frequências de aparição com os histogramas. As Figuras 70 e 71, mostram, respectivamente os histogramas da métrica custo correspondentes aos dados da execução dos tratamentos T1 e T2.

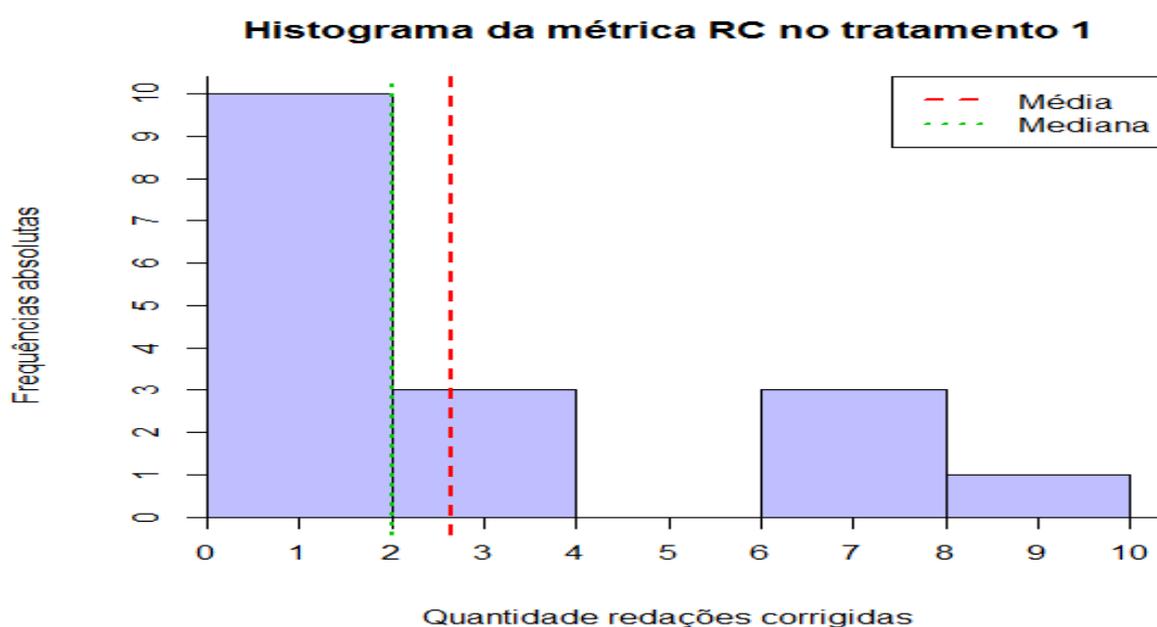


Figura 70 – Histograma para a métrica Redação Corrigida (RC) no tratamento T1

Fonte: elaborado pelo autor

Podemos observar nos dados do histograma do tratamento T1 que não há nenhuma semelhança dos dados com alguma distribuição conhecida. Estes mesmos dados apresentam uma cauda longa a esquerda, com uma alta frequência absoluta (10). Podemos notar que essa frequência vai caindo a medida que o número de correções vai crescendo de forma brusca. Por outro lado, os dados da distribuição em T2 apresentam características próximas a normalidade, mesmo possuindo cauda a esquerda e uma grande variação entre a média e a mediana. Podemos observar que a queda da frequência não ocorre de forma tão brusca quanto a queda nos dados do tratamento T1.

Podemos observar que as frequências absolutas para valores de correção não seguem um padrão de qual tratamento é maior, o que indica uma certa semelhança entre as distribuições. Porém, mais uma vez, testes de análise da amostra devem ser feitos para

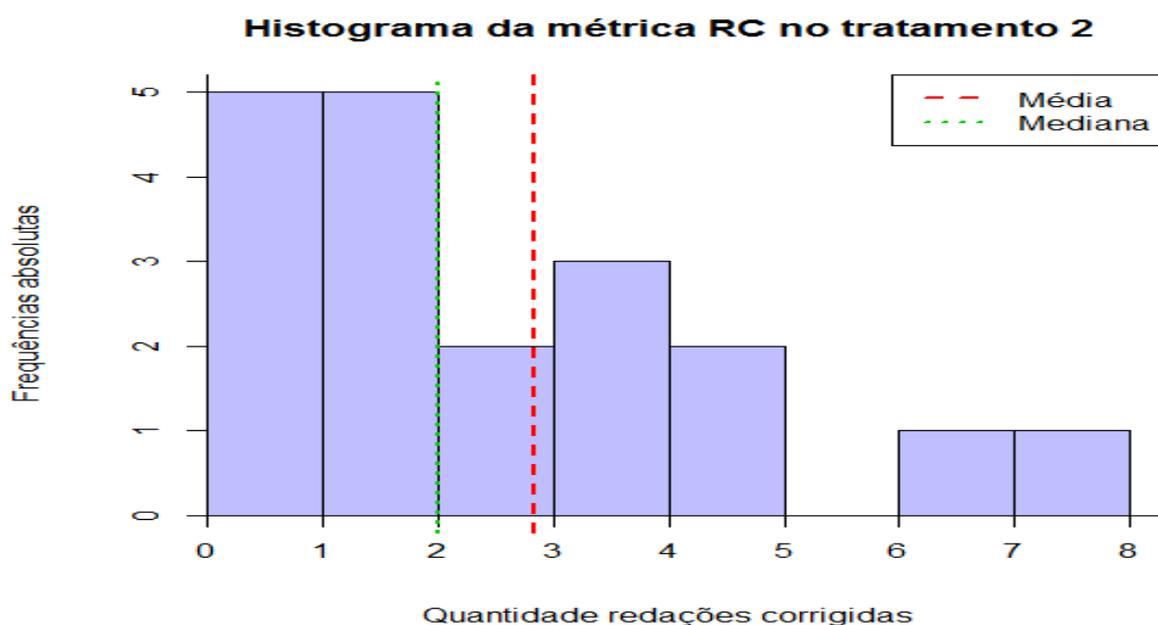


Figura 71 – Histograma para a métrica Redação Corrigida (RC) no tratamento T2

Fonte: elaborado pelo autor

validar a normalidade ou não das distribuições apresentadas nos dados antes de qualquer teste estatístico ser realizado.

Finalmente, de forma resumida, os dados apresentados nos gráficos para a métrica Redações Corrigidas (RC) obtidos com a execução dos tratamentos (T1 e T2) estão sumarizados na Tabela 27 para uma melhor visualização.

Tabela 27 – Sumarização dos dados relativos a variável Redações Corrigidas (RC).

Tratamento	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
T1	0.000	0.000	2.000	2.647	3.000	9.000
T2	0.000	1.500	2.000	2.842	4.000	8.000

Fonte: elaborado pelo autor.

6.3.3 Verificação das hipóteses

Assim como no experimento anterior, após uma análise descritiva dos dados, é necessário a realização de uma análise mais aprofundada com o objetivo de verificar a validade das hipóteses de pesquisa, descritas na Seção 6.1.1.

Para dar continuidade aos testes de verificação de hipótese, vale lembrar que estes serão feitos através da análise de intervalos de confiança e testes estatísticos. As condições serão iguais às do experimento anterior, isto é, o nível de confiança escolhido em nossas

análises deste experimento também é de 95%, ou seja, considerando um α (nível de significância) de 5% ou 0,05. Para analisar a normalidade dos dados usaremos o teste de Shapiro-Wilk (SHAPIRO; FRANCIA, 1972). Para casos de distribuições normais, será utilizado o T-Test (WELCH, 1938). Para distribuições não normais, será utilizado o teste de Wilcoxon (WILCOXON, 1945).

A primeira verificação de hipótese será feita com relação a métrica de acesso (A). Conforme a Seção 6.1.1, as hipóteses nula e alternativa são, respectivamente:

H1-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de acessos dos estudantes na plataforma.

H1-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de acessos dos estudantes na plataforma.

A primeira análise a ser feita é observando os intervalos de confiança dos tratamentos executados. A Figura 72 apresenta esses intervalos de confiança para os tratamentos T1 e T2 com 5% de significância.

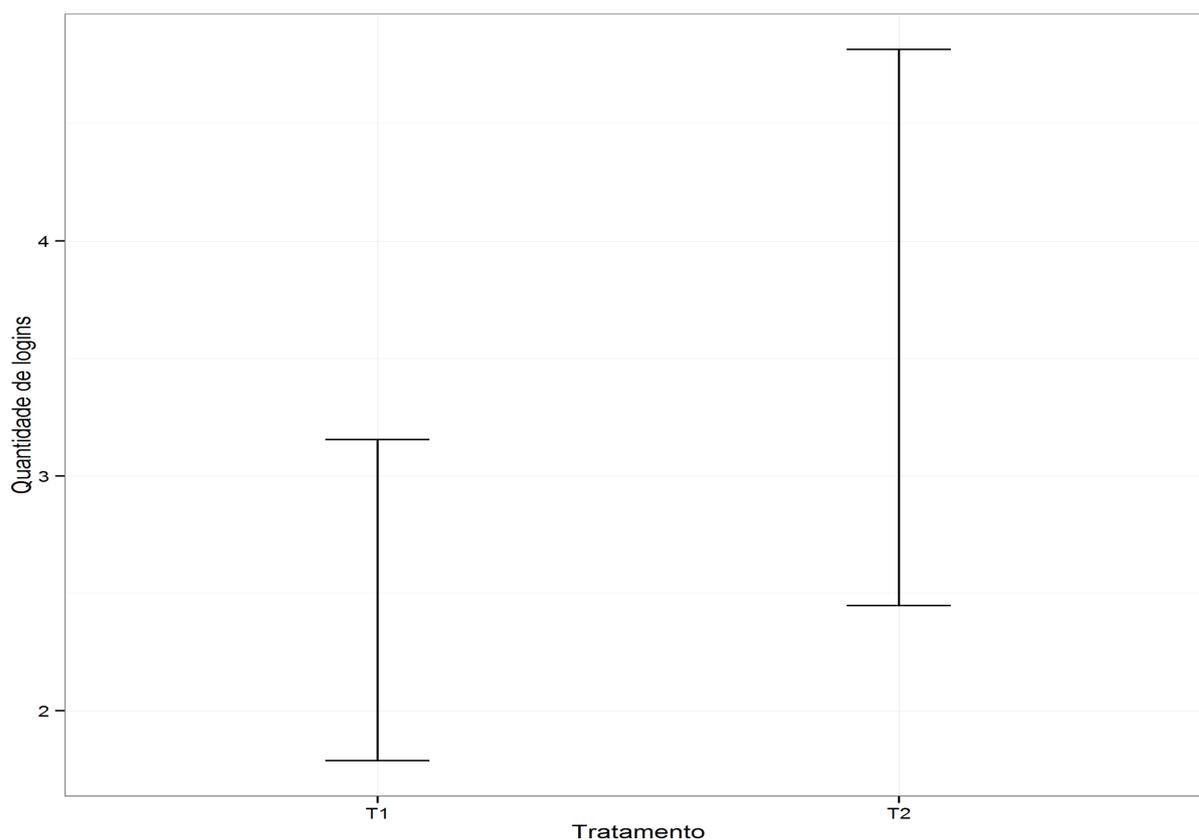


Figura 72 – Intervalo de confiança da métrica acesso (A) para os tratamentos T1 e T2

Fonte: elaborado pelo autor

A Figura apresenta os intervalos de confiança de cada tratamento comparando sob o efeito da métrica acesso com 5% de significância. Podemos observar que o intervalo de confiança do tratamento T2 se apresenta superior em grande parte se comparado ao

intervalo de confiança do tratamento T1. Contudo existe um intervalo de sobreposição. Dada esta sobreposição, ainda não há como fazer qualquer afirmação sobre essa métrica. Dessa forma, há uma necessidade real de verificar as hipóteses por meio de um teste de hipótese comparativo.

Nesse caso, em primeiro momento, precisamos aplicar o teste de Shapiro-Wilk para descobrir a normalidade ou não dos dados de cada tratamento. Os resultados da aplicação do teste de Shapiro-Wilk quando executado sobre os dados da métrica acesso nos tratamentos estão apresentados na tabela 28.

Tabela 28 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Acesso

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.71721	0.892	0.0001823	0.05
T2	0.74726	0.901	0.0002154	0.05

Fonte: elaborado pelo autor.

Ao analisar esses dados, pode-se notar que os valores de $W_{calculado}(T1) = 0.71721 < W_{\alpha}(T1) = 0.892$ e $P_{valor}(T1) = 0.0001823 < \alpha = 0.05$. Portanto, é possível refutar a hipótese nula do teste de Shapiro e, conseqüentemente, podemos afirmar com nível de significância de 5% que a amostra T1 não provém de uma população normal.

Da mesma forma, se analisarmos os dados do tratamento T2, temos que $W_{calculado}(T2) = 0.74726 < W_{\alpha}(T2) = 0.901$ e $P_{valor}(T2) = 0.0002154 < \alpha = 0.05$. Assim como acontece no tratamento T1, a hipótese nula do teste de Shapiro pode ser refutada e a distribuição dos dados de T2 também é não normal.

