

Avaliações em Jogos Educacionais: instrumentos de avaliação da reação, aprendizagem e comparação de jogos

Rháleff N. R. Oliveira¹, Guilherme D. Belarmino^{1,2}, Carla Rodriguez^{1,2}, Denise Goya^{1,2}, Mirtha Venero^{1,2}, Ailton P. Oliveira Junior², Rafaela V. Rocha^{1,2}

¹ Pós-graduação em Ciência da Computação, UFABC – Santo André – SP

² Centro de Matemática, Computação e Cognição, UFABC – Santo André – SP

{rhaleff.nascimento, g.dias, c.rodriguez, denise.goya, mirtha.lina, ailton.junior, rafaela.rocha}@ufabc.edu.br

Abstract. *Very often the assessments and evaluations are neglected in serious games. Therefore, this article presents and uses an adapted methodology, protocol (learning assessment and game evaluation) and data collection instruments (profile, reaction, learning and games comparison). The methodology was based on a model of the Training, Development and Education area, and the artifacts were applied in the evaluation of an educational game. Results indicate the benefits of using artifacts, among them the possibility to evaluate the reaction and learning with the game, as well as the comparison between the two versions / educational games. In addition, this work can serve as a basis for serious game evaluations.*

Resumo. *Muitas vezes as avaliações são negligenciadas em jogos educacionais. Por isso, este artigo propõe e usa uma metodologia adaptada, um protocolo de avaliação do jogo e aprendizagem e instrumentos de coleta de dados (perfil, reação, aprendizagem e comparação de jogos). A metodologia foi baseada em um modelo da área de Treinamento, Desenvolvimento e Educação, e os artefatos criados foram utilizados na avaliação de um jogo educacional. Os resultados apontam indícios dos benefícios de uso desses artefatos, entre eles a possibilidade de avaliar a reação e o aprendizado obtido com o jogo, bem como, a comparação entre duas versões / jogos educacionais. Além disso, esse trabalho pode servir de base para realização de avaliações em jogos sérios.*

1. Introdução

A avaliação é um tema recorrente na área de jogos educacionais digitais (JEDs), seja para avaliar a reação e aprendizagem do aluno ou para avaliar o jogo em si (como usabilidade, acessibilidade, jogabilidade) (Rocha *et al.*, 2015; Savi *et al.*, 2010). Entretanto, a avaliação da aprendizagem em JEDs é muitas vezes negligenciada, motivada pela falta de métodos e recursos consolidados (Oliveira *et al.*, 2018; Pereira Junior e Menezes, 2015). Além de que a literatura carece de relatos de comparação de jogos sérios com os mesmos objetivos (com a proposta de protocolos e instrumentos validados) (Salas *et al.*, 2009; Drummond *et al.*, 2017; Rocha, 2017).

Em relação à avaliação da aprendizagem, a avaliação visa estabelecer uma comparação entre os resultados desejados e os resultados obtidos para reforçar, orientar e/ou corrigir o comportamento do avaliado em suas atividades (Pontes, 1999). Há diferentes tipos de avaliação da aprendizagem, de acordo com o objetivo da mesma:

diagnóstica, formativa, somativa, autoavaliação. A *avaliação diagnóstica* é realizada no início do processo de ensino-aprendizagem para analisar a presença de pré-requisitos e colocar o aluno em um novo contexto de aprendizado (Miller, Imrie e Cox, 1998). A *avaliação formativa* é realizada durante o processo para desenvolver e aprimorar as competências do aprendiz (Perrenoud, 1999). A *somativa* ocorre após um período de aprendizagem, e visa mensurar a progressão do aluno (Miller, Imrie e Cox, 1998). Na *autoavaliação*, o aluno reflete sobre a sua própria evolução ao final do processo. A avaliação pode incluir ou não medidas de desempenho (dados quantitativos) somados a um julgamento de valor (dados qualitativos) (Salas *et al.*, 2009).

Em relação à avaliação do processo de ensino-aprendizagem, o modelo de avaliação da eficácia do programa de treinamento proposto por Kirkpatrick e Kirkpatrick (2006) é estruturado em quatro níveis para verificar se os objetivos foram alcançados: (1) reação, (2) aprendizagem, (3) comportamento e (4) resultados. Os dois primeiros baseados nos efeitos imediatos e os outros nos efeitos em longo prazo.

