



PUC

ISSN 0103-9741

Monografias em Ciência da Computação
nº 08/2018

**Entrevistas e mapeamento sistemático sobre
análise de *logs* de interação de alunos em
ambientes virtuais de aprendizagem**

André Luiz de Brandão Damasceno

Dalai dos Santos Ribeiro

Sérgio Colcher

Simone D. J. Barbosa

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900

RIO DE JANEIRO - BRASIL

Entrevistas e mapeamento sistemático sobre análise de *logs* de interação de alunos em ambientes virtuais de aprendizagem

André Luiz de Brandão Damasceno, Dalai dos Santos Ribeiro,
Sérgio Colcher, Simone D. J. Barbosa

Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro – Rio de Janeiro, RJ – Brasil

{adamasceno,dribeiro,colcher,simone}@inf.puc-rio.br

Abstract. Online Education has broadened the avenues of research on student’s behavior and performance. In this paper, we shed light on how logs of students’ behavior in virtual learning environments are analyzed. We conducted interviews with teachers and a systematic mapping of 109 papers on Education Data Mining and Learning Analytics, considering papers published between 2010 and 2018. The main contributions of this paper are on the discussion and analysis of the results that can direct and ground future researches, as the development of tools to evaluate logs of interaction on videos and analyze logs of teachers.

Keywords: Learning Analytics, Educational Data Mining, MOOC, VLE, e-learning, engagement, behavior, interaction, mapping.

Resumo. O ensino a distância tem ampliado as possibilidades de pesquisa em comportamento e desempenho dos estudantes. Neste artigo apresentamos um estudo sobre análise de logs dos alunos em ambientes virtuais de aprendizagem. Para isto, conduzimos entrevistas com professores e realizamos um mapeamento sistemático de 109 artigos das áreas de Educational Data Mining e Learning Analytics, publicados entre 2010 e 2018. A principal contribuição desse trabalho está nas discussões e análises dos resultados que podem direcionar e fundamentar investigações futuras, como a criação de ferramentas para avaliação de logs de interação em vídeos e análise de logs de professores.

Palavras-chave: Learning Analytics, Educational Data Mining, MOOC, AVA, e-learning, engajamento, comportamento, interação, mapeamento

In charge for publications:

Rosane Teles Lins Castilho

Assessoria de Biblioteca, Documentação e Informação

PUC-Rio Departamento de Informática

Rua Marquês de São Vicente, 225 - Gávea

22453-900 Rio de Janeiro RJ Brasil

Tel. +55 21 3114-1516 Fax: +55 21 3114-1530

E-mail: bib-di@inf.puc-rio.br

Web site: <http://bib-di.inf.puc-rio.br/techreports/>

Sumário

1	Introdução	1
2	Entrevistas com Professores de EaD	1
3	Mapeamento Sistemático das áreas de EDM e LA	4
4	Triangulação dos Resultados e Considerações Finais	7
	Referências	9

1 Introdução

A modalidade de ensino a distância (EaD) fez surgir novas oportunidades de pesquisa para a área de Informática na Educação. Toda interação dos estudantes em Ambientes Virtuais de Aprendizagem (AVAs) pode ser registrada em forma de *logs*. A análise desses *logs* permitem identificar padrões de comportamento, avaliar a aprendizagem e prever o desempenho dos estudantes [Romero and Ventura 2010].

As duas principais comunidades com objetivos e interesses similares na exploração de dados nessa área são: a *International Educational Data Mining Society* (IEDMS)¹ e a *Society for Learning Analytics* (SLA)², que definem suas áreas de pesquisa como *Educational Data Mining* (EDM) e *Learning Analytics* (LA), respectivamente. Ambas as áreas utilizam detecção automática de informação e análise humana através de visualizações e outros métodos. Contudo, a EDM utiliza a análise humana como ferramenta para melhorar a detecção automática de informação, enquanto a LA busca potencializar a análise humana através de ferramentas que realizam a detecção automática de informação [Siemens and Baker 2012]. Em outras palavras, LA direciona seus esforços para melhorar a visualização e interpretação humana dos dados, e EDM tem maior foco no aumento e aprimoramento dos métodos de detecção automática de informação.

