

Revisão Sistemática de Literatura sobre métodos, técnicas e critérios de avaliação de aprendizagem em jogos sérios sociointeracionistas

Rodrigo Dobbin Fellows¹, Juliana Cristina Braga¹, Silvia Cristina Dotta¹

¹Universidade Federal do ABC (UFABC)

Caixa Postal 09210-580 – Santo André – SP – Brasil

rodrigodobbinf@gmail.com, {juliana.braga,silvia.dotta}@ufabc.edu.br

Abstract. *Assessing the student is important to understand how the teaching-learning process occurs. In educational games, some methods are used to help in this evaluation, however, just a few consider humanistic pedagogical approaches. In socio-interactionist educational approach, creativity, interaction and the student's socio-historical-cultural context are valued, therefore, the evaluation of socio-interactionist games distances itself from behaviorist practices. This systematic review was carried out in order to understand the methods, techniques, criteria, frameworks and/or algorithms that can be used for the evaluation of socio-interactionist educational games.*

Resumo. *Avaliar o estudante é importante para entender como o processo de ensino e aprendizagem ocorre. Em jogos sérios, alguns métodos são utilizados para ajudar nessa avaliação, no entanto, poucos consideram abordagens pedagógicas humanistas. Na abordagem educacional sociointeracionista, são valorizados a criatividade, a interação e o contexto sócio-histórico-cultural do estudante, sendo assim, a avaliação de jogos sociointeracionistas distancia-se de práticas behavioristas. Essa revisão sistemática foi realizada no intuito de entender quais são os métodos, técnicas, critérios, frameworks e/ou algoritmos que podem ser utilizados para a avaliação de jogos sérios sociointeracionistas.*

1. Introdução

Em um modelo de educação interacionista e humanista, que valoriza a criatividade, o cidadão deve ser estimulado a aprender de forma significativa, colaborar, trabalhar em equipe, compartilhar, assumir iniciativas, agir com criatividade e inovação, além de desenvolver senso crítico para resolver problemas.

Os jogos, quando utilizados para fins educacionais, podem proporcionar um modelo de educação interacionista, gerando possibilidades para uma aprendizagem mais humanística.

No entanto, medir os resultados da aprendizagem em jogos sérios é algo considerado complexo. Existem estudos que apontam melhorias no desempenho acadêmico dos estudantes e outros que não apresentam evidências de melhoria no desempenho escolar.

Segundo [Del Blanco et al. 2010], um dos desafios mais relevantes enfrentados por professores e instrutores é avaliar os resultados de aprendizagem das sessões de jogo.

Métodos automatizados para avaliar a aprendizagem em jogos podem diminuir esse desafio e são utilizados para auxiliar o professor a mensurar o desempenho do aluno na interação com jogos. No entanto, poucos deles consideram abordagens pedagógicas humanistas onde são valorizados a criatividade, a interação e o contexto sócio-histórico-cultural do estudante.

A literatura apresenta trabalhos que consideram avaliação automatizada de jogos, mas as soluções aproximam-se mais das práticas behavioristas, como por exemplo, o tradicional questionário de retorno automático inserido no jogo para automatizar a avaliação da aprendizagem do estudante. Há em todo esse cenário uma contradição, pois se por um lado o jogo pode ser usado para aprendizagem humanista, por outro, as técnicas computacionais implementadas em jogos para avaliar aprendizagem distanciam-se dessa abordagem.

Diantes desse contexto, identificou-se como um problema a ser investigado a dificuldade de se automatizar avaliação de aprendizagem em jogos sérios sociointeracionistas, de modo que os componentes de avaliação não descaracterizem os principais elementos dessa abordagem. No intuito de apontar solução para esse problema, esse estudo propõe uma revisão sistemática que visa compreender quais são os métodos, frameworks e/ou algoritmos que podem ser utilizados para a avaliação de jogos sérios sociointeracionistas. Ao final da revisão, as seguintes questões de pesquisa serão respondidas: Qual ou quais técnicas de avaliação automática podem ser utilizadas para acompanhar o aprendizado do estudante em jogos sérios baseados na abordagem educacional sociointeracionista?