No contexto de JEDs, a maioria das avaliações realizadas ocorre no nível da reação, com o objetivo de coletar as percepções dos alunos em relação ao jogo, como jogabilidade, satisfação, motivação, entre outros (Savi *et al.*, 2010). Na literatura, há poucos relatos da avaliação da aprendizagem (nível 2) em si, o que há é a avaliação da percepção da aprendizagem, o que está no nível de reação (Oliveira *et al.*, 2018; Rocha, 2014). Para avaliar a aprendizagem, questionários de pré e pós-testes devem ser usados para mensurar o aumento do conhecimento).

Sendo assim, este trabalho objetiva propor e usar artefatos para avaliação da satisfação e da aprendizagem com JEDs. Isso inclui instrumentos para coleta de dados em relação ao perfil, reação, aprendizagem e comparação de jogos educacionais e um protocolo de avaliação, que integra esses aspectos, para ser executado com o público-alvo. Os instrumentos foram desenvolvidos por meio de uma metodologia adaptada pelos autores, baseada nas etapas do processo de construção de medidas de avaliação em Treinamento, Desenvolvimento e Educação, apresentado por Mourão e Menezes (2012). Os artefatos foram validados por três especialistas em desenvolvimento e avaliação de jogos e aplicados para a avaliação de um JED, nomeado Expedição Antártica, com 14 jogadores.

Este trabalho está organizado: na seção 2 são apresentados os trabalhos relacionados. A seção 3 descreve a metodologia, com as principais decisões tomadas no planejamento deste trabalho. Na seção 4 é apresentado o processo de criação do protocolo, tal como os seus instrumentos (questionários de perfil, avaliação da reação, aprendizagem e comparação). Na seção 5 são descritos os resultados e discussões, seguidos das considerações finais, na seção 6.

2. Trabalhos Relacionados

Nesta seção são apresentados trabalhos que apresentam modelos e instrumentos de coleta de dados usado para avaliação em jogos educacionais e sérios.

Fu *et al.* (2009) apresenta o *EGameFlow*, um método para avaliação de jogo educacional, com 42 itens divididos em oito dimensões: imersão, interação social, desafio, clareza de objetivos, *feedback*, concentração, controle e melhoria do conhecimento. Os itens são apresentados em escala *Likert* de 7 pontos. O método foi

validado observando a relevância dos itens, adequação das dimensões, ordem dos itens e gramática, e foi aplicado com quatro jogos diferentes.

Oliveira (2014) propõe um questionário de avaliação da qualidade de sistemas de *software* educacionais para o ensino-aprendizagem do tema Gerência de Projetos de *Software*. O questionário contém 32 questões em escala *Likert*, divididas em três categorias: (1) características específicas do tema de Gerência de Projetos de *Software* (Conceitos/Conteúdo); (2) aspectos lúdicos (Motivação); e (3) aprendizado focado nos usuários (Alunos). O questionário foi validado por uma técnica de avaliação apenas (não por especialistas), no entanto, não foi aplicado na avaliação de um jogo.

Rocha *et al.* (2015) apresentam um questionário para autoavaliação e avaliação da reação do aprendiz, com 48 itens, para coletar dados quanto aos aspectos de simulação, jogo e suas fases, aprendizagem e treinamento. O questionário se divide em quatro tópicos sobre reação do usuário a (1) simulação; (2) aprendizagem com o jogo; (3) jogo; e (4) aprendizagem durante as fases. O questionário foi validado por especialistas e aplicado após a interação de oito pessoas com um jogo sério, que visa treinar e avaliar o uso do protocolo de controle de vazamento de gás de cozinha.

Pires *et al.* (2015) apresentam um questionário para avaliação de jogos sérios dividido em quatro dimensões: (1) perfil do jogador; (2) aprendizagem (por associação, percepção, autoavaliação, apreensão de conceitos, interação e discussão em grupo/participação ativa); (3) emoções e atitudes (vontade e prazer na conquista, motivação para estudar, ambiguidade nos sentimentos de prazer e tensão, vontade de abandonar a partida, frustração); e (4) táticas do jogo (entendimento das regras, qualidade das cartelas, clareza do tabuleiro, reinvenção no jogo). O questionário foi validado por especialistas e a análise se dá pelo coeficiente *Kappa*, sendo usado para avaliar um jogo que discute a construção e o financiamento das redes de serviços de saúde no SUS.

Petri *et al.* (2018) apresenta o *MEEGA+* (*Model for the Evaluation of Educational Games*), um modelo que define instrumentos de medição para avaliar a qualidade de jogos (digitais ou não-digitais) em termos de experiência do jogador e percepção da aprendizagem do ponto de vista do aluno e instrutor. O questionário tem 36 itens divididos nos tópicos: atenção focada, diversão, desafio, interação social, confiança, relevância, satisfação, usabilidade e percepção da aprendizagem. Ele foi validado por especialistas, porém, não foi usado na avaliação de jogos.