Levando em consideração os últimos três anos, encontramos na literatura alguns trabalhos de revisão do estado arte em EDM e LA [Dutt et al. 2017, Na and Tasir 2017, Sergis and Sampson 2017, Vieira et al. 2018]. No entanto, esses artigos apresentam revisão de apenas uma das áreas (EDM ou LA) e levam em consideração somente um aspecto, como: intervenção durante o aprendizado, suporte ao professor, clusterização ou visualização de dados.

Nesse contexto, este trabalho busca identificar quais informações dos alunos os professores consideram relevantes (*e.g.*, desempenho, comportamento, motivação), como essas informações são extraídas e se há algum requisito para melhorar essa análise. Para isto, conduzimos entrevistas com professores que atuam no Brasil e realizamos um mapeamento sistemático das áreas de EDM e LA, com o objetivo de levantar os trabalhos que utilizam *logs* para analisar e prever o desempenho e o comportamento dos alunos. Em seguida, relacionamos as respostas obtidas nas entrevistas com os professores e os resultados dos trabalhos levantados no mapeamento sistemático.

O artigo está organizado como a seguir. A Seção 2 detalha a condução das entrevistas com os professores, assim como seus resultados. Na Seção 3 apresentamos a metodologia utilizada no mapeamento e as respostas obtidas às questões de pesquisa. A Seção 4 apresenta a discussão dos resultados do trabalho e as considerações finais.

2 Entrevistas com Professores de EaD

Entre os meses de novembro de 2017 e abril de 2018, realizamos entrevistas com professores utilizando um roteiro de perguntas estabelecido previamente³. A amostra selecionada consistiu de 18 professores universitários de instituições localizadas em seis estados do Brasil (RJ, MA, MG, PI, GO e PE). Todos esses professores informaram possuir experiência

¹<http://educationaldatamining.org/>

²<https://solaresearch.org/>

³<https://goo.gl/nXuui4>

no ensino a distância, sendo 9 deles em cursos inteiramente a distância, 1 em cursos do tipo *blended learning* e 8 nas duas modalidades. Os 18 professores citaram que haviam tido experiência com 10 AVAs diferentes, dentre os quais o Moodle foi o mais mencionado (por 15 professores).

A Tabela 1 sumariza os resultados das entrevistas. Em relação aos recursos usados nos AVAs, todos os professores informaram utilizar alguma ferramenta de comunicação com os alunos. Além disso, a maioria passa atividades para os alunos e posta materiais e notícias nos AVAs. Apesar de apenas um professor ter informado utilizar uma ferramenta para monitoramento de acesso dos alunos no AVA, 11 professores disseram que analisam os *logs* de acesso e 9 visualizam os *logs* de interação dos alunos com o conteúdo.

Tabela 1: Resultados gerais das entrevistas.

Prof.	Recursos Usados nos AVAs	Como analisa o uso que o aluno faz do conteúdo	Informações relevantes pra saber do aluno	Padrão de comportamento	Avaliação da motivação do aluno	Fatores que influenciam o aluno assistir uma videoaula	Interação e desempenho do aluno
01	Ferramentas de Comunicação (e.g., fórum, chat) Postagens (e.g., material, notícias) Atividades (e.g., quiz, exercícios) Repositório de conteúdo Referenciais materiais da Web Videconferência Wiki Blog Badges Monitoramento de acesso Outros	Análise de logs de acesso Análise de logs de interação com o conteúdo Não analisa Análise de postagens e participação no fórum Avalia atividades e provas Outros (e.g., feedback de tutores) Questionário de opinião Observa quem está online	Background Interação Desempenho Aprendizado Feedback Expectativas, Intenções e Motivações do aluno Utilidade do conhecimento Perfil Percepção Idade Dificuldades Participação Objetivos e Expectativas atingidas Desistências Outros (e.g., problemas de conexão, disponibilidade, repetência)	Acesso Interesse Fórum Interação Participação Desempenho Responsabilidade Outros (e.g., Materiais que gestou, Ritmo, Reação, Humor)	Comunicação (e.g., fórum, chat) Atividades Acesso Reação Participação Não sabe Desempenho Feedback (e.g., questionário qualitativo) Badges Interação Outros (e.g., uso de vídeo, reconhecimento facial)	Duração do vídeo Maneira como o professor se apresenta e se expressa no vídeo Formato como conteúdo é apresentado Acesso maior com a proximidade da prova Produção e Edição do vídeo Interatividade e o uso de imagens que representem o que está sendo abordado Conteúdo do vídeo Poder assistir o vídeo via <i>mobile</i>	Acesso aos materiais Participação dos alunos no fórum Atividades feitas Acesso do vídeo Não tem acesso a essa informação
Total	18 10 11 4 4 3 2 1 1 7	11 9 6 5 3 2 1	16 14 12 12 10 8 7 6 4 3 3 2 2 2 3	8 6 5 4 4 2 2 4	13 7 5 3 3 3 2 2 2 2	15 8 8 7 7 5 3 1	14 7 5 3 2