As contribuições desse estudo são: (i) sintetizar as técnicas e métodos de avaliação mais utilizados para medir a aprendizagem em jogos sérios, (ii) resumir quais são os critérios de avaliação mais utilizados em jogos sérios, e (iii) mostrar como podem ser utilizados em um ambiente sócio-interacionista.

2. Trabalhos relacionados

Com intuito de verificar se já existia alguma revisão sistemática para responder as questões de pesquisa desse trabalho, uma primeira revisão sistemática foi realizada. levantar o estado da arte para investigar, avaliar e interpretar o que está acontecendo em um assunto científico. Essa Revisão teve como ponto de partida a tentativa em descobrir como analisar o processo de aprendizagem de jogos sérios cuja proposta é baseada no conceito da teoria de aprendizagem sociointeracionista de Lev Vigotski. E o artigo Calderón, A. and Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, 87:396–422. O qual aborda a avaliação em jogos sérios.

Para tal foi feita então uma pesquisa por outras revisões sistemáticas de literatura sobre o mesmo tema nas bases: *DBLP*, *ACM Digital*, *IEEE Xplore* e *Google Scholar*. Utilizando a *String* de busca (*Learning Assessment OR Teaching Learning Process OR Teaching-learning process*) *AND* (*Serious Game OR Educational Games*) *AND* (*Systematic review OR Research review OR Systematic overview OR Systematic literature OR SLR*).

Não houve nenhuma similaridade significativa com os resultados apresentados nessa busca, para que invalidasse a necessidade dessa revisão. Contudo um estudo dessa

busca foi levado em consideração, Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4):1432–1462. O qual fez uma pesquisa e revisão sobre os trabalhos mais recentes com Mineração de dados educacionais, e apesar desse trabalho em específico não ter abordado o tema de aprendizagem ou processo de aprendizado em jogos sérios, ele citou outros quatro trabalhos que foram adicionados para leitura completa no processo dessa revisão.

3. Metodologia

Esse estudo seguiu as diretrizes propostas por *Kitchenham & Charters* [Keele et al. 2007].

A revisão sistemática teve como principal objetivo entender quais são os métodos, técnicas e critérios mais utilizados para avaliação do aprendizado e do processo de de aprendizado em um jogo sério sociointeracionista.

3.1. Questões de pesquisa

Oito questões de pesquisa específicas foram definidas baseada no objetivo especificado na introdução desse trabalho:

- **QP1:** Que tipo de jogo foi utilizado pelo estudo?
- **QP2:** Qual o público alvo do jogo?
- **QP3:** Qual a área de conhecimento do jogo (para que tipo de área de conhecimento ele foi feito)?
- **QP4:** Qual ou quais os objetivos pedagógicos do jogo?
- **QP5:** Qual ou quais técnicas de avaliação foram utilizadas nos estudos analisados?
- **QP6:** Qual ou quais métodos de avaliação foram utilizadas no jogo?
- **QP7:** Qual ou quais os critérios de avaliação foram utilizados?
- **QP8:** Qual a teoria de aprendizagem utilizada pelo jogo?

3.2. Estratégia de busca

A estratégia de busca foi baseada na busca por artigos científicos em reconhecidas base de dados digitais, sendo elas:

- <http://dblp.uni-trier.de>
- <http://dl.acm.org/dl.cfm>
- <https://ieeexplore.ieee.org>
- <https://scholar.google.com>

Para isso uma *String* de busca foi desenvolvida baseado nos cinco passos descritos por *Brereton, Kitchenham, Budgen, Turner, & Khalil* [Brereton et al. 2007], os quais são:

- Obter os principais termos das perguntas, identificando os conceitos principais;
- Identificar grafias alternativas e sinônimos para os termos principais;
- Verificar as palavras-chave em todos os artigos relevantes que já é possuído;
- Usar o “booleano” *OR* para adicionar alternativas de grafia e sinônimos;
- Usar o “booleano” *AND* para vincular os termos principais.

Com isso foi definida a *String* de busca (*Sociointeractionism OR Lev Vygotsky OR Vygotsky*) *AND* (*Serious Game OR Educational Games*) *AND* (*Stealth Assessment OR Game-Based Learning Assessments OR Learning Assessment OR Learning Analysis OR Teaching Learning Process OR Teaching-learning process OR Teaching-learning*), a qual será chamada *String* de busca completa.