De forma geral, os trabalhos analisados não apresentam um questionário de comparação de jogos. Ademais, nos trabalhos em que os artefatos foram aplicados para avaliar algum jogo, não apresentam a descrição de um protocolo para nortear a execução da avaliação, apenas o questionário é aplicado após o usuário jogar. Sendo que nenhum deles usa a técnica de pré e pós-teste para avaliar a aprendizagem, ou seja, os artefatos analisados se encontram no nível de reação. Neste contexto, este trabalho apresenta questionários para coletar dados dos alunos nos níveis de reação e aprendizagem, além de um questionário para comparação de jogos.

3. Metodologia

Para o planejamento, desenvolvimento, validação e execução da avaliação com os instrumentos propostos, foram realizadas as etapas descritas a seguir, baseadas no processo de construção de medidas de avaliação em TD&E (Mourão e Menezes, 2012) e adaptadas para as necessidades deste trabalho: (i) definir o que se pretende medir; (ii)

definir o público; (iii) escolher técnicas de coleta de dados; (iv) definir dimensões avaliativas; (v) definir o tipo de escala de julgamento; (vi) construir e validar os instrumentos de coleta; (vii) aplicar os instrumentos; e (viii) analisar os resultados.

(i) definir o que se pretende medir: foi definido o objetivo de avaliar a jogabilidade e as contribuições ao processo de ensino-aprendizagem e o processo de avaliação de jogos sérios e/ou educacionais, nos níveis de reação e aprendizagem; além da comparação de duas versões/jogos com os mesmos objetivos. O objeto de avaliação foi o JED nomeado "Expedição Antártica", com duas versões de uma missão.

(ii) definir o público: os participantes/avaliadores devem ser definidos conforme o público-alvo do jogo. Por exemplo, no caso da avaliação do jogo selecionado, foram definidos como avaliadores alunos da disciplina de programação de jogos, escolhidos por conveniência, pela familiaridade em desenvolvimento de jogos e por serem graduandos (no caso, o público-alvo do jogo é de alunos de licenciatura).

(iii) escolher técnicas de coleta de dados: foram adotadas (1) comunicação por meio de questionários (incluindo pré e pós-testes) e (2) observação humana (observação e registro de informações, pela equipe de desenvolvimento, padronizado por roteiro). Essas técnicas foram escolhidas por serem rápidas, fáceis e de baixo custo (construção e análise dos resultados). Os questionários são apresentados na seção 4.

(iv) definir dimensões avaliativas: foram definidos que os questionários deveriam ser divididos e organizados em categorias avaliativas (em níveis de reação e aprendizagem, de Kirkpatrick). Sendo assim, foi elaborado um protocolo de avaliação, disponível em <http://bit.ly/ProtocoloTestagem>, que consiste em um conjunto de etapas e atividades para orientar a equipe de desenvolvimento durante a execução da avaliação do JED. O protocolo organizou os questionários em: (i) questionário de perfil do jogador; (ii) questionários de pré e pós-testes (nível de aprendizagem); (iii) questionário de reação ao jogo (nível de reação); e (iv) questionário comparativo.

(v) definir o tipo de escala de julgamento: foi definida que algumas questões dos questionários deveriam usar a escala do tipo *Likert*, de 1 (discordo fortemente) a 5 (concordo fortemente), utilizada neste trabalho para mensurar as afirmações em grau de concordância, e a escala de medida ordinal, de 1 (menor pontuação) a 7 (maior pontuação), para os questionários com menor quantidade de questões, além de questões abertas e de múltipla escolha. Tais decisões foram baseadas no estudo de Vieira e Dalmoro (2008), no qual afirmam que a quantidade de questões, em relação aos valores da escala, pode interferir nos resultados (por exemplo, se houver muitas questões e for selecionada a escala de 7 pontos, haverá uma quantidade maior de escolhas).

(vi) construir e validar os instrumentos de coleta: os questionários foram criados com o Formulário do Google Docs®. A validação das assertivas foi realizada por três especialistas em desenvolvimento e avaliação de jogos sérios e tecnologia educacional (professores doutores). Todos os questionários criados foram submetidos à análise semântica, que consiste em identificar e corrigir falhas, de modo que os itens sejam precisos, claros e objetivos (Zerbini *et al.*, 2012). Para isso, os avaliadores, por meio de uma entrevista não estruturada, analisaram de maneira crítica cada item dos questionários. As sugestões de ajustes foram acatadas para melhoria dos instrumentos.