Dessa forma, como as distribuições dos dados dos dois tratamentos não são distribuições normais, podemos aplicar o teste de Wilcoxon, comparando os tratamentos. Os dados da execução do teste de Wilcoxon estão apresentados na Tabela 29.

Tabela 29 – Resultado da aplicação dos testes estatísticos com os dados da métrica Acesso

Tratamento	Teste Utilizado	v/w	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	96	0.03209

Fonte: elaborado pelo autor.

Podemos observar que o valor de $p\text{-value}(T1 \times T2) = 0.03209$, obtido com a execução do teste de Wilcoxon é menor que $\alpha = 0.05$. Portanto, com 95% de confiança, estatisticamente, pode-se refutar a hipótese nula, afirmando que os valores da quantidade de acesso obtidos entre os tratamentos possuem diferenças significativas entre si, isto é, há evidências estatísticas que mostrem a não equivalência dos dados.

O fato desse coeficiente ser menor que o α gera evidência estatística suficiente para refutar a hipótese nula $H1-0$ (onde os acessos seriam iguais) e aceitar a hipótese alternativa

H1-1, implicando que **O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de acessos dos estudantes na plataforma.**

A próxima verificação de hipótese será feita com relação a métrica Redações Feitas (RF). Conforme a Seção 6.1.1, as hipóteses nula e alternativa são, respectivamente:

H2-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações realizadas na plataforma.

H2-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de redações realizadas na plataforma.

De forma semelhante a métrica de acesso, será apresentado o gráfico com os intervalos de confiança dos tratamentos T1 e T2 para a métrica Redações Feitas (RF). A Figura 73 ilustra os intervalos de confiança citados com um valor de significância igual a 5%.

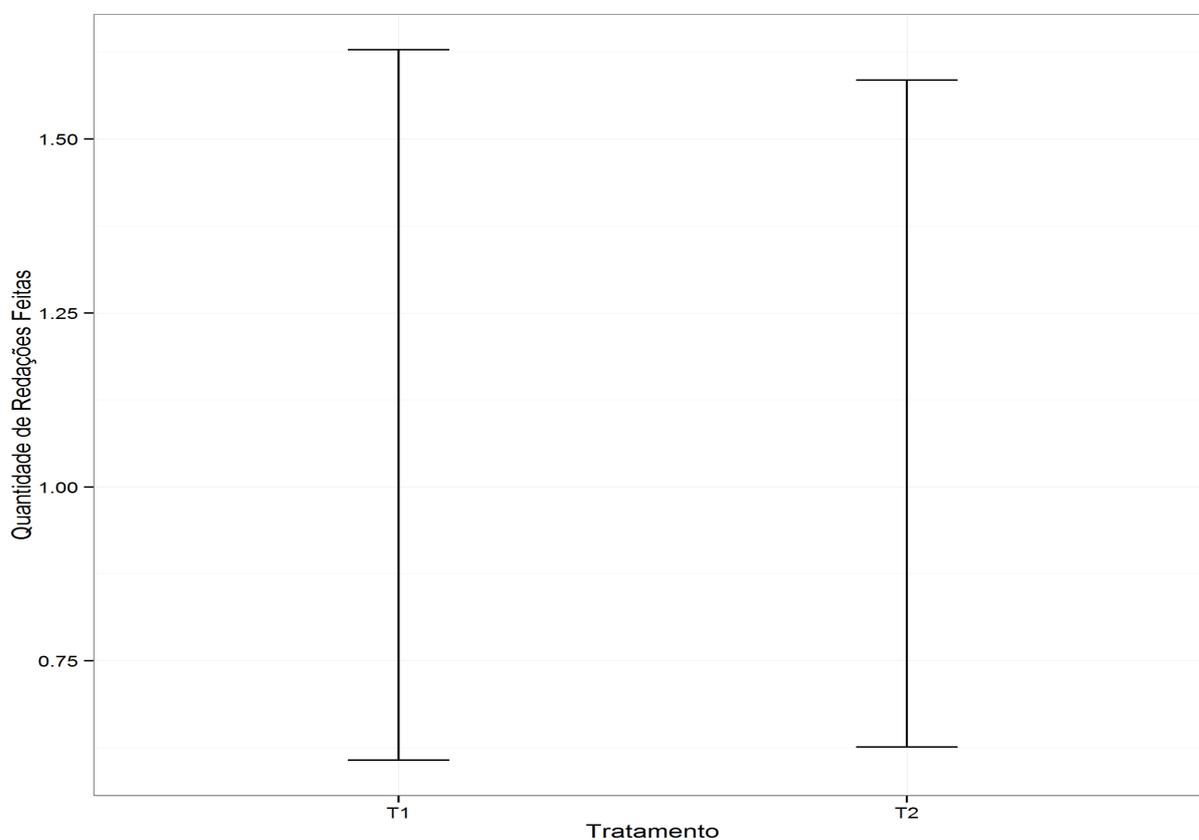


Figura 73 – Intervalo de confiança da métrica Redações Feitas (RF) para os tratamentos T1 e T2

Fonte: elaborado pelo autor

Podemos observar que o intervalo de confiança do tratamento T2 se sobrepõe totalmente se compararmos com o intervalo de confiança do tratamento T1. Essa sobreposição total indica fortemente que são semelhantes, porém não é possível concluir somente com observação visual. Aplicamos então o teste de Shapiro-Wilk para descobrir a normalidade ou não dos dados de cada tratamento. Os resultados da aplicação do teste estão apresentados na tabela 30.

Tabela 30 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Redações Feitas (RF)

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.86017	0.892	0.01539	0.05
T2	0.79264	0.901	0.0008953	0.05

Fonte: elaborado pelo autor.

Assim como ocorreu com a métrica anterior, tanto os dados do tratamento T1 quanto os dados do tratamento T2 são de distribuições não normais. Como vemos, $W_{calculado}(T1) = 0.86017 < W_{\alpha}(T1) = 0.892$ e $P_{valor}(T1) = 0.01539 < \alpha = 0.05$ e $W_{calculado}(T2) = 0.74726 < W_{\alpha}(T2) = 0.901$ e $P_{valor}(T2) = 0.0008953 < \alpha = 0.05$. Assim, em ambos os casos, a hipótese nula do teste de Shapiro pode ser refutada e a distribuição dos dados é não normal.

Com esse resultado de ambas as distribuições serem não normais, podemos aplicar o mesmo teste anterior (Teste de Wilcoxon) comparando os tratamentos. Os dados da execução do teste de Wilcoxon estão apresentados na Tabela 31.

Tabela 31 – Resultado da aplicação dos testes estatísticos com os dados da métrica Redação Feita (RF)

Tratamento	Teste Utilizado	v/w	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	164	0.9458

Fonte: elaborado pelo autor.

Podemos observar que o valor de $p\text{-value}(T1 \times T2) = 0.9458$ obtido com a execução do teste é maior do que $\alpha = 0.05$. Portanto, com 95% de confiança, estatisticamente, não pode se afirmar que os valores da quantidade de redações feitas obtidos entre os tratamentos possuem diferenças significativas entre si, isto é, não há evidências estatísticas que mostrem a não equivalência da quantidade de redações feitas.

O fato desse coeficiente ser maior que o α indica que os resultados dessa métrica são equivalentes (iguais), não gerando evidência estatística suficiente para refutar a hipótese nula H_{2-0} , implicando que **O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações realizadas na plataforma.**

A próxima verificação de hipótese será feita com relação a métrica Redações Corrigidas (RC). De acordo com a Seção 6.1.1, temos como hipótese nula e alternativa, respectivamente:

H3-0: O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações corrigidas na plataforma.

H3-1: O uso de técnicas de gamificação no modelo de avaliação por pares traz diferenças na quantidade de redações corrigidas na plataforma.

Vamos partir para a análise do intervalo de confiança da métrica redações corrigidas (RC) obtidos com a execução dos tratamentos T1 e T2. O gráfico está apresentado na Figura 74, com um valor de significância igual a 5%.

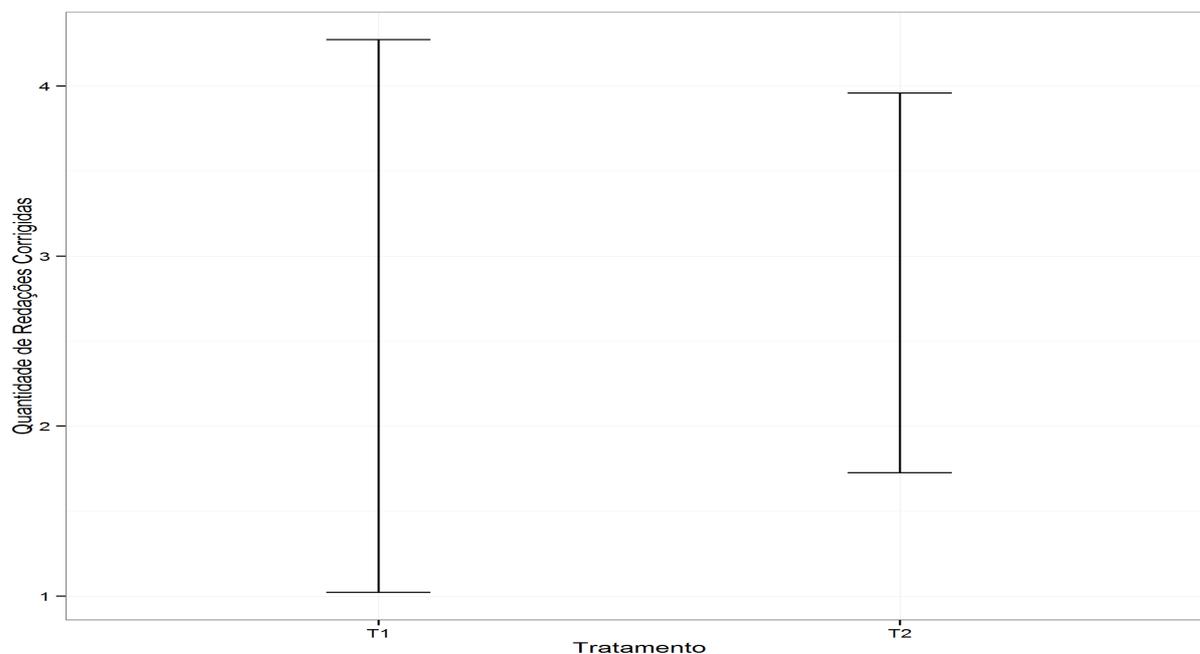


Figura 74 – Intervalo de confiança da métrica Redações Corrigidas (RC) para os tratamentos T1 e T2

Fonte: elaborado pelo autor

Da mesma forma como a métrica anterior, podemos observar que o intervalo de confiança do tratamento T2 se sobrepõe totalmente se compararmos com o intervalo de confiança do tratamento T1. Essa sobreposição indica semelhança, mas impede que consigamos tirar conclusões apenas com as observações visuais, necessitando dos testes estatísticos.

Nesse caso, aplicou-se o teste de Shapiro-Wilk com os dados desses tratamentos para analisar a normalidade dos dados e o resultado pode ser visto na Tabela 32.

Tabela 32 – Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Redações Corrigidas (RC)

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.79885	0.892	0.001956	0.05
T2	0.92381	0.901	0.1331	0.05

Fonte: elaborado pelo autor.

Ao se analisar os dados da tabela, temos o mesmo caso acontecendo no tratamento T1, uma vez que $W_{calculado}(T1) = 0.79885 < W_{\alpha}(T1) = 0.892$ e $P_{valor}(T1) = 0.001956 < \alpha = 0.05$. Com isso, podemos refutar a hipótese nula do teste de Shapiro-Wilk e consequentemente, podemos afirmar com nível de significância de 5% que a amostra não provém de uma população normal.

Contudo temos um caso especial no tratamento T2 que o difere dos outros. Se analisarmos os dados do tratamento T2, temos que $W_{calculado}(T2) = 0.92381 > W_{\alpha}(T2) = 0.901$ e $P_{valor}(T2) = 0.1331 > \alpha = 0.05$. Nesse caso, devido aos dois valores serem maiores, não conseguimos refutar a hipótese nula do teste de Shapiro e, portanto, os dados desse tratamento, com significância de 5%, provém de uma população normal.

Mesmo nesse caso do tratamento T2 ser de uma população normal, não podemos usar um teste paramétrico como o T-Test devido aos dados do tratamento T1 serem não normais e, por este motivo, continuaremos utilizando o Teste de Wilcoxon. Os dados da execução do teste de Wilcoxon estão apresentados na Tabela 33.

Tabela 33 – Resultado da aplicação dos testes estatísticos com os dados da métrica Redação Corrigida (RC)

Tratamento	Teste Utilizado	v/w	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	139	0.4769

Fonte: elaborado pelo autor.