Perguntamos aos professores quais informações dos alunos eles consideram relevantes. Eles mencionaram dados como interações dos alunos com os AVAs, desempenho, expectativas dos alunos e *feedbacks* do que está bom e sobre o que precisa melhorar no curso. A maioria dos professores consideraram importante saber o *background* de cada aluno, como por exemplo, o quanto ele já sabe do conteúdo e a sua formação. Segundo eles, isso ajudaria na organização da disciplina e na seleção do nível de dificuldade do conteúdo. Perguntamos também como eles avaliam a aprendizagem e a motivação do aluno. Um total de 15 professores responderam que utilizam o método de provas. Já em relação a motivação, 13 falaram que analisam a comunicação dos alunos, como o uso dos fóruns, chats e as contribuições e questionamentos que os alunos fazem. Alguns professores também disseram que avaliam a motivação observando quem realiza qual atividade.

Como dito anteriormente, apenas um professor disse utilizar uma ferramenta que ana-

lisa os *logs* de acesso dos alunos. A maioria dos professores, quando questionados sobre padrões de comportamento dos alunos, responderam que não possuem ferramentas de análise e que só conseguem visualizar os *logs* dos estudantes, o que torna inviável analisar a interação de uma turma em cursos cujo número de alunos é alto. Por isso, 12 professores apresentaram requisitos de ferramentas que analisassem, por exemplo, *logs* de acesso e interação dos alunos, tanto com os materiais, como com os recursos dos AVAs como fórum, *player* de vídeo, *quiz* e *ebooks*. Segundo eles, essas análises permitem criar sistemas de identificação de padrões de interação de grupos de alunos e predição de evasão e desempenho. Além disso, 2 professores mencionaram que essas informações devem ser apresentadas de maneira simples (*e.g.*, num *dashboard*), para eles dedicarem mais tempo nos métodos de ensino, pois hoje em dia o professor gasta muito tempo corrigindo exercícios, preparando e ministrando aulas. Os professores falaram que isso poderia ajudá-los a: (i) elaborar material de apoio de acordo com o perfil e desempenho de cada aluno; (ii) saber quando e quais materiais adicionar no AVA; (iii) entender por que certos alunos abandonam o curso; (iv) relacionar o desempenho de um aluno com a turma; e (v) tomar decisões pedagógicas para aumentar o desempenho e reduzir a evasão de alunos.

Dentre os 18 professores, somente 1 respondeu nunca ter construído conteúdo educacional online, porque ele sempre desempenhou o papel de tutor e o material do curso é elaborado pelo professor conteudista. Além disso, 14 professores relataram ter experiência na autoria de videoaulas. Alguns deles ressaltaram a importância do uso de vídeos na educação por ser esse o formato de conteúdo que os alunos mais gostam de acessar, por permitir que a explicação de um assunto possa ser vista mais de uma vez, sendo a videoaula, muitas vezes, a porta de entrada para o conhecimento. Quando perguntados sobre informações de interação dos alunos com as videoaulas, a maioria dos professores (17 ao total) disseram querer saber dados de acesso aos vídeos, como estatísticas de visualização de certos trechos dos vídeos, que trechos foram pulados e quais trechos os alunos assistiram mais de uma vez. Segundo eles, essas informações podem sinalizar ao professor o que os alunos acharam relevante e sobre quais trechos os alunos eventualmente tiveram mais dúvidas.