(vii) aplicar os instrumentos: os questionários foram aplicados de acordo com o protocolo descrito na seção 4.

(viii) **analisar os resultados:** os dados coletados foram analisados e realizadas discussões acerca dos instrumentos de coleta, descritos na seção 5.

4. Instrumentos de avaliação da reação, aprendizagem e comparação de JED

Nesta seção são apresentados os instrumentos criados para serem aplicados na avaliação da reação, aprendizagem do jogador, e comparação de versões/jogos com objetivos similares. Para este trabalho, os instrumentos contam com questões específicas para avaliação do JED nomeado "Expedição Antártica", porém pode ser adaptado para avaliação de outros jogos sérios ou educacionais, e em outros contextos.

4.1. Protocolo de Avaliação

O protocolo de avaliação tem como objetivo guiar os aplicadores na execução da avaliação com o público-alvo. Como há duas versões/jogos, os jogadores devem ser divididos em dois grupos. Em seguida, deve ser entregue um roteiro para cada um com uma breve introdução do que se trata a avaliação, seu objetivo e instruções do jogo. O protocolo é composto por: questionário de perfil, pré-teste, 1ª versão/jogo, questionário de reação, pós-teste, 2ª versão/jogo e questionário de comparação (os instrumentos são descritos nas próximas subseções). Para a realização da avaliação de modo anônimo, o roteiro deve ser entregue aos jogadores com um identificador alfanumérico, distinguindo grupos e indivíduos (por exemplo, formado pela letra do grupo e um número, tal como A0101 e B0101). O jogador deve inserir seu identificador em cada questionário. As versões do jogo (ou jogos similares) devem ser selecionados antecipadamente, e estarem disponíveis para uso, sendo que as instruções para acesso devem estar explícitas no roteiro. O protocolo descreve o papel do observador durante os testes, a saber: anotar o ID do jogador, em que parte do jogo ele estava com dúvida, a questão levantada pelo jogador e a resposta fornecida e registrar observações gerais.

4.2. Questionário de Perfil do Jogador

O questionário de perfil do jogador objetivou coletar informações em relação ao curso e a experiência com JEDs. Sendo assim, o questionário possui seis questões, a saber: (1) idade; (2) curso; (3) "*Qual sua experiência com desenvolvimento de jogos?*", em uma escala de 1 (nenhuma experiência) a 5 (muita experiência); (4) "*Você já jogou algum jogo educacional?* "; (5) "*Você já desenvolveu algum jogo educacional?*"; e (6) "*Se sua resposta foi SIM, faça uma breve descrição do jogo educacional que você desenvolveu*".

4.3. Questionário de Pré e Pós-Testes

O questionário de pré e pós-testes objetivou coletar informações em relação ao nível de aprendizagem de Kirkpatrick, no que diz respeito ao desempenho do jogador ao usar o jogo. O pré-teste foi aplicado antes e o pós-teste depois do uso do jogo. As perguntas dos questionários devem estar relacionadas às temáticas principais do jogo. No caso deste trabalho, o objetivo do JED é ensinar o processo de pesquisa científica, por meio da participação em um projeto de ciência cidadã para identificação de baleias na Antártica. Sendo assim, o questionário possui três questões abertas: (1) "*O que é um projeto de ciência cidadã? Cite exemplo(s)*"; (2) "*Descreva os passos do processo de pesquisa científica*"; e (3) "*Descreva os passos para cadastrar uma foto de baleia em um projeto*".

de ciência cidadã”. Para aplicação em outros jogos, os objetivos do jogo devem ser analisados, e as questões devem ser formuladas de acordo com eles, sendo que podem ser questões idênticas ou similares (avaliação dos mesmos conteúdos).