Após realizar as buscas nas bases com essa *String*, apenas 25 resultados foram encontrados, isso ocorreu por causa do primeiro bloco de termos, que trata sobre sociointeracionismo, o que nos levanta uma hipótese dele ser pouco abordado quando relacionado a Sistema da Informação. Para que uma nova busca fosse realizada, a *String* de busca foi adaptada (*Serious Game OR Educational Games*) *AND* (*Stealth Assessment OR Game-Based Learning Assessments OR Learning Assessment OR Learning Analysis OR Teaching Learning Process OR Teaching-learning process OR Teaching-learning*), a qual será chamada de *String* de busca adaptada. Portanto essa revisão foi feita com duas *Strings* diferentes, a *String* de busca completa e a *String* de busca adaptada.

Um fato aconteceu após analisar alguns estudos encontrados a partir da *String* de busca adaptada, sendo ele o descobrimento de um novo termo de busca, o *Dynamic Assessment* ou *Dynamic Testing*, esses termos foram incluídos e a busca foi refeita, mas não trouxe nenhum estudo a mais a ser considerado depois da leitura do Título e Resumo dos mesmos, portanto esses termos não foram adicionados como parte da busca das nossas *Strings* finais.

Para auxiliar no processo de armazenamento das informações adquiridas durante a leitura dos artigos foi utilizado o *Google Sheets*TM, e para organização dos artigos exportados e remoção dos duplicados foi utilizado o *Zotero* [Ahmed and Al Dhubaib 2011].

3.3. Seleção dos estudos e critérios de exclusão

Inicialmente para seleção dos estudos foi estabelecido um critério que aceitava apenas artigos feitos a partir de 2015, mas como a primeira pesquisa retornou poucos resultados, esse limite foi removido, e não foi definido data inicial para escrita dos artigos.

O processo de seleção portanto se deu em três etapas. A primeira etapa foi a validação das palavras-chave selecionadas para as **Strings** de busca, através da busca pelas palavras pesquisadas em uma seleção aleatória de alguns dos estudos encontrados pelas buscas. A segunda foi a seleção dos artigos que respeitaram os critérios de exclusão representados pela Tabela 1, a partir da leitura do seus Títulos e Resumos. Já a terceira se deu a partir dos artigos aprovados na primeira etapa, aqui foi feita a leitura do texto completo, e caso ele cumprisse os critérios da Tabela 1 seriam excluídos do estudo.

Tabela 1. Critérios de exclusão

Critérios de Exclusão	<ul style="list-style-type: none"> - Documentos Repetidos. - Publicações que não são artigo de periódicos ou Publicações em anais de conferência. - Publicações que não demonstraram como funciona ou como deve ser implementado seu método ou técnica de avaliação. - Publicações que não apresentem métodos ou técnicas de avaliação de jogos sérios.
-----------------------	---

3.4. Avaliação de qualidade

Cada artigo foi confrontado com um questionário contendo três perguntas, elas serviram para avaliar a qualidade do artigos perante o propósito desse estudo. Sendo elas:

- **AQ1:** O artigo contem informações sobre a técnica/método de avaliação que utilizou?
- **AQ2:** O artigo contém um experimento utilizando a técnica/método de avaliação apresentada?
- **AQ3:** O artigo menciona quais os critérios utilizados para a avaliação da aprendizagem?

Cada questão poderia ser respondida com “Sim” ou “Não”, sendo obrigatório para o estudo passar no critério de qualidade ter as três respostas “Sim”. É preciso portanto entender qual ou quais critérios, técnicas e métodos foram utilizadas e qual sua assertividade/acurácia, e para tal é necessário um experimento. Por isso as três avaliações de qualidade precisam ser respondidas de maneira positiva.

4. Resultados

Como já mencionado essa revisão possui uma peculiaridade em ter sido feita com duas strings diferentes, pois a primeira busca teve apenas o retorno de 25 artigos. Por conta disso foi necessário um novo processo de busca utilizando a *String* de busca adaptada, a qual retornou 481 resultados.