Os professores citaram como fatores que influenciam o aluno a assistir uma videoaula: (i) a proximidade da prova, (ii) a produção e a edição do vídeo, (iii) o formato de exibição do conteúdo e (iv) a maneira como o professor se apresenta e se expressa na videoaula. A duração do vídeo foi o fator mais mencionado por 15 professores. Segundo eles, o aluno não costuma assistir vídeos longos e, quando acessam, não assistem o vídeo todo. Contudo, não houve consenso sobre a duração ideal de uma videoaula. As respostas variaram de 5 a 30 minutos de duração máxima.

Sobre a relação entre o uso que o aluno faz no AVA e o seu desempenho no curso, 14 professores informaram que os acessos dos alunos aos materiais estão relacionados com suas notas, 7 disseram que os alunos que mais participam (*i.e.*, usam o fórum) são os que apresentam melhor desempenho e 5 utilizam o critério de conclusão das atividades como indício de que o aluno terá um bom desempenho no final do curso. Também perguntamos aos professores se eles já identificaram alguma relação entre a interação e evasão dos alunos nos AVAs. Metade dos professores informaram não terem identificado relação e 7 disseram terem percebido uma correlação entre o acesso dos alunos e a conclusão do curso. Outras 10 respostas foram fornecidas para essa pergunta, mas nenhuma delas foi mencionada por mais de 2 professores.

Ao final das entrevistas, alguns professores disseram que gostariam de relacionar o desempenho e a evasão com a interação dos alunos nos AVAs, mas que seria necessário haver uma ferramenta para isso. Eles sugeriram ferramentas que analisassem os *logs* de interação dos alunos nos fóruns, capturassem a forma como o aluno reage ao conteúdo do curso, detectassem padrões de navegação dos alunos nos materiais e recursos dos AVAs e identificassem a relação entre o acesso dos alunos e a taxa de desistência. Alguns deles também ressaltaram a importância da visualização desses dados por meio de *dashboards* que apresentassem, por exemplo, o resumo semanal da turma e de cada aluno (quem acessou, quem participou do fórum ou chat, quem enviou a atividade). Um deles destacou a dificuldade de se fazer alguma análise nesse sentido, porque, mesmo utilizando os filtros disponíveis na plataforma, os *logs* são apresentados de maneira misturada ou confusa.

3 Mapeamento Sistemático das áreas de EDM e LA

O procedimento utilizado na realização do mapeamento sistemático teve como referência o método proposto por Kitchenham and Charters 2007, para avaliar as pesquisas relevantes do domínio seguindo um protocolo bem definido a fim de evidenciar os principais problemas, objetivos, métodos, estudos de caso e resultados obtidos nos trabalhos levantados. O primeiro passo foi definir a questão de pesquisa que norteou o mapeamento:

- Quais métodos são utilizados para analisar, entender e prever o desempenho e o comportamento dos alunos?
- Quais ferramentas (*e.g.*, *dashboard*) são utilizadas pelos professores para analisar os *logs* de interação dos estudantes nos sistemas de EaD?

Em seguida, elaboramos a *string* de busca a partir de uma combinação de palavras-chave relacionadas com as questões de pesquisa: (*education OR course OR MOOC OR “e-learning” OR teaching OR “virtual learning environments” OR “virtual training environments” OR “learning management system” OR LMS*) **AND** (*engagement OR behavior OR behaviour*) **AND** (*analysis OR analyses OR analytics OR analytic OR visualisation OR visualization OR “data mining” OR “learning analytics”*).