Assim como na métrica anterior, podemos observar que o valor de $p\text{-value}(T1 \times T2) = 0.4769$ obtido com a execução do teste é maior do que $\alpha = 0.05$. Portanto, com 95% de confiança, estatisticamente, não pode se afirmar que os valores da quantidade de redações corrigidas obtidas entre os tratamentos possuem diferenças significativas entre si, isto é, não há evidências estatísticas que mostrem a não equivalência da quantidade de redações corrigidas.

O fato desse coeficiente ser maior que o α indica que os resultados dessa métrica são equivalentes (iguais), não gerando evidência estatística suficiente para refutar a hipótese nula $H3-0$, implicando que **O uso de técnicas de gamificação no modelo de avaliação por pares não traz diferenças na quantidade de redações corrigidas na plataforma.**

6.4 Principais conclusões

O experimento planejado, executado e descrito nesse capítulo teve como principal objetivo avaliar o impacto da gamificação dentro do modelo de avaliação por pares proposto no trabalho, do ponto de vista de três métricas (Acesso (A), Redações Feitas (RF) e Redações Corrigidas (RC)) importantes. Assim como no experimento anterior, os dados de

cada variável foram coletados ao longo da realização do experimento e permitiram que essas variáveis pudessem ser avaliadas separadamente.

Como resultado das análises dos dados apresentadas acima (tanto análise descritiva quando as verificações de hipóteses), foi possível concluir que:

- Houve variações estatísticas significativas na quantidade de acesso (logins) dos estudantes entre o modelo com gamificação e o modelo sem gamificação. Isso indica que a gamificação incentivou os alunos a usarem mais a plataforma. Nas análises preliminares, o aumento da quantidade de login foi 64.28%.
- As análises preliminares indicaram que houve cerca de 10.53% mais redações realizadas no tratamento T2. Contudo, nas análises estatísticas não houve variações significativas da quantidade de redações realizadas na plataforma entre os modelos.
- As análises preliminares indicaram que houve cerca de 20% mais redações corrigidas no tratamento T2. Contudo, nas análises estatísticas não houve variações significativas da quantidade de redações corrigidas na plataforma entre os modelos.

7 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho apresentou um modelo de avaliação por pares com técnicas de gamificação incorporadas que tem como objetivo prover uma solução para a inclusão de avaliações escritas em ambientes educacionais online de forma eficiente. Tal modelo foi apresentado através de diagramas de entidade relacionamentos, diagrama de classes, imagens contendo seus fluxos e sua estrutura, que ilustravam tanto o seu processo de implementação quanto seu processo de execução.

A necessidade da criação desse modelo surgiu do grande número de estudantes que estavam presentes nos ambientes e da grande dificuldade em prover provas discursivas neles, uma vez que isso gerava um alto custo e uma grande sobrecarga nos professores envolvidos no processo. Dessa forma, as atividades discursivas eram limitadas, dificultando a aprendizagem do aluno no processo.

O modelo proposto foi construído utilizando tecnologias atuais, como banco de dados semânticos (ontologias), implementação em JAVA baseada em componentes no modelo COSMOS e provimento de seus serviços via interfaces web (serviços web), para que pudesse ser facilmente integrado aos diversos ambientes educacionais. O processo de integração também foi ilustrado com a integração do modelo proposto junto ao ambiente educacional MeuTutor-ENEM (um sistema web inteligente que prepara alunos do Ensino Médio para a prova do ENEM). Essa integração permitiu a inclusão da disciplina de redação para que os alunos pudessem melhorar seus conhecimentos.

O modelo proposto atingiu o objetivo proposto nesse trabalho, possibilitando a inclusão de avaliações discursivas nos ambientes online de maneira viável, como podemos observar em sua integração com o ambiente educacional MeuTutor e seu uso com provas de redação aplicadas para cerca de 30 alunos. Os experimentos apresentados também deram comprovações estatísticas suficientes de que os resultados obtidos com o modelo proposto são semelhantes aos resultados do modelo tradicional.

O primeiro experimento que foi proposto, planejado e executado teve como objetivo avaliar e comparar os resultados obtidos do modelo de avaliação por pares proposto com os resultados da aplicação com o modelo tradicional (correção por especialistas) sob as métricas de nota final, tempo de correção e custo associado. Os resultados foram bastante satisfatórios, tendo em vista que foi comprovado, estatisticamente, que as notas dos modelos são equivalentes, permitindo que a substituição entre eles possa ser feita, sem comprometer os resultados finais. Da mesma forma, os resultados da métrica Tempo indicaram uma possibilidade de o tempo de correção ser menor usando o modelo proposto, contudo em alguns casos esse tempo seja semelhante. Por fim, os resultados da métrica Custo foram bem favoráveis a aplicação do modelo, tendo em vista que não é necessária uma correção por especialista no modelo proposto. O modelo de regressão criado mostrou

que o custo é cerca de 72.4% menor.

O segundo experimento realizado tinha como objetivo analisar apenas o impacto da gamificação dentro do modelo de avaliação por pares proposto. A avaliação foi feita sob três variáveis: quantidade de acesso, número de redações feitas e número de redações corrigidas. Com relação a quantidade de acesso, foi comprovado estatisticamente que a gamificação influencia de forma positiva, aumentando a participação dos alunos. Os dados preliminares indicaram um aumento de 64.28% na quantidade de logins. Já com relação a quantidade de redações feitas e corrigidas, não foi possível comprovar estatisticamente que houve influência da gamificação, porém os dados preliminares indicaram que houve um aumento de 10.53% na métrica de redações feitas e cerca de 20% no número de redações corrigidas.

Diante do apresentado, uma das contribuições foi o modelo de avaliação por pares gamificado para avaliações escritas, que poderá ser incrementado em ambientes educacionais online, o qual utiliza avaliação por pares aplicado na correção de atividades discursivas em ambientes educacionais online. Este modelo de avaliação por pares é funcional, envolvendo desde a etapa de seleção dos alunos para realizar a avaliação, passando pelo gerenciamento das correções e métodos para garantir confiabilidade, até a entrega do relatório final ao estudante. Este modelo poderá ser aplicado em todo ou adaptado em diversos outros ambientes educacionais.

Com o modelo de avaliação por pares proposto juntamente com técnicas de gamificação é possível unir as vantagens de ambos, diminuindo a sobrecarga no professor, economizando seu tempo e esforço para o que realmente é necessário, e consequentemente, diminuindo o custo associado as correções e possibilitando a expansão dos cursos para centenas ou até mesmo milhares de usuários, assim como ocorre nos MOOCs. Além disso, o uso de gamificação deixa os alunos mais motivados a participarem das atividades, e consequentemente, há um aumento na aprendizagem do aluno.

Com a abordagem de avaliação por pares aplicada na correção destas atividades nestes ambientes online, os alunos poderão praticar seus conhecimentos de uma maneira "ilimitada", seja enviando suas respostas bem como corrigindo as respostas de seus pares. Para que o modelo de avaliação por pares funcione com todo o seu potencial foi acrescentado técnicas de gamificação como aspecto motivacional para o estudante.

Outra contribuição dessa proposta foi a realização de uma revisão sistemática de literatura a fim de analisar o estado da arte da área no qual o trabalho está inserido (avaliação por pares aplicadas em ambientes online). Como resultado dessa revisão temos dos 3114 artigos importados, apenas 44 atenderam os critérios e foram analisados. Concluímos que o principal contexto envolvido nos trabalhos é voltado à educação do ponto de vista acadêmico aplicados em nível superior. Há diversas finalidades de uso, porém a grande maioria se aplica na correção de atividades nos cursos. Houveram evidências de uma melhora no desempenho dos alunos em 60% dos artigos e benefícios aos professores em

pelo menos 34%. Em alguns estudos foram apresentadas dificuldades do uso da técnica de avaliação por pares.

Além disso, os resultados gerados pelos trabalhos realizados ao longo do processo puderam ser compartilhados com os pesquisadores de diversas áreas através da publicação de artigos científicos em periódicos e conferências internacionais. Os artigos gerados ao longo do período do mestrado estão ilustrados no Apêndice F, dando destaque aos dois primeiros artigos, que são diretamente relacionados ao trabalho proposto nessa dissertação.

Contudo, existem algumas limitações e possibilidades de trabalhos futuros. A decisão da implementação de quais serviços estariam disponíveis na camada de serviços web foi pensada de uma maneira mais abstrata na tentativa de que todos os ambientes pudessem utilizar estes serviços sem problemas. No entanto, alguns ambientes educacionais, devido às suas limitações tecnológicas, podem requerer serviços específicos que não foram implementados inicialmente. Nesse sentido temos como trabalho futuro implementar mecanismos que tornem o modelo e sua implementação mais flexível do ponto de vista de funcionalidades, permitindo que as adições de novas funcionalidades sejam feitas de forma mais rápida.

A implementação do modelo, apesar de ter sido pensada e construída de maneira abstrata e independente de ambiente, ainda sofreu uma forte influência tecnológica do ambiente educacional MeuTutor (o qual tinha a necessidade de sua criação). Isso implica no uso de diversas tecnologias do MeuTutor, como o uso de ontologias, implementação em Java, estilo arquitetural em camadas e baseado em componentes COSMOS, e provimento de serviços WEB em WSDL e SOAP. Apesar de não ter problemas específicos ao usar essas tecnologias, isso pode gerar dificuldades na implantação com outros ambientes educacionais que não foram pensados a priori, devido a incompatibilidade de tecnologias e/ou devido às tecnologias ultrapassadas de alguns modelos.

Neste sentido, como trabalho futuro temos o planejamento de implantar o modelo em outros ambientes educacionais diferentes do MeuTutor, como por exemplo, o Moodle. Com a implementação/integração em outros ambientes, as dificuldades que por ventura poderão aparecer servirão como entrada para possíveis melhorias na implementação do modelo, deixando-o mais completo, flexível e compatível com a maioria dos ambientes educacionais.

Os experimentos realizados do modelo foram incorporados no ambiente do Meututor. Apesar de ter sido planejado de forma imparcial, o fato de que o modelo foi proposto sob as necessidades do Meututor, bem como a tecnologia de sua implementação ser compatível com o Meututor, pode ter influenciado de forma positiva na implantação do modelo e seus resultados. Sendo assim, temos como trabalho futuro o planejamento de realizar novos experimentos em condições diferentes, tanto com relação ao ambiente utilizado, bem como os usuários envolvidos e a forma de avaliação do modelo. Também pretendemos aumentar a amostra utilizada.

Além disso, é preciso avaliar novas métricas de comparação do modelo proposto com o modelo tradicional que podem ser relevantes, como por exemplo, o nível de aprendizagem do aluno, melhorias das suas capacidades de julgamento, entre outros benefícios que um modelo de avaliação por pares pode trazer. Com isso, pretende-se realizar novos experimentos mais completos a fim de avaliar tais métricas.

Por fim, outro ponto que temos como trabalho futuro diz respeito à gamificação aplicada no modelo. Sabe-se que a gamificação, em geral, tem um papel importante, principalmente como um aspecto motivacional aos usuários envolvidos. Contudo, se aplicada de forma errada, pode influenciar de forma negativa nos resultados. Na construção do modelo, todo o plano de gamificação foi construído pensando em um usuário padrão de um ambiente. Porém, sabemos que os usuários apresentam perfis de gamificação diferentes, onde cada um desses tem características próprias diferentes do padrão.

Nesse sentido, como trabalho futuro, temos que incorporar e implementar mecanismos ao modelo que tornem o seu plano de gamificação adaptável ao perfil do usuário, de tal maneira que sempre apresente aspectos positivos. No entanto, para se conseguir tal mecanismo, é necessário avaliar quais e como os elementos de gamificação influenciam em cada perfil de usuário. Sendo assim, planejaremos um experimento controlado com o objetivo de descobrir essas influências e criarmos planos de gamificação adaptáveis.