A análise dos dois resultados foi feita de acordo com os seguintes passos: primeiro é obtido os artigos das bases, os arquivos duplicados são removidos, posteriormente é feita a primeira etapa da revisão que é constituída da leitura do Resumo e do Título, caso eles não contenham nenhuma informação relevante para alguma das questões de pesquisa, eles são removidos, em seguida os artigos selecionados são lidos por completo. Esses artigos que foram lidos por completo são analisados de acordo com as questões de qualidade AQ1, AQ2 e AQ3, para que assim o estudo possa entrar como selecionado para essa revisão. Esse processo está demonstrado na Tabela 2.

Tabela 2. Quantidade de artigos selecionados após cada etapa do processo de busca e seleção

Etapa	<i>String</i> Completa	<i>String</i> Adaptada
Busca	25	481
Eliminação dos livros e artigos duplicados	24	432
Primeira revisão (leitura do Título e Resumo)	1	59
Segunda revisão (leitura do artigo completo)	1	29
Avaliação de qualidade	1	18

A seleção portanto resultou em 29 estudos, os quais foram confrontados com as três questões de análise de qualidade pré definidas (AQ1, AQ2 e AQ3). Para a **AQ1** 24 estudos obtiveram **sim** e 5 **não**, já para a **AQ2** 26 estudos obtiveram **sim** e 3 **não**, e para a **AQ3** 22 estudos obtiveram **sim** e 7 **não**. Com isso mais 11 estudos foram excluídos dessa revisão.

5. Discussão

Nessa revisão foram encontrados 29 estudos primários, dos quais apenas 18 foram considerados no processo para responder as questões de pesquisa, pois os 11 restantes não passaram no critério de avaliação. De acordo com os resultados obtidos na busca e nos estudos analisados, foi observado que está acontecendo um aumento na quantidade de pesquisas relacionadas ao que chamam de *Stealth Assessment*, uma avaliação que é feita de forma não disruptiva dentro do contexto do próprio jogo. A qual segundo [Mislevy and Haertel 2006] é normalmente desenvolvida usando a estrutura de design centrado em evidências (ECD) que visa estabelecer um argumento logicamente coerente entre o domínio que está sendo aplicado (modelo de competência), design de tarefa de avaliação (modelo de tarefa) e interpretação (modelo de evidência).

Com a evolução na capacidade dos computadores, e consequente aprimoramento dos algoritmos de aprendizado de máquina, também foi observado um uso mais crescente dessa técnica para esse tipo de avaliação, ou seja, há uma tentativa de automatizar esse processo de avaliação, para que não precise ser feito de forma manual.

O domínio de aplicação mais explorado por jogos sérios que continham alguma forma de avaliação do aprendizado foi o da Educação, mais especificamente o da Educação fundamental.

Foi percebido também que a maioria dos jogos possuem os elementos de avaliação da aprendizagem com cunho behaviorista, ou seja na hora de medir o aprendizado existem respostas certas e erradas, eles não conseguem medir o processo de aprendizagem.

Abaixo iremos discutir as questões de pesquisa propostas por esta revisão, e fazer algumas considerações sobre cada uma delas.

5.1. QP1: Que tipo de jogo foi utilizado pelo estudo?

Todos os jogos foram feitos ou para serem executados em computadores ou dispositivos móveis. E os jogos mais utilizados pelos estudos analisados, são do tipo “Mundo Virtual”, os quais fazem o jogador imergir em um novo mundo de interação, um mundo paralelo e não real. Os tipos de jogos analisados pelos estudos são:

- **Jogos para Computador** [Del Blanco et al. 2010];
- **Quebra-cabeça** [Rowe et al. 2017];
- **Mundo Virtual** [Moreno-Ger et al. 2007], [Thomas et al. 2012], [Harpstead et al. 2013], [Baron et al. 2014], [de Lima et al. 2016], [Mota and Pimentel 2017], [Falakmasir et al. 2016], [Georgiadis et al. 2019a], [Georgiadis et al. 2019c], [Min et al. 2019];
- **Jogos Mobile** [Kam et al. 2007], [Kam et al. 2009];
- **MMORPGs** [Conrad et al. 2014].