Utilizamos as máquinas de buscas avançadas de três bibliotecas digitais: ACM⁴, Elsevier⁵ e IEEE⁶. Durante esse processo, definimos em todas as bibliotecas digitais que os artigos deveriam ter o ano de publicação maior que 2009, estar no formato PDF e terem sido escritos em inglês ou português. Conforme resume a Figura 1, o processo de seleção de artigos seguiu quatro passos: (1) a busca nas bibliotecas digitais retornou 1710 artigos; (2) removemos os artigos duplicados, ficando com 1427 artigos; (3) analisamos os títulos e *abstracts* de cada artigo, descartando os artigos conforme os critérios de inclusão e exclusão, resultando em 264 artigos; (4) lemos 264 artigos na íntegra, novamente descartando os artigos conforme os critérios de inclusão e exclusão. Ao final desse processo, restaram 109 artigos, reportados neste trabalho.

Os critérios de inclusão e exclusão nos ajudaram a descartar os artigos que, embora contivessem as palavras-chaves definidas na *string* de busca, não contribuíam para responder as questões de pesquisa. Os critérios foram:

⁴<http://dl.acm.org/advsearch.cfm>

⁵<http://sciencedirect.com/science/search>

⁶<http://ieeexplore.ieee.org/search/advsearch.jsp>

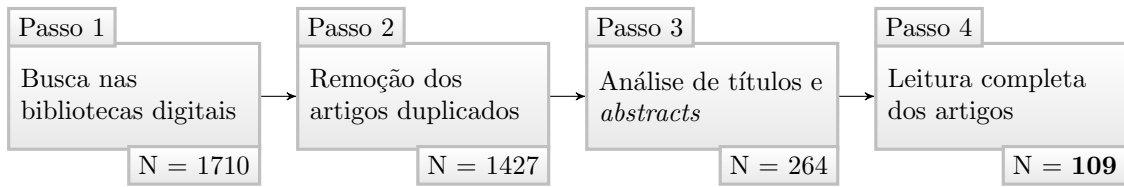


Figura 1: Processo iterativo de seleção dos artigos e resultados de cada etapa.

Inclusão:

1. Trabalhos que apresentam/discutem soluções/metodologias/estudos de caso relacionadas a análise de dados (*e.g.*, *logs*) para mensurar o desempenho, motivação, participação ou evasão dos alunos em sistemas de EaD (*e.g.*, MOOCs, AVAs);
2. Trabalhos que apresentam a detecção de padrões de comportamento dos estudantes em sistemas de EaD ou acervos de videoaula;
3. Trabalhos que mostram soluções/metodologias/estudos de visualização dos *logs* dos alunos em sistemas de EaD ou acervos de videoaula;
4. Trabalhos que avaliam problemas de interação dos estudantes em sistemas de EaD ou acervos de videoaula.

Exclusão:

1. Chamadas de trabalho ou *keynotes*;
2. Trabalhos que enfocam o ensino presencial ou na recomendação de cursos;
3. Trabalhos que objetivam melhorar a acessibilidade para pessoas com deficiência;
4. Trabalhos que apresentam resultados relacionados à análise de *emoticons*;
5. Trabalhos que analisam dados provenientes apenas de questionários;
6. Trabalhos que não foram escritos em português ou inglês;
7. Trabalhos que não apresentam resultados.

A Tabela 2 apresenta a distribuição por ano dos artigos encontrados. Vale observar que realizamos as buscas no dia 21 de fevereiro de 2018, o que explica termos encontrado apenas 1 artigo publicado neste ano. A Tabela 3 lista as fontes de publicações (congressos, simpósios e periódicos) com 5 ou mais artigos selecionados no nosso mapeamento.

Tabela 2: Distribuição dos artigos encontrados por ano e biblioteca digital.

	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
ACM	0	2	3	6	14	13	16	21	0	75
Elsevier	1	0	0	3	2	6	2	3	1	18
IEEE	1	0	0	0	3	2	2	8	0	16
Total	2	2	3	9	19	21	20	32	1	109

Tabela 3: Fonte das publicações dos artigos.