REFERÊNCIAS

- ACM. *The 2012 ACM Computing Classification System*. 2012. Acessado em 27/01/2015. Disponível em: <<http://www.acm.org/about/class/class/2012>>.
- ANDRADE, F. R. et al. Desafio do uso de gamificação em sistemas tutores inteligentes baseados em web semântica. In: *XXXIII Congresso da Sociedade Brasileira de Computação*. [S.l.: s.n.], 2013. p. 1453–1462.
- BAKER, R. et al. Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research*, v. 19, n. 2, p. 185–224, 2008.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. *Estatística: para cursos de engenharia e informática*. [S.l.]: Atlas São Paulo, 2004. v. 3.
- BIGGS, J.; TANG, C. *Teaching for quality learning at university: What the student does*. [S.l.]: McGraw-Hill Education (UK), 2011.
- BITTENCOURT, I. I. et al. Research directions on semantic web and education. *Interdisciplinary Studies in Computer Science*, v. 19, n. 1, p. 60–67, 2008.
- BLACK, P. et al. *Assessment for learning: Putting it into practice*. [S.l.]: McGraw-Hill International, 2003.
- BLACK, P.; WILIAM, D. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, Springer, v. 21, n. 1, p. 5–31, 2009.
- BUSCHMANN, F.; HENNEY, K.; SCHIMDT, D. *Pattern-oriented Software Architecture: On Patterns and Pattern Language*. [S.l.]: John Wiley & Sons, 2007. v. 5.
- CARDOSO, J. The semantic web vision: Where are we? *Intelligent Systems, IEEE*, IEEE, v. 22, n. 5, p. 84–88, 2007.
- CARVALHO, A. d. L.; LEON, F. d. Ponce de et al. Grandes desafios da pesquisa em computação no brasil–2006–2016. *São Paulo: Sociedade Brasileira de Computação*, 2006.
- CAVALCANTI, S. R. *Veer: Um algoritmo de seleção de pares em redes ad hoc veiculares*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, 2008.
- CHANG, C.-C.; TSENG, K.-H.; LOU, S.-J. A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students. *Computers & Education*, Elsevier, v. 58, n. 1, p. 303–320, 2012.
- CORDOVA, D. I.; LEPPER, M. R. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology*, American Psychological Association, v. 88, n. 4, p. 715, 1996.
- COSTA, M. T. C. d. *Uma arquitetura baseada em agentes para suporte ao ensino à distância*. Dissertação (Doutorado) — Universidade Federal de Santa Catarina, 1999.

- CUNHA, E.; FIGUEIRA, Á. A web-based tool for assessing online peer-reviews. In: ACADEMIC CONFERENCES LIMITED. *8th European Conference on E-Learning, University of Bari, Italy, 29-30 October 2009*. [S.l.], 2009. p. 132.
- CUNHA, L. F. D.; GASPARINI, I.; BERKENBROCK, C. D. M. Investigando o uso de gamificação para aumentar o engajamento em sistemas colaborativos. In: *WAIHCWS*. [S.l.: s.n.], 2013. p. 28–33.
- CUNHA, L. F. D.; GASPARINI, I.; BERKENBROCK, C. D. M. Investigando o uso de gamificação para aumentar o engajamento em sistemas colaborativos. In: *WAIHCWS*. [S.l.: s.n.], 2013. p. 28–33.
- CYTOWIC, R. E. *Synesthesia: A union of the senses*. [S.l.]: MIT press, 2002.
- DENNY, P. The effect of virtual achievements on student engagement. In: ACM. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S.l.], 2013. p. 763–772.
- DETERDING, S. et al. From game design elements to gamefulness: defining gamification. In: ACM. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. [S.l.], 2011. p. 9–15.
- DEVEDŽIC, V. *Semantic web and education*. [S.l.]: Springer Science & Business Media, 2006. v. 12.
- DOMINGUEZ, C. et al. Online peer assessment: An exploratory case study in a higher education civil engineering course. In: IEEE. *Interactive Collaborative Learning (ICL), 2012 15th International Conference on*. [S.l.], 2012. p. 1–8.
- EASTERBROOK, S. et al. Selecting empirical methods for software engineering research. In: *Guide to advanced empirical software engineering*. [S.l.]: Springer, 2008. p. 285–311.
- ESPINHEIRA, P. L.; SILVA, L. C. M. da; SILVA, A. d. O. Prediction measures in beta regression models. *arXiv preprint arXiv:1501.04830*, 2015.
- FALCHIKOV, N. Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, Taylor & Francis, v. 11, n. 2, p. 146–166, 1986.
- FERMELIS, J.; TUCKER, R.; PALMER, S. Online self and peer assessment in large, multi-campus, multi-cohort contexts. In: ASCILITE. *Ascilite 2007: ICT: Providing Choices for Learners and Learning*. [S.l.], 2007. p. 271–281.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004.
- FLATLA, D. R. et al. Calibration games: making calibration tasks enjoyable by adding motivating game elements. In: ACM. *Proceedings of the 24th annual ACM symposium on User interface software and technology*. [S.l.], 2011. p. 403–412.
- FLETCHER, P. R. A teoria da resposta ao item: medidas invariantes do desempenho escolar. *Ensaio*, revista da Fundação Cesgranrio Rio de Janeiro, v. 1, n. 2, p. 21–27, 1994.

GALVÃO, C.; SAWADA, N.; TREVIZAN, M. Revisão sistemática. *Rev Latino-am enfermagem*, SciELO Brasil, v. 12, n. 3, p. 549–56, 2004.

GEE, J. P. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, ACM, v. 1, n. 1, p. 20–20, 2003.

GIRAFFA, L. M. *Uma arquitetura de tutor utilizando estados mentais*. Dissertação (Doutorado) — Universidade Federal do Rio Grande do Sul, 1999.

GOULI, E.; GOGOULOU, A.; GRIGORIADOU, M. Supporting self-, peer-, and collaborative-assessment in e-learning: the case of the peer and collaborative assessment environment (pecasse). *Journal of Interactive Learning Research*, v. 19, n. 4, p. 615–647, 2008.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, Academic Press, Inc., Duluth, MN, USA, v. 43, n. 5-6, p. 907–928, dez. 1995. ISSN 1071-5819. Disponível em: <<http://dx.doi.org/10.1006/ijhc.1995.1081>>.

GUARINO, N. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. [S.l.]: IOS press, 1998. v. 46.

HAMARI, J.; LEHDONVIRTA, V. Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science & Applied Management*, v. 5, n. 1, p. 14–29, 2010.

HEATH, T. *Linked Data - Connect Distributed Data across the Web*. 2013. Acessado em 27/02/2015. Disponível em: <www.linkeddata.org>.

HOLANDA, O. C. N. *JOINT-DE: Sistema de Mapeamento Objeto-Ontologia com Suporte a Objetos Desconectados*. Dissertação (Mestrado) — Universidade Federal de Alagoas - UFAL, 2014.

HONEYCHURCH, S. et al. Peer assessment assisted by technology. *International Journal of e-Assessment*, v. 3, n. 1, 2013.

HUOTARI, K.; HAMARI, J. Defining gamification: a service marketing perspective. In: ACM. *Proceeding of the 16th International Academic MindTrek Conference*. [S.l.], 2012. p. 17–22.

Instituto Nacional de Estudos e Pesquisas Educacionais (INEP). *Guia do participante do ENEM*. 2013. Acessado em 15/08/2014. Disponível em: <<http://portal.inep.gov.br/web/enem/enem>>.

JAIN, R. *The art of computer systems performance analysis*. [S.l.]: John Wiley & Sons, 2008.

JONASSEN, D. H.; WANG, S. The physics tutor: Integrating hypertext and expert systems. *Journal of Educational Technology Systems*, Baywood, v. 22, n. 1, p. 19–28, 1993.

KAPP, K. M. *The gamification of learning and instruction: game-based methods and strategies for training and education*. [S.l.]: John Wiley & Sons, 2012.

- KAWAI, G. Collaborative peer-based language learning in unsupervised asynchronous online environments. In: IEEE. *Creating, Connecting and Collaborating through Computing, 2006. C5'06. The Fourth International Conference on*. [S.l.], 2006. p. 35–41.
- KEELE, S. *Guidelines for performing systematic literature reviews in software engineering*. [S.l.], 2007.
- KIGOZI, E. K. et al. Modelling a peer assignment review process for collaborative e-learning. *Journal of Interactive Online Learning*, v. 11, n. 2, p. 67–79, 2012.
- KIM, A. J. *Beyond Player Types: Kim's Social Action Matrix*. 2014. Acesso em 15/02/2015. Disponível em: <<http://amyjokim.com/2014/02/28/beyond-player-types-kims-social-action-matrix/>>.
- KLOPFER, E. *Augmented learning: Research and design of mobile educational games*. [S.l.]: Mit Press, 2008.
- KNIGHT, L.; STEINBACH, T. Adapting peer review to an online course: An exploratory case study. *Journal of Information Technology Education: Research*, v. 10, n. 1, p. 81–100, 2011.
- KOBSA, A. User modeling and user-adapted interaction. In: ACM. *Conference companion on Human factors in computing systems*. [S.l.], 1994. p. 415–416.
- LI, C. et al. Engaging computer science students through gamification in an online social network based collaborative learning environment. *International Journal of Information and Education Technology*, v. 3, n. 1, p. 72–77, 2013.
- LOINAZ, M. Sistemas inteligentes em el ámbito de la educación. *Disponível em: <http://aepia.dsic.upv.es/revista/numeros/12/Urretavizcaya.pdf>* Acesso em, v. 5, 2001.
- MALEHORN, H. Ten measures better than grading. *The Clearing House*, Taylor & Francis, v. 67, n. 6, p. 323–324, 1994.
- MALONE, T. W.; LEPPER, M. R. Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, v. 3, n. 1987, p. 223–253, 1987.
- MCCUINNESS, D. L. Ontologies come of age. *Spinning the semantic web: bringing the World Wide Web to its full potential*, MIT Press, p. 171, 2005.
- MCDOWELL, L.; MOWL, G.; GIBBS, G. Innovative assessment: its impact on students. *Improving student learning through assessment and evaluation*. Oxford: Oxford Centre for Staff Development. McLean, M., Henson, Q., Hiles, L.(2003) *The possible contribution of student drawings to evaluation in a new problem-based learning medical programme: a pilot study*. *Medical Education*, v. 37, n. 10, p. 895–906, 1995.
- MERCERON, A.; YACEF, K. Tada-ed for educational data mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, v. 7, n. 1, p. 267–287, 2005.
- MESSA, W. C. Utilização de ambientes virtuais de aprendizagem-avas: A busca por uma aprendizagem significativa. *Revista Brasileira de Aprendizagem Aberta ea Distância*, v. 9, p. 01–49, 2010.

- MIAO, Y.; KOPER, R. An efficient and flexible technical approach to develop and deliver online peer assessment. In: INTERNATIONAL SOCIETY OF THE LEARNING SCIENCES. *Proceedings of the 8th international conference on Computer supported collaborative learning*. [S.l.], 2007. p. 506–515.
- MITCHELL, M. *An introduction to genetic algorithms*. [S.l.]: MIT press, 1998.
- MIYAKE, N. Computer supported collaborative learning. *The SAGE handbook of e-learning research*, p. 248–265, 2007.
- MORAN, J. M. *O que é educação a distância*. 2008.
- MOREIRA, B. G. Desenvolvimento de uma ferramenta de avaliação por pares para disciplinas de algoritmos e programação. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*. [S.l.: s.n.], 2014. v. 3, n. 1.
- MULLIGAN, A.; HALL, L.; RAPHAEL, E. Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, v. 64, n. 1, p. 132–161, 2013.
- NICHOLSON, S. A user-centered theoretical framework for meaningful gamification. *Games+ Learning+ Society*, v. 8, n. 1, 2012.
- NIKKILA, S. et al. Taskville: visualizing tasks and raising awareness in the workplace. In: ACM. *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. [S.l.], 2013. p. 151–156.
- ORSMOND*, P.; MERRY, S.; CALLAGHAN, A. Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, Taylor & Francis, v. 41, n. 3, p. 273–290, 2004.
- PAIVA¹, R. et al. Mineração de dados e a gestão inteligente da aprendizagem: Desafios e direcionamentos. 2012.
- PAVLUS, J. The game of life. *Scientific American*, Nature Publishing Group, v. 303, n. 6, p. 43–44, 2010.
- PIECH, C. et al. Tuned models of peer assessment in moocs. *arXiv preprint arXiv:1307.2579*, 2013.
- PURHONEN, O. *Split Testing Guide for Online Stores*. 2012. Acessado em 01/07/2015. Disponível em: <<http://www.webics.com.au/blog/google-adwords/split-testing-guide-for-online-retailers/>>.
- READING, U. of. *Engage in Assessment - An online toolkit enabling staff to enhance all aspects of assessment*. 2014. <<http://www.reading.ac.uk/engageinassessment/>>. Acessado em 20/08/2014.
- ROSA, K. D.; ESKENAZI, M. Self-assessment in the reap tutor: Knowledge, interest, motivation, & learning. *IJ Artificial Intelligence in Education*, v. 21, n. 4, p. 237–253, 2013.