5.2. QP2: Qual o público alvo do jogo?

A maior parte dos jogos foram feitos voltados para área de educação, e tendo como seu principal público alvo crianças de 6 até 15 anos. O público alvo dos jogos analisados são:

- **Crianças até 6 anos** [Baron et al. 2014];

- **Crianças de 6 até 15 anos** [Rowe et al. 2017], [Del Blanco et al. 2010], [Asbell-Clarke et al. 2013], [Harpstead et al. 2013], [Conrad et al. 2014], [Mota and Pimentel 2017], [Falakmasir et al. 2016], [Min et al. 2019], [Kam et al. 2007], [Kam et al. 2009];
- **Pessoas com interesse por gastronomia** [Moreno-Ger et al. 2007];
- **Bancários** [Thomas et al. 2012];
- **Ensino superior** [de Lima et al. 2016], [Georgiadis et al. 2019a].

5.3. QP3: Qual a área de conhecimento do jogo (para que tipo de área de conhecimento ele foi feito)?

Corroborando o fato explanado na resposta da QP2, a área de conhecimento mais abordada pelos estudos são as áreas vistas no Ensino fundamental, o que faz sentido, uma vez que essa área é geralmente frequentada por crianças de 6 até 15 anos. A análise completa com todas as áreas de conhecimento pode ser vista abaixo:

- **Educação fundamental** [Del Blanco et al. 2010], [Rowe et al. 2017], [Asbell-Clarke et al. 2013], [Harpstead et al. 2013], [Mota and Pimentel 2017], [Falakmasir et al. 2016], [Kam et al. 2007], [Kam et al. 2009], [Conrad et al. 2014];
- **Gastronomia** [Moreno-Ger et al. 2007];
- **Bancária** [Thomas et al. 2012];
- **Saúde** [de Lima et al. 2016], [Georgiadis et al. 2019a];
- **Psicologia** [Georgiadis et al. 2019c].

Uma análise das áreas do Ensino fundamental mostra que há uma diversidade entre elas, sendo Física 33.3%, Matemáticas 22.2%, Língua Inglesa 22.2%, STEM 11.1% e História 11.1% das áreas analisadas.

5.4. QP4: Qual ou quais os objetivos pedagógicos do jogo?

Sobre os objetivos pedagógicos, a área de ciências exatas superou as demais áreas. A lista com todos os objetivos pedagógicos dos estudos analisados na revisão é:

- **Ensinar História** [Del Blanco et al. 2010];
- **Conceitos de Física** [Rowe et al. 2017], [Asbell-Clarke et al. 2013], [Harpstead et al. 2013];
- **Conceitos de matemática** [Mota and Pimentel 2017], [Falakmasir et al. 2016];
- **Inglês como segunda língua (ESL)** [Kam et al. 2007], [Kam et al. 2009];
- **STEM** [Conrad et al. 2014];
- **Ensino da culinária** [Moreno-Ger et al. 2007];
- **Conhecimentos bancários** [Thomas et al. 2012];
- **Melhoria no ensino da saúde** [de Lima et al. 2016], [Georgiadis et al. 2019a];
- **Definição de personalidade do jogador** [Georgiadis et al. 2019c].

5.5. QP5: Qual ou quais técnicas de avaliação foram utilizadas nos estudos analisados?

Há um crescente aumento no uso de técnicas como aprendizado de máquina e mineração de dados educacionais e as mais usadas são Redes Baysianas, Redes Neurais e *Hidden Markov Model (HMM)*. Isso foi observado principalmente nos estudos mais recentes (dos últimos 5 anos). Todas as técnicas encontradas nos estudos são:

- **Gravação da interação do usuário com o jogo** [Rowe et al. 2017], [Asbell-Clarke et al. 2013], [Conrad et al. 2014];
- **Parâmetros implícitos** [Rowe et al. 2017], [Del Blanco et al. 2010], [de Lima et al. 2016], [Falakmasir et al. 2016];
- **Mineração de dados educacionais** [Rowe et al. 2017], [Asbell-Clarke et al. 2013], [Georgiadis et al. 2018], [Georgiadis et al. 2019b], [Georgiadis et al. 2019a], [Georgiadis et al. 2019c], [Min et al. 2019];
- **Petri nets (PN)** [Thomas et al. 2012];
- **Modelo de equação estrutural** [Baron et al. 2014];
- **Aprendizado de máquina** [Harpstead et al. 2013], [Conrad et al. 2014], [Falakmasir et al. 2016], [Georgiadis et al. 2018], [Georgiadis et al. 2019b], [Georgiadis et al. 2019a], [Georgiadis et al. 2019c], [Min et al. 2019].