4.4. Questionário de Reação ao Jogo

O questionário de reação ao jogo objetivou coletar informações em relação ao nível de reação de Kirkpatrick, no que diz respeito à percepção da (1) aprendizagem, (2) jogabilidade, (3) interface, (4) imersão e diversão e (5) facilidade e confiança ao usar o jogo. Sendo assim, foram elaboradas 16 questões, sendo 8 com escala de 1 (discordo fortemente) a 5 (concordo fortemente), 7 questões abertas e 1 do tipo caixa de seleção. Os oito itens que usaram a escala são: **(R1)** “*Eu estou confiante com o aprendizado que obtive no jogo*”; **(R2)** “*Depois do jogo, eu consigo compreender o processo de fotoidentificação de uma baleia*”; **(R3)** “*Houve momentos em que eu queria desistir do jogo*”; **(R4)** “*Eu alcancei rapidamente os objetivos do jogo*”; **(R5)** “*Eu jogaria novas missões desse jogo*”; **(R6)** “*Fiquei torcendo pelo jogo acabar logo*”; **(R7)** “*O jogo foi mais difícil de entender do que eu gostaria*”; e **(R8)** “*O design da interface do jogo é atraente*”. Os itens R1, R2, R4, R5 e R8 são considerados questões positivas, ou seja, os maiores valores, dentro da escala *Likert*, são vistos de forma favorável ao proposto. Já os itens R3, R6 e R7 são consideradas questões negativas e, portanto, foram normalizadas antes da análise. Neste trabalho, apenas a questão sobre a aprendizagem (**R2**) é específica, devendo ser alterada conforme o objetivo do jogo avaliado.

4.5. Questionário de Comparação das Versões do Jogo

Conforme descrito no protocolo, foi aplicado um questionário de comparação com objetivo de analisar qual versão é mais adequada aos objetivos propostos, em relação à jogabilidade, aprendizagem, motivação, diversão e satisfação. O questionário de comparação buscou coletar a opinião (críticas e sugestões) dos alunos em comparação às duas versões, por meio de sete questões, sendo quatro abertas, uma de múltipla escolha e duas em escala *Likert*. O questionário é dividido em duas partes, a saber: (1) questões sobre opiniões de forma geral de ambas versões, no que diz respeito a comentários, sugestões e críticas, e a opinião em relação à faixa etária do público-alvo do jogo; e (2) escala *Likert* de pontuação da versão A e B, separadamente, no que diz respeito ao **(C1)** Aprendizado, **(C2)** Jogabilidade, **(C3)** Desafio, **(C4)** Divertimento, **(C5)** Satisfação em jogar, **(C6)** Motivação para estudar mais sobre Ciência Cidadã e **(C7)** Motivação para estudar mais sobre Pesquisa Científica. Neste último, os critérios são pontuados em uma escala de 1 (menor pontuação) e 7 (maior pontuação). Da mesma forma, neste caso, as questões **(C6 e C7)** sobre a motivação para estudar são específicas, devendo ser alteradas conforme o objetivo do jogo avaliado.

5. Resultados e Discussões

Esta seção apresenta a descrição das duas versões do protótipo do jogo Expedição Antártica, e os resultados da aplicação dos questionários de perfil, reação, aprendizagem e comparação do mesmo. O protocolo foi aplicado com 14 alunos da disciplina de “Introdução à Programação de Jogos”, que foram selecionados por conveniência pela familiaridade em desenvolvimento de jogos e por serem graduandos. Além disso, quatro observadores participaram da avaliação, sendo dois alunos e duas professoras, que fazem parte da equipe de desenvolvimento do jogo. Foram realizadas análises qualitativas (para

questões abertas) e quantitativas dos dados coletados (por meio de medidas estatísticas de média e desvio padrão, quando cabíveis).

5.1. Aplicação com Duas Versões de um Protótipo do Jogo Expedição Antártica

Expedição Antártica é um JED inclusivo para o ensino interdisciplinar de ciências, matemática e educação científica, para licenciandos, com ênfase no ambiente antártico. É um jogo de gênero *RPG top-down* multiplataforma, no qual o jogador aventura-se pela Antártica em busca de coletar, analisar e enviar dados para pesquisas de Ciência Cidadã (CC). Os dois protótipos, desenvolvidos usando o motor de jogo *Unity3D*, são parte de uma das missões do jogo, e tem como objetivo conhecer o processo de fotoidentificação de baleias (observar, fotografar e identificar a cauda em um catálogo), ao participar de um projeto de CC. Sendo assim, nas duas versões, o jogador precisa navegar (em 8 direções) no cenário do navio, dialogar com mentores (5 na versão A e 2 na versão B) para obter informações sobre CC, para então ir ao desafio de fotoidentificação das baleias.

Nas duas versões, ao iniciar a missão, o jogador tem uma vista do mar, onde aparecem esporadicamente as caudas das baleias. Ao fotografar uma cauda, inicia-se o processo de fotoidentificação. Na versão A, o processo resume-se em comparar a foto tirada com um catálogo de caudas (vide **Figuras 1 A1 e A2**). Caso o jogador encontre essa baleia no catálogo, clica na foto correspondente onde é apresentado o nome da baleia identificada e, por fim, clica no botão *catalogar* (vide **Figura 1 A3**). Caso o jogador não encontre a cauda no catálogo, ele clica no botão + (significa que uma nova baleia foi encontrada e então deve nomeá-la antes de catalogar (vide **Figura 1 A4**)).