- ROURKE, A. J. et al. Did i mention it's anonymous? the triumphs and pitfalls of online peer review. In: *Hello! Where are you in the landscape of educational technology? Proceedings of ascilite conference*. [S.l.: s.n.], 2008. v. 30, p. 830–840.
- SADLER, P. M.; GOOD, E. The impact of self-and peer-grading on student learning. *Educational assessment*, Taylor & Francis, v. 11, n. 1, p. 1–31, 2006.
- SALEN, K.; ZIMMERMAN, E. *Rules of play: Game design fundamentals*. [S.l.]: MIT press, 2004.
- SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Braz. J. Phys. Ther.(Impr.)*, v. 11, n. 1, p. 83–89, 2007.
- SCHELL, J. *The Art of Game Design: A book of lenses*. [S.l.]: CRC Press, 2014.
- SCIENCES, N. I. G. M. *Protégé. A free, open-source ontology editor and framework for building intelligent systems*. 2013. Acessado em 27/01/2015. Disponível em: <<http://protege.stanford.edu/overview/>>.
- SCOTT, A. Peer review and the relevance of science. *Futures*, Elsevier, v. 39, n. 7, p. 827–845, 2007.
- SHAFFER, D. W. *How computer games help children learn*. [S.l.]: Macmillan, 2006.
- SHAPIRO, S. S.; FRANCA, R. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 67, n. 337, p. 215–216, 1972.
- SILVA, A. P. d. et al. Ontology-based software product line for building semantic web applications. In: ACM. *Proceedings of the 2010 Workshop on Knowledge-Oriented Product Line Engineering*. [S.l.], 2010. p. 1.
- SILVA, M. C. d. *COSMOS : um modelo de estruturação de componentes para sistemas orientados a objetos*. Dissertação (Mestrado) — Universidade Estadual de Campinas - UNICAMP, 2003.
- SLEEMAN, D.; BROWN, J. S. *Intelligent Tutoring Systems*. London : Academic Press, 1982. 345 pages p. Disponível em: <<http://hal.archives-ouvertes.fr/hal-00702997>>.
- SMITH, M. K.; WELTY, C.; MCGUINNESS, D. L. *Protégé. A free, open-source ontology editor and framework for building intelligent systems*. 2004. Acessado em 15/02/2015. Disponível em: <<http://www.w3.org/TR/owl-guide/>>.
- STERBINI, A.; TEMPERINI, M. Openanswer, a framework to support teacher's management of open answers through peer assessment. In: IEEE. *Frontiers in Education Conference, 2013 IEEE*. [S.l.], 2013. p. 164–170.
- SUNG, Y.-T. et al. The design and application of a web-based self-and peer-assessment system. *Computers & Education*, Elsevier, v. 45, n. 2, p. 187–202, 2005.
- TOPPING, K. J. Peer assessment. *Theory into practice*, Taylor & Francis, v. 48, n. 1, p. 20–27, 2009.

- TOSIC, M.; NEJKOVIC, V. *Trust-based peer assessment for virtual learning systems*. [S.l.]: Springer, 2010.
- TRAHASCH, S. From peer assessment towards collaborative learning. In: IEEE. *Frontiers in Education, 2004. FIE 2004. 34th Annual*. [S.l.], 2004. p. F3F–16.
- VANLEHN, K. et al. The affective meta-tutoring project: How to motivate students to use effective meta-cognitive strategies. In: *19th International Conference on Computers in Education, Chiang Mai, Thailand*. [S.l.: s.n.], 2011.
- VICARI, R. M. Tutor inteligente para a programação em lógica: idealização, projecto e desenvolvimento. 1989.
- VICARI, R. M.; GIRAFFA, L. M. Fundamentos dos sistemas tutores inteligentes. *Barone, D.; et alii. Sociedades artificiais: a nova fronteira da inteligência nas máquinas*. Porto Alegre: Bookman, 2003.
- WANG, Y.; VASSILEVA, J. Trust and reputation model in peer-to-peer networks. In: IEEE. *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*. [S.l.], 2003. p. 150–157.
- WANG, Y.; VASSILEVA, J. Bayesian network trust model in peer-to-peer networks. In: *Agents and Peer-to-Peer Computing*. [S.l.]: Springer, 2005. p. 23–34.
- WELCH, B. L. The significance of the difference between two means when the population variances are unequal. *Biometrika*, JSTOR, p. 350–362, 1938.
- WERBACH, K.; HUNTER, D. *For the win: How game thinking can revolutionize your business*. [S.l.]: Wharton Digital Press, 2012.
- WILCOXON, F. Individual comparisons by ranking methods. *Biometrics bulletin*, JSTOR, p. 80–83, 1945.
- WITTEN, I. H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. [S.l.]: Morgan Kaufmann, 2005.
- WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer Science & Business Media, 2012.
- WOOLF, B. P. *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. [S.l.]: Morgan Kaufmann, 2010.
- WULF, J. et al. Massive open online courses. *Business & Information Systems Engineering*, v. 6, n. 2, p. 111–114, 2014.
- YANG, K.; EL-HAIK, B. S. *Design for six sigma*. [S.l.]: McGraw-Hill New York, 2003.
- YANG, Y.-F.; MENG, W.-T. The effects of online feedback training on students' text revision. *Language Learning & Technology*, v. 17, n. 2, p. 220–238, 2013.
- ZAMBONI, A. et al. Start uma ferramenta computacional de apoio à revisão sistemática. In: *Proc.: Congresso Brasileiro de Software (CBSOft-10), Salvador, Brazil*. [S.l.: s.n.], 2010.
- ZICHERMANN, G.; CUNNINGHAM, C. *Gamification by design: Implementing game mechanics in web and mobile apps*. [S.l.]: "O'Reilly Media, Inc.", 2011.

APÊNDICE A – LISTA COM OS ARTIGOS SELECIONADOS PARA REVISÃO SISTEMÁTICA

Os artigos selecionados para a revisão sistemática foram:

1. Phielix, Chris, Frans J. Prins, and Paul A. Kirschner. “Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection.” *Computers in Human Behavior* 26.2 (2010): 151-161.
2. Kulkarni, Chinmay, et al. “Peer and self assessment in massive online classes.” *Design Thinking Research*. Springer International Publishing, 2015. 131-168.
3. Sterbini, Andrea, and Marco Temperini. “Analysis of open answers via mediated peer-assessment.” *System Theory, Control and Computing (ICSTCC)*, 2013 17th International Conference. IEEE, 2013.
4. Sterbini, Andrea, and Marco Temperini. “OpenAnswer, a framework to support teacher’s management of open answers through peer assessment.” *Frontiers in Education Conference, 2013 IEEE*. IEEE, 2013.
5. Honeychurch, Sarah, et al. “Peer assessment assisted by technology.” *International Journal of e-Assessment* 3.1 (2013).
6. Rosa, Kevin Dela, and Maxine Eskenazi. “Self-Assessment in the REAP Tutor: Knowledge, Interest, Motivation and Learning.” *International Journal of Artificial Intelligence in Education* 21.4 (2013): 237-253.
7. Long, Yanjin, and Vincent Aleven. “Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment.” *Artificial intelligence in education*. Springer Berlin Heidelberg, 2013.
8. Yang, Yu-Fen, and Wen-Ting Meng. “THE EFFECTS OF ONLINE FEEDBACK TRAINING ON STUDENTS’TEXT REVISION.” *Language Learning and Technology* 17.2 (2013): 220-238.
9. Kahiigi Kigozi, Evelyn, et al. “Modelling a peer assignment review process for collaborative e-learning.” *Journal of Interactive Online Learning* 11.2 (2012): 67-79.
10. Dominguez, Caroline, et al. “Online peer assessment: An exploratory case study in a higher education civil engineering course.” *Interactive Collaborative Learning (ICL)*, 2012 15th International Conference on. IEEE, 2012.

11. Sterbini, Andrea, and Marco Temperini. "Supporting assessment of open answers in a didactic setting." *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on*. IEEE, 2012.
12. Chang, Chi-Cheng, Kuo-Hung Tseng, and Shi-Jer Lou. "A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students." *Computers and Education*. 58.1 (2012): 303-320.
13. Knight, Linda, and Theresa Steinbach. "Adapting peer review to an online course: An exploratory case study." *Journal of Information Technology Education: Research* 10.1 (2011): 81-100.
14. De Graaff, E., and Gillian Saunders-Smits. "Peer evaluation and Peer review to support PBL teamwork." *Wireless Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology (Wireless VITAE), 2011 2nd International Conference on*. IEEE, 2011.
15. Chang, Chi-Cheng, et al. "Reliability and validity of Web-based portfolio peer assessment: A case study for a senior high school's students taking computer course." *Computers and Education* 57.1 (2011): 1306-1316.
16. Liu, Feng-Jung, Chun-Wei Tseng, and Wen-Chang Tseng. "Constructing problem-based learning activities using self-assessment system." *Computational Collective Intelligence. Technologies and Applications*. Springer Berlin Heidelberg, 2010. 227-235.
17. Yoshizawa, Sayuri, Takao Terano, and Atsushi Yoshikawa. "Analyzing the effects of peer review activities in the EFL writings." *Proceedings of the 18th International Conference in Computers in Education*. 2010.
18. Tasic, Milorad, and Valentina Nejkovic. *Trust-based peer assessment for virtual learning systems*. Springer Berlin Heidelberg, 2010.
19. Yang, Yu-Fen. "Student's reflection on online self-correction and peer review to improve writing." *Computers and Education* 55.3 (2010): 1202-1210.
20. Giannoukos, Ioannis, et al. "An adaptive mechanism for author-reviewer matching in online peer assessment." *Semantics in Adaptive and Personalized Services*. Springer Berlin Heidelberg, 2010. 109-126.
21. Phielix, Chris, Frans J. Prins, and Paul A. Kirschner. "The design of peer feedback and reflection tools in a CSCL environment." *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 1*. International Society of the Learning Sciences, 2009.

22. Bauer, Christine, et al. "The student view on online peer reviews." *ACM SIGCSE Bulletin*. Vol. 41. No. 3. ACM, 2009.
23. Joglar, Nuria, et al. "iTest: an online tool for assessment and self-assessment in Mathematics." *Multimedia*, 2009. ISM'09. 11th IEEE International Symposium on. IEEE, 2009.
24. Chen, Nian-Shing, et al. "Effects of high level prompts and peer assessment on online learners' reflection levels." *Computers and Education* 52.2 (2009): 283-291.
25. Cunha, Elisabete, and Álvaro Figueira. "A Web-Based Tool for Assessing Online Peer-Reviews." 8th European Conference on E-Learning, University of Bari, Italy, 29-30 October 2009. Academic Conferences Limited, 2009.
26. Rourke, Arianne Jennifer, et al. "Did I mention it's anonymous? The triumphs and pitfalls of online peer review." *Hello! Where are you in the landscape of educational technology? Proceedings of ascilite conference*. Vol. 30. 2008.
27. Gouli, Evangelia, Agoritsa Gogoulou, and Maria Grigoriadou. "Supporting self-, peer-, and collaborative-assessment in e-learning: the case of the PEer and Collaborative ASSESSment Environment (PECASSE)." *Journal of Interactive Learning Research* 19.4 (2008): 615-647.
28. Wood, Denise, and Martin Friedel. "Peer review of online learning and teaching: New technologies, new challenges." *Ascilite*, 2008.
29. Fermelis, Jan, Richard Tucker, and Stuart Palmer. "Online self and peer assessment in large, multi-campus, multi-cohort contexts." *Ascilite 2007: ICT: Providing Choices for Learners and Learning*. Ascilite, 2007.
30. Miao, Yongwu, and Rob Koper. "An efficient and flexible technical approach to develop and deliver online peer assessment." *Proceedings of the 8th international conference on Computer supported collaborative learning*. International Society of the Learning Sciences, 2007.
31. Lan, Chung Hsien, Sabine Graf, and K. Robert Lai. "Effect of Learning Styles on Peer Assessment in an Agent-based Collaborative Learning Environment." *Advanced Learning Technologies, 2007. ICALT 2007*. Seventh IEEE International Conference on. IEEE, 2007.
32. de Raadt, Michael, Mark Toleman, and Richard Watson. "Electronic peer review: A large cohort teaching themselves?." *Proceedings ASCILITE 2005: 22nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Balance, Fidelity, Mobility-Maintaining the Momentum?*. Vol. 1. Queensland University of Technology, Teaching and Learning Support Services, 2005.

33. Kali, Yael, and Miky Ronen. "Design principles for online peer-evaluation: Fostering objectivity." Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!. International Society of the Learning Sciences, 2005.
34. Chiang, I-Tsun, et al. "Using game-based learning and interactive peer assessment to improve career goals and objectives for college students." *Eduentertainment technologies. Educational games and virtual reality/augmented reality applications*. Springer Berlin Heidelberg, 2011. 507-511.
35. Roll, Ido, et al. "Metacognitive practice makes perfect: Improving student's self-assessment skills with an intelligent tutoring system." *Artificial Intelligence in Education*. Springer Berlin Heidelberg, 2011.
36. Jimenez-Romero, Cristian, Jeffrey Johnson, and Ricardo De Castro. "Machine and social intelligent peer-assessment systems for assessing large student populations in massive open online education." *European Conference on e-Learning*. Academic Conferences International Limited, 2013.
37. Sterbini, Andrea, and Marco Temperini. "Dealing with open-answer questions in a peer-assessment environment." *Advances in Web-Based Learning-ICWL 2012*. Springer Berlin Heidelberg, 2012. 240-248.
38. Wang, Yanqing, et al. "Assessment of programming language learning based on peer code review model: Implementation and experience report." *Computers and Education* 59.2 (2012): 412-422.
39. Trahasch, Stephan. "From peer assessment towards collaborative learning." *Frontiers in Education*, 2004. FIE 2004. 34th Annual. IEEE, 2004.
40. Sterbini, Andrea, and Marco Temperini. "Peer-assessment and grading of open answers in a web-based e-learning setting." *Information Technology Based Higher Education and Training (ITHET)*, 2013 International Conference on. IEEE, 2013.
41. Kawai, Goh. "Collaborative peer-based language learning in unsupervised asynchronous online environments." *Creating, Connecting and Collaborating through Computing*, 2006. C5'06. The Fourth International Conference on. IEEE, 2006.
42. Coit, Caroline. "Peer review in an online college writing course." *Advanced Learning Technologies*, 2004. Proceedings. IEEE International Conference on. IEEE, 2004.
43. Colwell, Joy L., and Carl F. Jenks. "Using peer evaluations and teams in online classes." *Frontiers in Education*, 2004. FIE 2004. 34th Annual. IEEE, 2004.