Fonte da publicação	Quantidade
International Conference on Learning Analytics and Knowledge – LAK	29
Proceedings of the ACM Conference on Learning @ Scale - L@S	21
Computers in Human Behavior	13
IEEE Frontiers in Education Conference (FIE)	5
Outras	41

Conforme pode ser visto na Tabela 4, agrupamos os artigos por problemas, objetivos, casos de estudo e resultados em comum. A maioria desses artigos abordam os problemas de análise de comportamento e performance dos alunos. Como consequência, o principal objetivo desses artigos é identificar padrões de comportamento e performance dos estudantes. A maior parte desses artigos tiveram como estudo de caso disciplinas relacionadas a computação. Cerca de 13 artigos relataram que seus casos de estudo eram de múltiplas áreas, mas sem especificar quais.

Agrupamos os artigos também por métodos utilizados. Contudo, não apresentamos esses métodos neste artigo devido a limitações de espaço. Ao todo, foram encontrados 50 métodos diferentes. A maioria desses artigos apresentam trabalhos utilizando métodos de clusterização (*e.g.*, K-means) e predição (*e.g.*, Regressão logística, Árvore de decisão). Encontramos também 14 trabalhos que relatam o uso de indicadores definidos pelos próprios autores e que fazem correlações entre o uso de ferramentas de discussão (*e.g.*, fórum, chat) e as notas obtidas pelos alunos. Há 13 artigos que correlacionam dados de questionários qualitativos com a interação/performance do aluno.

Quanto aos 115 tipos de resultados apresentados, agrupamos os artigos em 14 tópicos, também apresentados na Tabela 4. Um total de 46 artigos mostram que é possível agrupar (*cluster*) os estudantes com base nos seus padrões de acesso e interação. Em relação a performance dos estudantes e ao uso do fórum, destacamos resultados como: modelos de predição de performance dos alunos; correlações entre acesso, atividades realizadas, participação e performance dos alunos; e evidências de que (i) materiais apresentados nos cursos a distância são ignorados até mesmo por estudantes com performance satisfatória; e (ii) estudantes que mais usam o fórum tendem a ter uma boa performance no curso. Em relação ao tópico de Ferramentas, classificamos os resultados dos artigos encontrados em 4 categorias diferentes, cujas principais (*i.e.*, com 5 ou mais artigos) são: sistemas que podem ser usados para fornecer análise e intervenção na aprendizagem do estudante; ferramentas que ajudam o professor a analisar o comportamento do estudante; e artigos que sugerem o uso de *logs* de AVAs para o professor monitorar o comportamento, motivação e performance dos estudantes.

A maioria dos artigos que apresentam resultados relacionados ao tópico de Conclusão de curso mostram que a taxa de conclusão em MOOCs é baixa e que a realização de atividades pode ser utilizada como preditor de conclusão. Engajamento é um termo que aparece em 15 artigos e é mensurado pela maioria deles como: número de atividades realizadas nos AVAs; número de materiais acessados; e número de acessos, interações e postagens em fóruns. Alguns artigos também relacionam engajamento a: o tempo de visualização de vídeos; e o número de notas de aulas dos alunos. Em relação ao tópico de Vídeos, foram encontradas 22 categorias de resultados. Contudo, a categoria que agrupa

os artigos que relatam que os estudantes não costumam assistir a videoaula inteira é a única que possui mais de 5 artigos. As outras categorias possuem apenas 1 ou 2 artigos.

Em relação ao tópico de Desistência, todos os artigos buscam prever a evasão do aluno. Contudo, apenas 4 artigos apresentam modelos de previsão através da análise de *logs* de interação dos estudantes. Outras 6 categorias de resultados foram atribuídas a esse tópico. Entretanto, todas elas possuem no máximo 2 trabalhos.

Os artigos do tópico de Alunos disciplinados (relacionado ao termo em inglês *self-regulated*) afirmam que alunos indisciplinados tendem a ter uma performance insatisfatória no curso. Além disso, há 2 artigos relatando que alunos disciplinados não necessariamente obtêm bons resultados no final curso. Os 2 artigos do tópico de Personalidade apresentam resultados relacionados à identificação da personalidade do aluno através de seus *logs* de interação. O tópico Participação apresenta resultados relacionados à detecção de participação do estudante e à correlação de participação com dados dos alunos (*e.g.*, nacionalidade e escolaridade). Os 2 artigos do tópico *Log* de cliques afirmam que o aumento e a redução do número de cliques no AVA estão relacionados com a probabilidade de passar e falhar no curso, respectivamente. Já o tópico de Intenção está relacionado com 2 artigos, que relacionam a intenção do estudante no AVA, coletada por meio de questionários, com seu comportamento e performance no curso. Por fim, o tópico Outros apresenta 5 categorias, nas quais apenas 1 possui mais de 5 artigos, que afirmam que o número de acesso nos AVAs aumenta perto de provas e datas de entrega de trabalho.