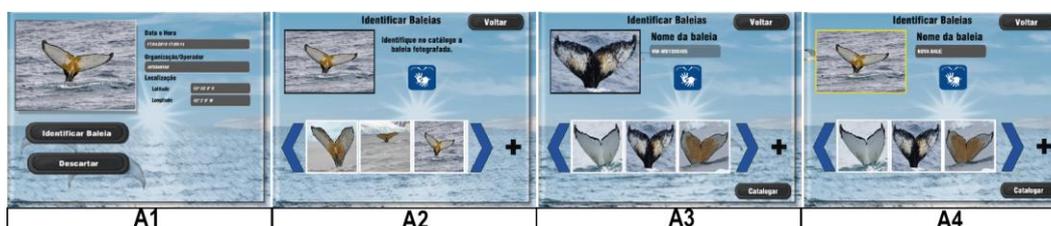


Figura 1 - Telas do protótipo da versão A

As duas versões se diferem em relação ao conteúdo dos diálogos e ao processo de fotoidentificação. Na versão B, os diálogos são mais específicos e diretos, no que concerne ao projeto de CC. Na versão B, foi acrescentada a descrição e o processo de fotoidentificação, em que os jogadores devem identificar as características da cauda da baleia, que são: pigmentação, marcas, manchas, riscos, borda, ponta e entalhe. Após fotografar a cauda, é apresentada ao jogador uma tela comum em ambas as versões (vide **Figura 1 A1**) e ao clicar em *identificar baleia*, inicia-se o processo de fotoidentificação (vide **Figuras 2 B1 a B7**), i.e., o jogador analisa as características da foto tirada, selecionando a opção que julgar correta (caso acerte a borda fica verde e é habilitado um botão de *próximo* levando-o para a próxima característica, caso contrário a borda fica vermelha). Há também a opção *não identificado*, caso não seja identificada uma opção a partir da foto e, então, já passa para a próxima característica. Ao finalizar, é apresentado um catálogo (semelhante da **Figura 1 A2**) em que o jogador deve comparar a foto tirada com fotos em um catálogo filtrado (com as características selecionadas). Caso identifique a foto no filtro, o jogador cataloga a baleia existente (**Figura 2 B8**). Caso não, será preciso nomeá-la antes de catalogá-la.

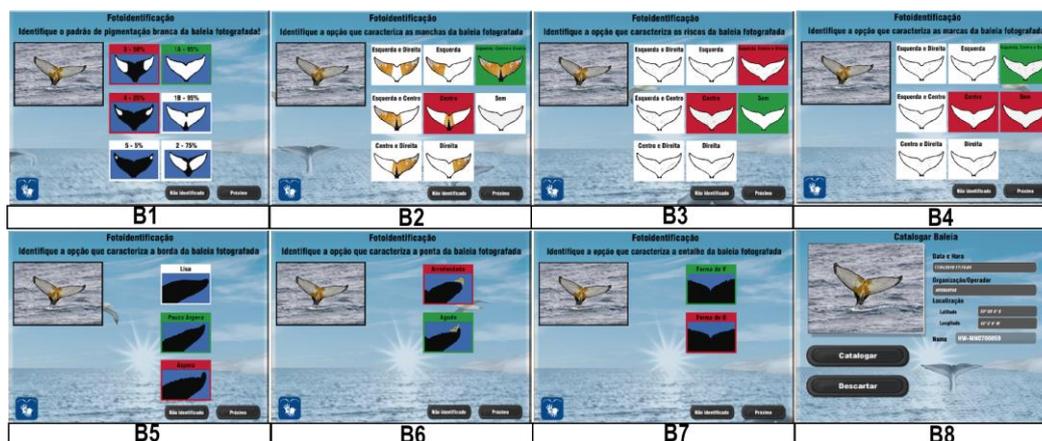


Figura 2 - Telas do protótipo da versão B

5.2. Questionário de Perfil do Jogador

Os participantes da avaliação são em maioria do curso de Bel. em Ciência e Tecnologia (64,3%), com alguns alunos do curso de Ciência da Computação (28,6%) e Engenharia da Informação (7,1%), com idade de entre 20 e 43 anos, sendo 71,4% entre 21 e 25 anos. De acordo com o questionário, 78,5% dos alunos possuem pouca experiência com desenvolvimento de jogos, no entanto, 57,1% relataram já terem jogado algum JED.