44. Topping, Keith. "Peer assessment between students in colleges and universities."
Review of educational Research 68.3 (1998): 249-276.

APÊNDICE B – MODELADEM USANDO ONTOLOGIAS

Conforme explanado, a modelagem da proposta foi realizada usando ontologias. Este Apêndice tem como objetivo apresentar as ontologias adaptadas para o modelo de avaliação por pares que foram adaptadas dos modelos do ambiente educacional MeuTutor.

As principais ontologias modificadas foram Learner.owl, Resource.owl e Gamification.owl. A ontologia Learner representa a modelagem dos usuários do sistema. Neste sentido ela importa as ontologias SIOC¹ e FOAF². As principais classes da ontologia Learner que foram adaptadas são apresentadas na Figura 75.

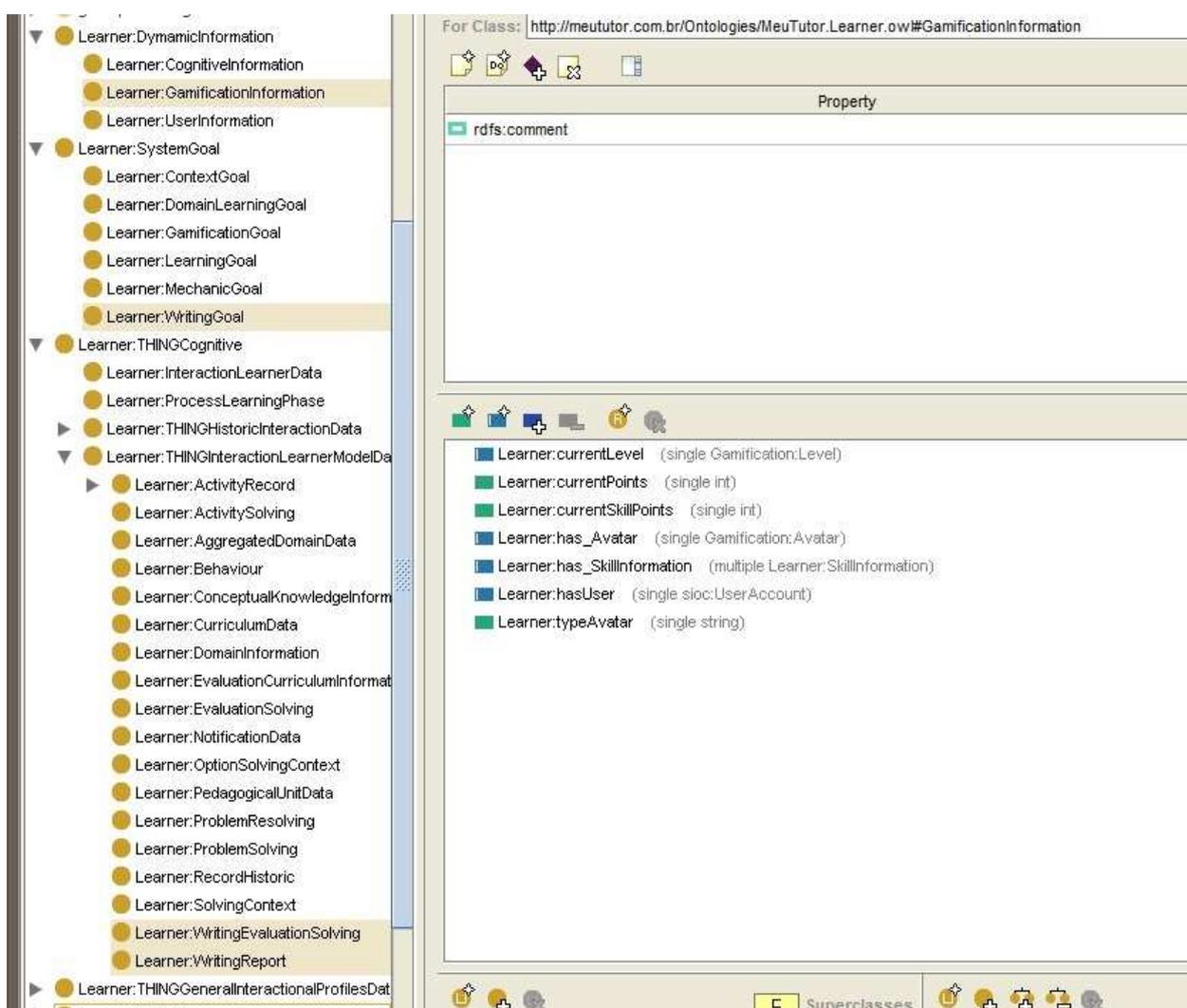


Figura 75 – Classes da Ontologia Learner.owl

Fonte: elaborado pelo autor

As classes destacadas foram as classes incluídas na ontologia para suportar as novas

¹ Especificação disponível em <http://www.w3.org/Submission/sioc-spec/>

² Disponível em <http://www.foaf-project.org/>

informações do usuário. A classe *GamificationInformation* representa as informações de gamificação do usuário. A classe *WritingGoal* representa as informações de reputação e confiança do usuário, isto é, as Informações de Objetivo. As classes *Writing evaluation solving* e *Writing Report* apresentam, respectivamente, a modelagem de uma resposta do usuário para um problema discursivo e seu respectivo relatório. A figura também ilustra os atributos e relação da classe *Gamification Information*.

Por outro lado, a ontologia *Resource.owl* representa todos os recursos disponíveis no ambiente. Neste sentido foi necessário incluir um novo tipo de recurso - os problemas discursivos. A figura 76 apresenta as classes da ontologia citada.

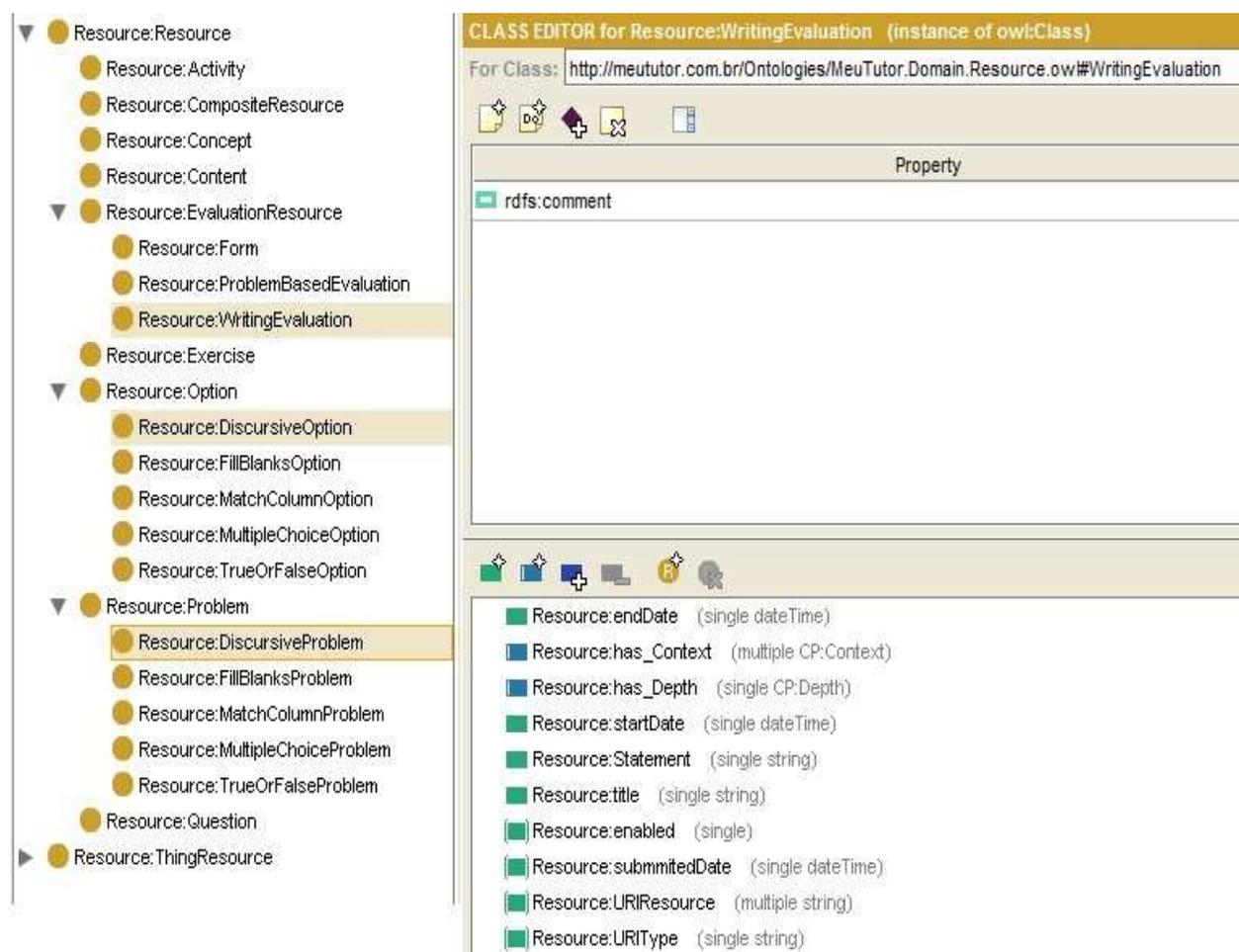


Figura 76 – Classes da Ontologia Resource.owl

Fonte: elaborado pelo autor

As classes *Writing Evaluation*, *Discursive Problem* e *Discursive Option* representam a modelagem para questões discursivas no ambiente. Pode-se ver na figura os atributos e relacionamentos da classe *Writing Evaluation*, conforme explicado e detalhado no diagrama de entidade e relacionamento da Figura 28.

Por fim, a última adaptação necessária foi na ontologia *Gamification.owl*. Esta ontologia tinha como objetivo representar os elementos de gamificação suportados pelo ambi-

ente. Neste sentido, foi preciso adaptar para suportar o framework 6D, com a inclusão do suporte a Loops de atividade, bem como suas especificações. A figura 77 apresenta as classes da ontologia de gamificação.