5.6. QP6: Qual ou quais métodos de avaliação foram utilizadas no jogo?

As técnicas que foram previamente demonstradas, são utilizadas de acordo com métodos pré-definidos. Como a maior parte dos trabalhos tem uma abordagem mais behaviorista, os métodos mais utilizados consequentemente são a aplicação de testes tradicionais, como se vê em instituições de ensino, podendo ser aplicado antes do experimento, depois, ou tanto antes quanto depois. Os métodos encontrados nos estudos estão listados abaixo:

- **Pré teste** [Moreno-Ger et al. 2007], [Del Blanco et al. 2010];
- **Pré e pós teste** [Rowe et al. 2017], [Kam et al. 2007], [Kam et al. 2009], [Harpstead et al. 2013], [Baron et al. 2014];
- **Testes in-game** [Del Blanco et al. 2010], [de Lima et al. 2016];
- **Avanço no jogo** [Del Blanco et al. 2010], [Harpstead et al. 2013], [Nyamsuren et al. 2017], [Mota and Pimentel 2017];
- **Logs dos jogos** [Rowe et al. 2017], [Falakmasir et al. 2016], [Georgiadis et al. 2019a], [Georgiadis et al. 2019c], [Min et al. 2019];
- **Avaliação dos aplicadores do experimento (professores)** [Rowe et al. 2017];
- **Narrativa do jogo** [Moreno-Ger et al. 2007];
- **Cognitive Reliability and Error Analysis Method (CREAM)** [Thomas et al. 2012];
- **Comparação do resultados de pré testes vs. números apresentados pela técnica aplicada** [Thomas et al. 2012], [Conrad et al. 2014];
- **Análise do gameplay** [Asbell-Clarke et al. 2013].

5.7. QP:7 Qual ou quais os critérios de avaliação foram utilizados?

Já como critério de avaliação o que predomina é um modelo conhecido como *Evidence Centered Assessment Design (ECD)*. Esse modelo é uma abordagem para construir avaliações em termos de argumentos probatórios, tem como base de avaliação as evidências produzidas pelo usuário participante do experimento. O que faz sentido pois uma técnica que foi bastante utilizada foi a de Mineração de dados educacionais, e como o *ECD* consegue trabalhar facilmente com logs, se integra muito bem no contexto de mineração de dados. Todos os critérios de avaliação dos estudos analisados são:

- **Validação das ações do jogador no decorrer do jogo** [Rowe et al. 2017], [Baron et al. 2014];

- **Validação das ações do usuário vs. o domínio de conhecimento** [Rowe et al. 2017], [Moreno-Ger et al. 2007], [Asbell-Clarke et al. 2013];
- **Acertar o questionário teste** [Kam et al. 2007], [Kam et al. 2009], [Del Blanco et al. 2010];
- **Evidence Centered Assessment Design (ECD)** [Harpstead et al. 2013], [Falakmasir et al. 2016], [Georgiadis et al. 2018], [Georgiadis et al. 2019b], [Georgiadis et al. 2019a], [Georgiadis et al. 2019c], [Min et al. 2019];
- **Experiment Centered Assessment Design (XCD)** [Conrad et al. 2014].

O estudo feito por [Conrad et al. 2014], vale ser ressaltado, pois ele foi o único que apresentou um critério conhecido como *Experiment Centered Assessment Design (XCD)*, critério se baseia no *ECD*, mas difere na questão que ele não toma como base ações ou eventos individuais, ele armazena um conjunto de ações do usuário, e chama esse conjunto de experimento, fazendo sua análise baseado do experimento, pois ele fala que não é ideal analisar evidências isoladas, pois no mundo aberto (ou no mundo real) a aleatoriedade (imprevistos) pode auxiliar no processo de aprendizado, e que nem tudo é de simples interação. Isso remete ao aprendizado baseado na experiência, baseado na vivência do todo.