5.3. Questionário de Reação ao Jogo

Os dados obtidos com o questionário, já normalizados para efeitos de comparação, podem ser encontrados na **Tabela 1**, onde maiores valores de média representam melhor desempenho. Os jogadores reagiram mais favoravelmente à versão B do que à versão A, na maioria das questões, exceto nos itens R6 (acabar logo o jogo) e R8 (*design* da interface - que é a mesma para ambas as versões).

Tabela 1 - Visão geral das respostas do questionário de reação

Itens	R1		R2		R3		R4		R5		R6		R7		R8	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Média	3,00	3,83	2,71	4,33	2,43	3,17	3,29	3,33	2,57	3,17	3,43	3,33	2,29	3,00	2,57	2,50
DP	0,58	0,98	1,11	0,82	1,27	0,98	1,38	1,37	1,27	1,17	1,40	0,82	1,38	1,41	0,98	1,38
Diferença (B - A)	0,83		1,62		0,74		0,05		0,60		-0,10		0,71		-0,07	

5.4. Questionário de Pré e Pós-Teste (Aprendizagem)

As análises dos questionários foram realizadas de maneira qualitativa comparando as respostas do pré e pós-teste (resultados na **Tabela 2**). A maioria dos jogadores não sabia responder as questões do pré-teste e era esperado que no pós-teste os acertos fossem maiores em ambas as versões. Na versão A, a quantidade de acertos no pós-teste sempre foi inferior que na versão B; na questão (2), houve um aumento na porcentagem de pessoas que afirmaram desconhecer; além disso, uma pessoa errou a questão (3). Na versão B, a maioria (57%) acertou as questões (1) e (2), e 100% acertaram a (3), porém, houve um aumento de 28% de erro na resposta da questão (1).

Tabela 2 - Visão geral das respostas do questionário de comparação

Questões	1				2				3			
	A		B		A		B		A		B	
Versão	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós
Desconheço (%)	100	71,5	85,7	0	71,5	85,7	85,7	28,6	100	42,9	100	0
Acerto (%)	0	28,5	0	57,1	28,5	14,3	0	57,1	0	42,9	0	100
Erro (%)	0	0	14,3	42,9	0	0	14,3	14,3	0	14,3	0	0

5.5. Questionário de Comparação

Dentre as questões abertas, é possível apurar que 14,3% dos jogadores gostariam de jogar novas fases da versão A e 64,3% da versão B, enquanto que 21,4% não gostariam de jogar novas fases. Outros aspectos que foram avaliados em questões fechadas (vide seção 4.5), são apresentados na forma de média na **Tabela 3**. Vale notar que a versão B obteve médias maiores e melhores do que a versão A em todos os itens comparados.

Tabela 3 - Visão geral das respostas do questionário de comparação

Itens	C1		C2		C3		C4		C5		C6		C7	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Média	2,86	5,36	3,21	4,00	2,14	4,79	2,71	3,93	2,36	4,00	2,71	4,07	2,71	4,21
DP	1,29	0,84	1,42	1,18	1,29	1,37	1,68	1,27	1,50	1,24	1,27	1,14	1,38	0,97

6. Considerações Finais

A avaliação em jogos educacionais deve ocorrer para garantir a sua efetividade. Neste trabalho foi apresentada a metodologia de desenvolvimento de questionários de perfil, avaliação de reação, aprendizagem e comparação em jogos educacionais. Também foi descrito um protocolo de avaliação, que objetivou guiar a execução e aplicação dos instrumentos para avaliação do jogo Expedição Antártica.

Por meio do protocolo de avaliação proposto, foi possível realizar uma análise comparativa entre dois protótipos do jogo Expedição Antártica. As duas versões possuem *design* semelhante com diferenças principais nas interações com os mentores (diálogos) e no processo de fotoidentificação. Ambos os jogos buscam alcançar o mesmo objetivo pedagógico de ensino sobre a ciência cidadã, metodologia científica e identificação de baleias. A partir dos questionários de reação e comparação, foi possível observar que a versão B se sobressai em relação à versão A em quase todos os aspectos levantados, seja relacionado ao desafio ou a aprendizagem, por exemplo. A diferença se mostra maior no questionário de comparação, no qual ambos os grupos já haviam jogado as duas versões. Isso possibilita uma escolha de uso de uma versão, bem como aprimoramento dela (a partir dos *feedbacks* qualitativos).