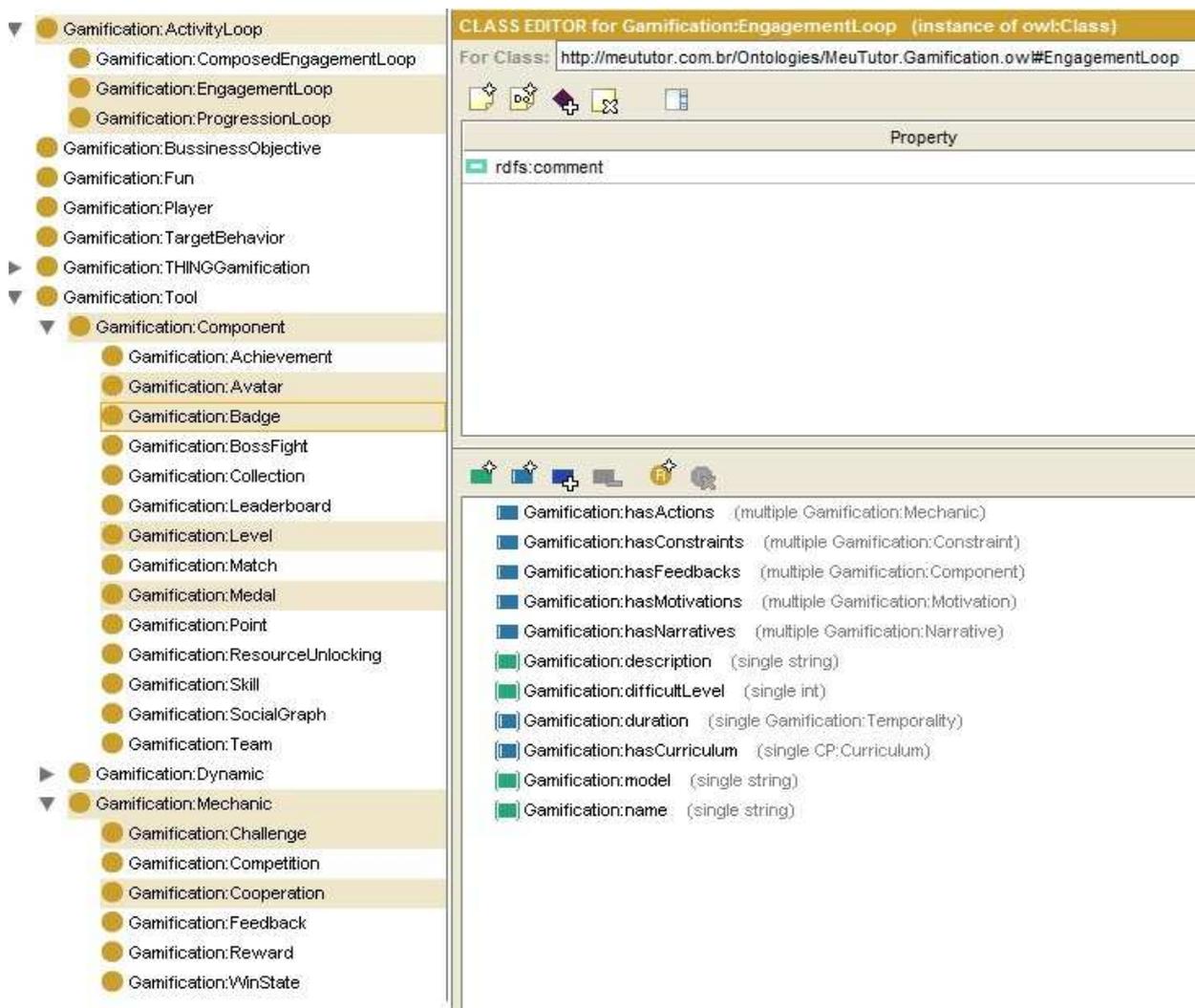


Figura 77 – Classes da Ontologia Gamificação.owl

Fonte: elaborado pelo autor

As classes destacadas na figura referem-se as classes que sofreram algumas alterações. Na figura, há um destaque maior para as propriedades e relacionamentos da classes *Engagement Loop*. Esta representa um loop de engajamento que o usuário pode realizar no sistema e suas propriedades são baseadas no modelo de entidade e relacionamento apresentado na figura 32.

APÊNDICE C – EXEMPLO INTERFACE WS PROVIDA PELO MODELO - IWRITINGEVALUATIONWS

G:\repositorio-meututor\trunks\MeuTutorAdhoc\src\java\sistema\microkernel\writingevaluation\spec\prov\IWritingEvaluationWS.java

```

1 /*
2 * To change this license header, choose License Headers in Project Properties.
3 * To change this template file, choose Tools | Templates
4 * and open the template in the editor.
5 */
6 package sistema.microkernel.writingevaluation.spec.prov;
7
8 import java.util.ArrayList;
9 import javax.jws.WebParam;
10 import javax.jws.WebService;
11 import model.dto.AssessmentFormDTO;
12 import model.dto.CodeDTO;
13 import model.dto.ContextDTO;
14 import model.dto.WritingEvaluationDTO;
15 import model.dto.WritingEvaluationSolvingDTO;
16 import model.dto.WritingReportDTO;
17 import utils.exceptions.InvalidAuthenticationException;
18
19 /**
20 *
21 * @author Thyago
22 */
23 @WebService(name = "Writing_Evaluation_Service")
24 public interface IWritingEvaluationWS {
25
26     /**
27      * Método para criar uma questão discursiva no sistema
28      *
29      * @author Thyago Tenório
30      * @param titulo String - Título da questão a ser criada
31      * @param enunciado String - Texto da questão a ser criada
32      * @param idsContext String - ID do Tema da questão ( ; para vários)
33      * @param login String - Id do usuário que criou a questão
34      * @param idForm String - Id do formulário a ser vinculado na questão
35      * @param autenticacao String - Código de autenticação para validar a
36      * segurança
37      *
38      * @return WritingEvaluationDTO - Questão criada
39      */
40     public WritingEvaluationDTO criarQuestao(@WebParam(name = "titulo") String titulo,
41         @WebParam(name = "enunciado") String enunciado,
42         @WebParam(name = "idsContext") String idsContext,
43         @WebParam(name = "login") String login,
44         @WebParam(name = "idForm") String idForm,
45         @WebParam(name = "autenticacao") String autenticacao);
46
47     /**
48      * Método para editar uma questão
49      *
50      * @author Thyago Tenório
51      * @param idQuestao String - ID da questão a ser editada
52      * @param newTitle String - Novo título que irá substituir o título antigo
53      * @param newEnunciado String - Novo enunciado que irá substituir o
54      * enunciado antigo
55      * @param autenticacao String - Código de autenticação para validar a
56      * segurança
57      *
58      * @return CodeDTO - Resultado da operação
59      */
60     public CodeDTO editarQuestao(@WebParam(name = "idQuestao") String idQuestao, @WebParam(name = "titulo") String newTitle,
61         @WebParam(name = "enunciado") String newEnunciado, @WebParam(name = "autenticacao") String autenticacao);
62
63     /**
64      * Método para vincular uma questão a um tema
65      *
66      * @author Thyago Tenório
67      * @param idQuestao String - ID da questão a ser vinculada
68      * @param idsContext String - IDs dos temas a serem vinculados
69      * @param autenticacao String - Código de autenticação para validar a
70      * segurança
71      *
72      * @return CodeDTO - Resultado da operação
73      */
74     public CodeDTO vincularTemaQuestao(@WebParam(name = "idQuestao") String idQuestao, @WebParam(name = "idsContext") String idsContext,
75         @WebParam(name = "autenticacao") String autenticacao);
76
77     /**
78      * Método para vincular uma questão a um tema
79      *
80      * @author Thyago Tenório
81      * @param idQuestao String - ID da questão a ser desvinculada
82      * @param idContext String - ID do temas a ser desvinculado

```

```

83 * @param autenticacao String - Código de autenticação para validar a
84 * segurança
85 *
86 * @return CodeDTO - Resultado da operação
87 */
88 public CodeDTO desvincularTemaQuestao(@WebParam(name = "idQuestao") String idQuestao, @WebParam(name = "idContext") String idContext,
89     @WebParam(name = "autenticacao") String autenticacao);
90
91 /**
92 * Método para enviar um formulario de peer Assessment para uma questão
93 *
94 * @author Thyago Tenório
95 * @param idForm String - ID do formulário a ser associado
96 * @param idQuestao String - Id da questão a ser associada
97 * @param autenticacao String - Código de autenticação para validar a
98 * segurança
99 *
100 * @return CodeDTO - código indicando o resultado da operação
101 */
102 public CodeDTO setarFormularioParaQuestao(@WebParam(name = "idForm") String idForm,
103     @WebParam(name = "idQuestao") String idQuestao,
104     @WebParam(name = "autenticacao") String autenticacao);
105
106 /**
107 * Método para setar uma data inicial e uma data final para uma questão
108 *
109 * @author Thyago Tenório
110 * @param startDate String - Data inicial da questão
111 * @param idQuestao String - Id da questão a ser associada com as datas
112 * @param endDate String - Data final da questão
113 * @param autenticacao String - Código de autenticação para validar a
114 * segurança
115 *
116 * @return CodeDTO - código indicando o resultado da operação
117 */
118 public CodeDTO setarDataParaQuestao(@WebParam(name = "idQuestao") String idQuestao,
119     @WebParam(name = "startDate") String startDate,
120     @WebParam(name = "endDate") String endDate,
121     @WebParam(name = "autenticacao") String autenticacao);
122
123 /**
124 * Método para recuperar uma questão dado uma determinada data. Usado para
125 * recuperar a questão da semana
126 *
127 * @author Thyago Tenório
128 * @param date String - Data inicial da questão
129 * @param autenticacao String - Código de autenticação para validar a
130 * segurança
131 *
132 * @return WritingEvaluationDTO - redação da semana
133 * @throws InvalidAuthenticationException - Caso a autenticação esteja
134 * inválida
135 */
136 public WritingEvaluationDTO recuperarQuestaoPorData(@WebParam(name = "date") String date,
137     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
138
139 /**
140 * Método para recuperar um formulario de uma questão
141 *
142 * @author Thyago Tenório
143 * @param idQuestao String - Id da questão a ser recuperada
144 * @param autenticacao String - Código de autenticação para validar a
145 * segurança
146 *
147 * @return CodeDTO - código indicando o resultado da operação
148 */
149 public AssessmentFormDTO recuperarFormularioDeQuestao(@WebParam(name = "idQuestao") String idQuestao,
150     @WebParam(name = "autenticacao") String autenticacao);
151
152 /**
153 * Método para recuperar possíveis questões para um determinado usuário
154 * corrigir
155 *
156 * @author Thyago Tenório
157 * @param login String - Login do usuário no qual deseja recuperar as
158 * questões
159 * @param autenticacao String - Código de autenticação para validar a
160 * segurança
161 *
162 * @return ArrayList<WritingEvaluationSolvingDTO> - Questões possíveis para
163 * um determinado usuário corrigir
164 * @throws InvalidAuthenticationException - Caso a autenticação esteja
165 * inválida
166 */

```

```

167 public ArrayList<WritingEvaluationSolvingDTO> recuperarQuestoesParaCorrigir(@WebParam(name = "login") String login,
168     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
169
170 /**
171  * Método para recuperar todos os temas das questões
172  *
173  * @author Thyago Tenório
174  * @param autenticacao String - Código de autenticação para validar a
175  * segurança
176  *
177  * @return ArrayList<ContextDTO> - Temas das questões
178  *
179  * @throws InvalidAuthenticationException - Caso a autenticação esteja
180  * inválida
181  */
182 public ArrayList<ContextDTO> recuperarTemasQuestoes(@WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
183
184 /**
185  * Método para recuperar um tema pelo ID
186  *
187  * @author Thyago Tenório
188  * @param idContext String - ID do tema a ser recuperado
189  * @param autenticacao String - Código de autenticação para validar a
190  * segurança
191  *
192  * @return ContextDTO - Temas da questão solicitada
193  */
194 public ContextDTO recuperarTema(@WebParam(name = "idContext") String idContext, @WebParam(name = "autenticacao") String autenticacao);
195
196 /**
197  * Método para editar um tema de uma determinada questão
198  *
199  * @author Thyago Tenório
200  * @param idContext String - ID do tema a ser editado
201  * @param tema String - Novo tema que irá substituir o tema antigo
202  * @param autenticacao String - Código de autenticação para validar a
203  * segurança
204  *
205  * @return CodeDTO - Resultado da operação
206  */
207 public CodeDTO editarTema(@WebParam(name = "idContext") String idContext, @WebParam(name = "tema") String tema,
208     @WebParam(name = "autenticacao") String autenticacao);
209
210 /**
211  * Método para criar um novo tema de questão
212  *
213  * @author Thyago Tenório
214  * @param tema String - tema da questão
215  * @param autenticacao String - Código de autenticação para validar a
216  * segurança
217  *
218  * @return ContextDTO - Tema da questão criado
219  */
220 public ContextDTO criarTemaQuestao(@WebParam(name = "tema") String tema, @WebParam(name = "autenticacao") String autenticacao);
221
222 /**
223  * Método para recuperar todas as questões dado um tema
224  *
225  * @author Thyago Tenório
226  * @param idContext String - ID do Tema da questão
227  * @param autenticacao String - Código de autenticação para validar a
228  * segurança
229  *
230  * @return ArrayList<WritingEvaluationDTO> - Lista contendo todas as
231  * questões com o tema especificado
232  *
233  * @throws InvalidAuthenticationException - Caso a autenticação esteja
234  * inválida
235  */
236 public ArrayList<WritingEvaluationDTO> recuperarQuestoesPorTema(
237     @WebParam(name = "idContext") String idContext,
238     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
239
240 /**
241  * Método para recuperar uma determinada questão no sistema
242  *
243  * @author Thyago Tenório
244  * @param idQuestao String - Id da questão a ser recuperada
245  * @param autenticacao String - Código de autenticação para validar a
246  * segurança
247  *
248  * @return DiscursiveProblemDTO - A questão especificada
249  */
250 public WritingEvaluationDTO recuperarQuestao(@WebParam(name = "idQuestao") String idQuestao,