5.8. QP8: Qual a teoria de aprendizagem utilizada pelo jogo?

Apesar de nenhum dos estudos analisados ter citado explicitamente que seguiu a diretriz proposta por alguma teoria de aprendizagem cabe fazer algumas menções.

O estudo feito por [Rowe et al. 2017] citou a zona de desenvolvimento proximal de Vigotski [Vygotsky 1997], como ponto para ser observado pelo *game designer*, ao falar sobre a possibilidade de inferir alterações em tempo real no jogo para aumentar o aprendizado do jogador. Já os estudos de [Moreno-Ger et al. 2007] e [Conrad et al. 2014], apesar de não terem citado a teoria sociointeracionista, introduziram suas avaliações dentro da própria narrativa do jogo, o que se aproxima de uma experiência de aprendizado baseado em vivências.

Já os estudos de [Kam et al. 2007], [Kam et al. 2009], [Del Blanco et al. 2010], [de Lima et al. 2016] e [Mota and Pimentel 2017] não citaram nenhuma teoria de aprendizado, mas utilizaram técnicas *behavioristas* em seu método e/ou critério de avaliação.

6. Conclusão

Os estudos analisados por esse artigo uma crescente em relação a tentativa de automatizar o processo de avaliação de aprendizado, utilizando técnicas com aprendizado de máquina e mineração de dados. Assim como também mostram que a grande maioria dos jogos ainda utilizam métodos *behavioristas* na hora de realizarem essas avaliações nos jogos, aplicando testes que contem respostas “certas” e “erradas”.

Mas quando falamos de um ambiente de aprendizado de interação, o qual o indivíduo não aprende apenas acertando, mas sim explorando, errando, trocando experiências com seu contexto sócio e cultural. Estamos falando de um cenário sócio-interacionista, o qual foi o ponto de partida dessa análise. Nesse cenário há uma diferença entre medir o aprendizado e medir o processo de aprendizagem.

Para que essa aferição possa ser feita sem impactar o desempenho e sem causar ruídos para o estudante, toda avaliação deve estar inserida dentro do jogo, de forma não

disruptiva. Ela tem que estar inserida na narrativa do próprio jogo, fazendo com que o jogador consiga imergir nesse contexto, obtendo aprendizado através da interação com o ambiente do jogo.

Com tudo isso em mente, a Avaliação Silenciosa (*Stealth Assessment*) mostra potencial para realizar essa tarefa de avaliação de aprendizagem em um contexto socio-interacionista, pois seu princípio é não ser disruptiva e não "parar/bloquear" o processo/andamento do jogo. Assim como utilizar *logs* e mineração de dados educacionais, aplicados a um ambiente que utiliza como critério de avaliação o XCD - *Experiment Centered Assessment Design* [Conrad et al. 2014], uma evolução do ECD - *Evidence Centered Assessment Design*, temos um ambiente montado para uma avaliação baseada na interação do usuário em todo o contexto do jogo e não só focado em respostas de momentos pontuais do jogo.

Como trabalho futuro, é sugerido um aprofundamento na Avaliação Silenciosa assim como nos critérios de avaliação ECD e XCD, com intuito de produzir um modelo, algoritmo ou proposta de avaliação sociointeracionista. A qual seja capaz de ser aplicada em um contexto de aprendizado não disruptivo, um contexto que contenha troca de experiências e vivências dentro do jogo.