Os artefatos propostos e o uso deles, bem como sua análise, podem servir de apoio para a realização de avaliações de outros jogos sérios. Como trabalhos futuros, pretende-se aprimorar o protocolo para novas análises e/ou comparações de jogos, além de aperfeiçoar a análise estatística dos dados quantitativos.

Referências

- Drummond, D.; Monnier, D.; Tesnière, A. H. A. (2017). “A systematic review of serious games in asthma education”. *Pediatr Allergy Immunol*, v. 28, p. 257-265.
- Fu, F. L., Su, R. C.; Yu, S. C. (2009). “EGameFlow: a scale to measure learners’ enjoyment of e-learning games”. In: *Computers & Education*, v. 52(1), p. 101-112.
- Kirkpatrick, D.L.; Kirkpatrick, J.D. (2006). “Evaluating Training Programs: The Four Levels”. 3rd Edition. San Francisco, USA: Berrett-Koehler, p. 1-568.
- Miller, A.; Imrie, B. W.; Cox, K. (1998). Functions of Assessment. In: “Student Assessment in Higher Education: a handbook for assessing performance” London: Kogan Page, p. 23-40.
- Mourão, L.; Meneses, P. P. M. (2012). Construção de Medidas em TD&E. In: Abbad, G. S.; *et al.* “Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: ferramentas para gestão de pessoas”. Porto Alegre: Artmed. p. 50-63.
- Oliveira, A. R. A. (2014). “Questionário para Avaliação de Sistemas de Software Educacionais no Apoio do Processo de Ensino-Aprendizagem em Gerência de Projetos de Software”. Monografia, Universidade Federal de Lavras, Departamento de Ciência da Computação.
- Oliveira, R.N.R., Cardoso, R.P., Braga, J.C., Rocha, R.V. (2018). “Frameworks para Desenvolvimento de Jogos Educacionais: uma revisão e comparação de pesquisas recentes”. In: SBIE 2018, p. 854-863.
- Pereira Junior, H. A.; Menezes, C. S. (2015). “Modelo para um Framework Computacional para Avaliação Formativa da aprendizagem em jogos digitais”. In: SBGames 2015, p. 819-828.
- Perrenoud, P. (1999). “Avaliação: da excelência à regulação das aprendizagens entre duas lógicas”. Porto Alegre: Artes Médicas, p. 1-183.
- Petri, G.; von Wangenheim, C. G.; Borgatto, A. F. (2018). MEEGA+, Systematic Model to Evaluate Educational Games. In: Lee N. (eds), "Encyclopedia of Computer Graphics and Games". Springer.
- Pires, M. R. G. M., Göttems, L. B. D., Silva, L. V. S., Carvalho P. A., Melo G. F., Fonseca R. M. G. S. (2015). “Desenvolvimento e validação de instrumento para avaliar a ludicidade de jogos em saúde”. In: *Rev. Esc. Enferm USP*, v. 49(6), p. 981-990.
- Pontes, B.R. (1999). “Avaliação de Desempenho: nova abordagem”. 7. ed. SP: LTr.
- Rocha, R. V. (2017). “Critérios para a construção de jogos sérios”. In: SBIE 2018, p. 947-956.
- Rocha, R. V., Bittencourt, I. I., Isotani, S. (2015). “Avaliação de Jogos Sérios: questionário para autoavaliação e avaliação da reação do aprendiz”. In: SBGames 2015, p. 648-657.
- Rocha, R.V. (2014). “Metodologia iterativa e modelos integradores para desenvolvimento de jogos sérios de treinamento e avaliação de desempenho humano”. Tese - UFSCar, Departamento de Computação.
- Salas, E.; *et al.* (2009). “Performance Measurement in Simulation-Based Training: A Review and Best Practices”. In: *Simulation Gaming*, v. 40(3), p. 328-376.
- Savi, S., Wangenheim, C.G.; Ulbricht, V.; Vanzin, T. (2010). “Proposta de um modelo para avaliação de jogos educacionais”. In: *RENTE*, v. 8(3). p. 1-10.
- Vieira, K. M.; Dalmoro, M. (2008). “Dilemas na Construção de Escalas Tipo Likert: o Número de Itens e a Disposição Influenciam nos Resultados?”. In: ANPAD 2008, p. 1-16.
- Zerbini, T.; Borges-Ferreira, M. F.; Abbad, G. S. (2012). Medidas de Reação a Cursos a Distância. In: ABBAD, G. S. “Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: ferramentas para gestão de pessoas”. Porto Alegre: Artmed, p. 91-107.