```

```

251     @WebParam(name = "autenticacao") String autenticacao);
252
253 /**
254  * Método para recuperar uma determinada questão no sistema de um curriculum
255  * dado um determinado usuário
256  *
257  * @author Thyago Tenório
258  * @param login String - Id do usuário que irá receber a questão
259  * @param idContext String - Id do tema no qual se deseja recuperar uma
260  * questão para o usuário
261  * @param autenticacao String - Código de autenticação para validar a
262  * segurança
263  *
264  * @return DiscursiveProblemDTO - A questão especificada
265  *
266  * @throws InvalidAuthenticationException - Caso a autenticação esteja
267  * inválida
268  */
269 public WritingEvaluationDTO enviarQuestãoParaUsuarioEmTema(@WebParam(name = "login") String login,
270     @WebParam(name = "idContext") String idContext,
271     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
272
273 /**
274  * Método para um aluno responder uma determinada questão
275  *
276  * @author Thyago Tenório
277  * @param idQuestao String - Id da questão que foi respondida
278  * @param resposta String - Resposta da questão
279  * @param login String - Id do usuário que respondeu a questão
280  * @param autenticacao String - Código de autenticação para validar a
281  * segurança
282  *
283  * @return DiscursiveProblemDTO - A questão especificada
284  */
285 public CodeDTO responderQuestão(@WebParam(name = "idQuestao") String idQuestao,
286     @WebParam(name = "resposta") String resposta,
287     @WebParam(name = "login") String login,
288     @WebParam(name = "autenticacao") String autenticacao);
289
290 /**
291  * Método para recuperar as respostas de todas as questões de um determinado
292  * aluno
293  *
294  * @author Thyago Tenório
295  * @param login String - Id do usuário que se deseja as correções
296  * @param autenticacao String - Código de autenticação para validar a
297  * segurança
298  *
299  * @return ArrayList<DiscursiveProblemSolvingDTO> - Todas as correções de
300  * questões de um determinado usuário
301  * @throws InvalidAuthenticationException - Caso a autenticação esteja
302  * inválida
303  */
304 public ArrayList<WritingEvaluationSolvingDTO> recuperarQuestoesFeitasPorUsuario(
305     @WebParam(name = "login") String login,
306     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
307
308 /**
309  * Método para recuperar uma resposta de uma questão de um determinado aluno
310  *
311  * @author Thyago Tenório
312  * @param idEvaluationSolving String - Id do evaluation Solving do aluno
313  * @param autenticacao String - Código de autenticação para validar a
314  * segurança
315  *
316  * @return DiscursiveProblemSolvingDTO - Uma correção de uma questão de um
317  * determinado usuário
318  */
319 public WritingEvaluationSolvingDTO recuperarQuestaoFeitaPorUsuario(
320     @WebParam(name = "idEvaluationSolving") String idEvaluationSolving,
321     @WebParam(name = "autenticacao") String autenticacao);
322
323 /**
324  * Método para checar se um aluno pode responder ou não uma questão
325  *
326  * @author Thyago Tenório
327  * @param login String - Login do usuário
328  * @param autenticacao String - Código de autenticação para validar a
329  * segurança
330  *
331  * @return Boolean - indicando se o usuário está permitido (true) ou não
332  * (false) a fazer uma questão
333  *
334  * @throws InvalidAuthenticationException - Caso a autenticação esteja

```

```

335 * inválida
336 */
337 public Boolean verificarPermissaoResponderQuestao(@WebParam(name = "login") String login,
338     @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
339
340 /**
341 * Método para checar se um aluno respondeu uma determinada questão
342 *
343 * @author Thyago Tenório
344 * @param login String - Login do usuário
345 * @param idQuestao String - ID da questão a ser verificada
346 * @param autenticacao String - Código de autenticação para validar a
347 * segurança
348 *
349 * @return ArrayList<WritingEvaluationSolvingDTO> - lista contendo as
350 * resoluções da questão que ele já fez
351 * @throws InvalidAuthenticationException - Caso a autenticação esteja
352 * inválida
353 */
354 public ArrayList<WritingEvaluationSolvingDTO> verificarUsuarioRespondeuQuestao(@WebParam(name = "login") String login,
355     @WebParam(name = "idQuestao") String idQuestao, @WebParam(name = "autenticacao") String autenticacao) throws InvalidAuthenticationException;
356
357 /**
358 * Método para recuperar o resultado final de uma questao
359 *
360 * @author Thyago Tenório
361 * @param login String - Id da questão que foi respondida
362 * @param idEvaluationSolving String - Id da questão que foi respondida
363 * @param autenticacao String - Código de autenticação para validar a
364 * segurança
365 *
366 * @return WritingReportDTO - Relatório final daquela questão
367 */
368 public WritingReportDTO recuperarResultadoQuestao(@WebParam(name = "login") String login,
369     @WebParam(name = "idEvaluationSolving") String idEvaluationSolving,
370     @WebParam(name = "autenticacao") String autenticacao);
371
372 }
373

```

APÊNDICE D – PROPOSTA DE REDAÇÃO CRIADA PELO ESPECIALISTA

PROPOSTA DE REDAÇÃO

O futuro da inteligência artificial

Você pode não perceber, mas você já é rodeado de inteligência artificial. A inteligência artificial é um ramo de pesquisa da ciência da computação que busca, através de símbolos computacionais, construir mecanismos e/ou dispositivos que simulem a capacidade do ser humano de pensar, resolver problemas, ou seja, de ser inteligente. O estudo e desenvolvimento desse ramo de pesquisa tiveram início na Segunda Guerra Mundial e vem evoluindo significativamente ao longo dos últimos anos. O desejo de construir máquinas capazes de reproduzir a capacidade humana de pensar e agir vem de muitos anos, contudo há um grande receio em que os computadores podem se tornar. Para muitos a existência de máquinas com o poder de pensar, sentir e até ter a capacidade de realizar atividades humanas é um fato inconcebível. Considerando os textos abaixo, bem como seus próprios conhecimentos sobre o assunto, produza um texto dissertativo-argumentativo apresentando qual, em seu ponto de vista, será o futuro da inteligência artificial.



Nova inteligência artificial do Facebook aprende e responde perguntas

O Facebook quer tornar seus algoritmos ainda mais eficientes para que seus sistemas de IA consigam ser capazes de não só entender sobre o que as pessoas falam na internet, mas aprender e responder perguntas.

As aplicações podem variar, permitindo uma melhor ferramenta de buscas para a plataforma, mas também pode transformar o Facebook em um assistente pessoal útil, como são hoje Google Now, Cortana e Siri.

O esforço já é antigo e vem desde 2013, quando a rede social criou um grupo de pesquisa dedicado ao aprendizado profundo. Mais do que aprender texto, a rede social quer conseguir identificar melhor o que as pessoas tentam comunicar em fotos e vídeos. Schroepfer diz que o algoritmo já é bom o suficiente para diferenciar patinação, patinação de velocidade, patinação artística e hockey no gelo.

[Olhar Digital]

Elon Musk Sobre Robôs Inteligentes: Teremos Sorte Se Eles Não Nos Escravizarem

Não é segredo algum que o intelectual Elon Musk teme o que pode acontecer se robôs superinteligentes passarem a se atualizar – ele disse certa vez que estamos “invocando o demônio” com pesquisas em inteligência artificial – mas neste domingo Musk falou em detalhes sobre como crê que a IA poderia dar fim à humanidade.

“Estou muito preocupado com superinteligência artificial atualmente. Acho que é algo que talvez seja mais perigoso que armas nucleares”, disse Musk. “Deveríamos ter muito cuidado com isso. Se existisse uma superinteligência digital que pudesse ativar uma espécie de auto-melhoria recursiva de forma não-logarítmica, que pudesse se reprogramar para ser mais inteligente e fazer iterações mais rápidas e fizesse isso 24 horas por dia em milhões de computadores, seria o fim.”

Se a inteligência artificial irá nos matar, esta é a forma prevista geralmente em filmes de ficção científica e gente que pensa nesse tipo de coisa: a inteligência se atualizará e se ensinará a como se tornar muito mais esperta que nós, e então será incontrolável

[Motherboard]

Stephen Hawking adverte inteligência artificial poderia terminar a humanidade

Prof Stephen Hawking, um dos cientistas proeminentes da Grã-Bretanha, disse que os esforços para criar máquinas pensantes representam uma ameaça para a nossa própria existência.

Ele disse à BBC: "O desenvolvimento da inteligência artificial plena poderia significar o fim da raça humana."

Sua advertência veio em resposta a uma pergunta sobre uma reformulação da tecnologia que ele usa para se comunicar, o que envolve uma forma básica de IA.

Prof Hawking diz que as formas primitivas de inteligência artificial desenvolvidas até agora já provaram ser muito úteis, mas ele teme as consequências da criação de algo que pode igualar ou superar os seres humanos.

[BBC]

Figura 78 – Proposta de redação criada para a unidade de experimento

Fonte: elaborado pelo autor

APÊNDICE E – FORMULÁRIO DE AVALIAÇÃO PADRÃO PARA REDAÇÕES DO ENEM

No formulário de avaliação do ENEM, os professores avaliam o desempenho do aluno nas redações de acordo com cinco critérios (competências) de avaliação pré-definidos, conforme especificados abaixo:

- **Competência 1:** Demonstrar domínio da modalidade escrita formal da Língua Portuguesa
- **Competência 2:** Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa
- **Competência 3:** Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.
- **Competência 4:** Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.
- **Competência 5:** Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.

Cada avaliador atribuirá uma nota entre 0 (zero) e 200 (duzentos) pontos para cada uma das cinco competências, e a soma desses pontos comporá a nota total de cada avaliador, que pode chegar a 1000 (mil) pontos.

Além disso, existem algumas condições (critérios) no qual a redação receberá a nota 0 (zero). Esses critérios são:

- Fuga total ao tema;
- Não obediência à estrutura dissertativo-argumentativa;
- Texto com até 7 (sete) linhas;
- Impropérios, desenhos e outras formas propositais de anulação ou parte do texto deliberadamente desconectada do tema proposto;
- Desrespeito aos direitos humanos; e
- Folha de redação em branco, mesmo que haja texto escrito na folha de rascunho.

Nesse sentido, construímos um formulário que atende aos requisitos elicitados acima (os parâmetros de correção do ENEM). A Figura 79 apresenta o formulário de redação criado.

Formulário de Avaliação de Redação ENEM

Critérios de eliminação

- Fuga total ao tema
- Não obediência à estrutura dissertativo-argumentativa
- Desrespeito aos direitos humanos
- Improprios, desenhos e outras formas propositais de anulação ou parte do texto deliberadamente desconectada do tema proposto

Notas por área

Competências	Notas	Descrição
Demonstrar domínio da modalidade escrita formal da Língua Portuguesa	 0	
Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa	 0	
Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista	 0	
Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação	 0	
Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos	 0	

Enviar correção

Figura 79 – Formulário de avaliação criado com base nos parâmetros de correção do ENEM

Fonte: elaborado pelo autor

Podemos observar na figura que alguns critérios de eliminação foram retirados do formulário. Isso acontece porque os critérios já podem ser avaliados de forma automática pelo próprio sistema, como por exemplo, o critério de texto com menos de 7 linhas e folha de redação em branco. Outro fator que podemos destacar no formulário é que a variação encontrada no ENEM de 0 a 200 é mantida, porém utilizamos um intervalo de 40 pontos, ou seja, os possíveis valores são [0,40,80,120,160,200].

APÊNDICE F – ARTIGOS PUBLICADOS/PREVISTOS AO LONGO DO MESTRADO

Alguns artigos que foram desenvolvidos como os resultados desse trabalho foram:

1. A gamified peer assessment model for written evaluations in on-line learning environments. Submetido a *Computers and education*
2. A systematic review of use of peer assessment model in on-line learning environments. Em processo de revisão. Após, submissão para *Computers and education*

Outros artigos publicados ao longo do mestrado:

1. HOLANDA, OLAVO ; ISOTANI, SEIJI ; BITTENCOURT, IG IBERT ; ELIAS, ENDHE ; TENÓRIO, THYAGO . JOINT: Java ontology integrated toolkit. *Expert Systems with Applications*, v. 40, p. 6469-6477, 2013.
2. DERMEVAL, DIEGO ; CASTRO, JAELSON ; SILVA, CARLA ; PIMENTEL, JOÃO ; BITTENCOURT, IG IBERT ; BRITO, PATRICK ; ELIAS, ENDHE ; TENÓRIO, THYAGO ; PEDRO, ALAN . On the use of metamodeling for relating requirements and architectural design decisions. In: the 28th Annual ACM Symposium, 2013, Coimbra. *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13*. New York: ACM Press. p. 1278.
3. TENÓRIO, THYAGO ; DERMEVAL, D. ; BITTENCOURT, I. I. . On the Use of Ontology for Dynamic Reconfiguring Software Product Line Products. In: The Ninth International Conference on Software Engineering Advances (ICSEA), 2014, Nice - França. *Proceedings of the Ninth International Conference on Software Engineering Advances*, 2014. p. 545-550.
4. DERMEVAL, DIEGO ; TENÓRIO, THYAGO ; BITTENCOURT, IG IBERT ; SILVA, ALAN ; ISOTANI, SEIJI ; RIBEIRO, MÁRCIO . Ontology-based Feature Modeling: An Empirical Study in Changing Scenarios. *Expert Systems with Applications*, p. 4950-4964, 2015.

Para mais informações a respeito dos projetos envolvidos, apresentações e diversas outras informações a cerca desses e outros artigos, acesse o curriculum lattes¹.

¹ Disponível pelo endereço <http://lattes.cnpq.br/3202277585004694>