Referências

- Ahmed, K. M. and Al Dhubaib, B. (2011). Zotero: A bibliographic assistant to researcher. *Journal of Pharmacology and Pharmacotherapeutics*, 2(4):303.
- Asbell-Clarke, J., Rowe, E., and Sylvan, E. (2013). Assessment design for emergent game-based learning. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 679–684.
- Baron, H. B., Salinas, S. C., and Crespo, R. G. (2014). An approach to assessment of video game-based learning using structural equation model. In *2014 9th Iberian conference on information systems and technologies (CISTI)*, pages 1–6. IEEE.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4):571–583.
- Conrad, S., Clarke-Midura, J., and Klopfer, E. (2014). A framework for structuring learning assessment in a massively multiplayer online educational game: Experiment centered design. *International Journal of Game-Based Learning (IJGBL)*, 4(1):37–59.
- de Lima, R. M., de Medeiros Santos, A., Neto, F. M. M., de Sousa Neto, A. F., Leão, F. C. P., de Macedo, F. T., and de Paula Canuto, A. M. (2016). A 3d serious game for medical students training in clinical cases. In *2016 IEEE International Conference on Serious Games and Applications for Health (SegAH)*, pages 1–9. IEEE.
- Del Blanco, A., Torrente, J., Marchiori, E. J., Martínez-Ortiz, I., Moreno-Ger, P., and Fernandez-Manjon, B. (2010). Easing assessment of game-based learning with and lams. In *Proceedings of the second ACM international workshop on Multimedia technologies for distance learning*, pages 25–30.
- Falakmasir, M. H., González-Brenes, J. P., Gordon, G. J., and DiCerbo, K. E. (2016). A data-driven approach for inferring student proficiency from game activity logs. In

- Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 341–349.
- Georgiadis, K., Faber, T., and Westera, W. (2019a). Bolstering stealth assessment in serious games. In *International Conference on Games and Learning Alliance*, pages 211–220. Springer.
- Georgiadis, K., van Lankveld, G., Bahreini, K., and Westera, W. (2018). Accommodating stealth assessment in serious games: Towards developing a generic tool. In *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pages 1–4. IEEE.
- Georgiadis, K., van Lankveld, G., Bahreini, K., and Westera, W. (2019b). Learning analytics should analyse the learning: proposing a generic stealth assessment tool. In *2019 IEEE Conference on Games (CoG)*, pages 1–8. IEEE.
- Georgiadis, K., van Lankveld, G., Bahreini, K., and Westera, W. (2019c). Reinforcing stealth assessment in serious games. In *International Conference on Games and Learning Alliance*, pages 512–521. Springer.
- Harpstead, E., Myers, B. A., and Alevén, V. (2013). In search of learning: facilitating data analysis in educational games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 79–88.
- Kam, M., Kumar, A., Jain, S., Mathur, A., and Canny, J. (2009). Improving literacy in rural india: Cellphone games in an after-school program. In *2009 International Conference on Information and Communication Technologies and Development (ICTD)*, pages 139–149. IEEE.
- Kam, M., Ramachandran, D., Devanathan, V., Tewari, A., and Canny, J. (2007). Localized iterative design for language learning in underdeveloped regions: the pace framework. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1097–1106.
- Keele, S. et al. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, Citeseer.
- Min, W., Frankosky, M. H., Mott, B. W., Rowe, J. P., Smith, A., Wiebe, E., Boyer, K. E., and Lester, J. C. (2019). Deepstealth: Game-based learning stealth assessment with deep neural networks. *IEEE Transactions on Learning Technologies*, 13(2):312–325.
- Mislevy, R. J. and Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational measurement: issues and practice*, 25(4):6–20.
- Moreno-Ger, P., Burgos, D., Sierra, J. L., and Manjón, B. F. (2007). A game-based adaptive unit of learning with ims learning design and. In *European Conference on Technology Enhanced Learning*, pages 247–261. Springer.
- Mota, L. F. and Pimentel, E. P. (2017). Environment for students performance monitoring through a digital game for teaching mathematics. In *2017 Twelfth Latin American Conference on Learning Technologies (LACLO)*, pages 1–4. IEEE.
- Nyamsuren, E., Van der Vegt, W., and Westera, W. (2017). Automated adaptation and assessment in serious games: A portable tool for supporting learning. In *Advances in computer games*, pages 201–212. Springer.

- Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., Brown, R. A., and Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*, 76:617–630.
- Thomas, P., Labat, J.-M., Muratet, M., and Yessad, A. (2012). How to evaluate competencies in game-based learning systems automatically? In *International Conference on Intelligent Tutoring Systems*, pages 168–173. Springer.
- Vygotsky, L. S. (1997). *The collected works of LS Vygotsky: Problems of the theory and history of psychology*, volume 3. Springer Science & Business Media.