4 Triangulação dos Resultados e Considerações Finais

As seções anteriores apresentam os resultados obtidos nas entrevistas com os professores e no mapeamento sistemático das áreas de EDM e LA. Notamos que há relações entre esses resultados. Devido à limitação de espaço, vamos destacar as principais relações, referenciando o número do professor da Tabela 1 (no formato *P99*) com o número do artigo da Tabela 4 (no formato *A99*).

P04 constatou que pessoas mais velhas interagem com os AVAs de forma diferente das pessoas mais novas, o que corrobora os resultados de A19, que afirma que estudantes mais velhos tendem a obter melhor performance, a participar mais nos fóruns e a realizar mais *backjumps* na *timeline* dos vídeos do que estudantes mais novos. Ainda segundo P04, ele utiliza técnicas de gamificação para motivar o aluno e já notou uma melhora da turma depois que passou a utilizar *badges* e *rankings* de desempenho. Isto está em linha com A30, que identificou uma melhora no engajamento dos alunos quando comparou cursos de EaD que utilizam e que não utilizam *badges*.

P02, P03 e P05 disseram identificar o interesse dos alunos através das interações dos alunos no fórum. De acordo com P09, o uso do fórum pelo aluno é um sinal de que ele está ativo no curso. Relacionados com essas respostas estão resultados de que os estudantes que mais postam no fórum são os mais prováveis de completar os cursos (A42, A66, A102 e A107) e que tanto os estudantes que mais iniciam tópicos nos fóruns (A41, A52, A80 e A92), quanto os que respondem a esses tópicos (A42, A52 e A80) tendem a ter uma boa performance. O aumento dos acessos em recursos e materiais dos AVAs (*e.g.*, vídeos) em períodos próximos às provas ou às datas de entrega de trabalhos também foram observados tanto por P01, P02, P03, P04, P06, P12 e P13, como por A8, A11, A20, A27, A88 e A90. Complementando esse resultado, A80 afirma que o número de postagens nos fóruns cresce

nesses mesmos períodos.

Em relação ao uso de vídeos, a duração foi apontada pela maioria dos entrevistados (exceto P02, P08 e P13) como um dos principais fatores que influenciam o aluno a assistir uma videoaula. Essa observação está alinhada com os resultados encontrados por A17 e A59, que afirmam que vídeos curtos proporcionam mais engajamento. Além disso, A59 encontrou uma correlação entre a diminuição da retenção de audiência com a duração do vídeo. Outro resultado relevante foi identificado por P18, A66 e A76, que encontraram uma correlação entre a quantidade de tempo que os alunos passam vendo os vídeos e o desempenho obtidos por eles.

Há indícios de que a retenção dos alunos nos vídeos está relacionada com a autoria do vídeo. P01 acredita que vídeos no formato de slides não obtêm um bom resultado. P01, P06, P09 e P16 já constataram que há pouco engajamento dos alunos em vídeos mais teóricos, sem demonstração ou resolução de exercícios. Corroborando com esses indícios, A17 verificou que os alunos assistem mais tutoriais múltiplas vezes, e A59 verificou que os vídeos de teoria retêm menos alunos do que os vídeos de demonstração de código. Além disso, o baixo engajamento foi observado também em vídeos que os instrutores falam devagar (P04, P12 e A17).

P01 e A17 observaram que filmagens de aulas presenciais têm pouco engajamento dos alunos. A17 também revela que os alunos se engajam mais com vídeos filmados informalmente com o professor sentado em sua mesa, quando comparados com vídeos produzidos no formato de produções de alto custo de TV. Esse resultado apoia a resposta de P03, que acredita que a super-produção de um vídeo tem que ser justificável. Além disso, A17 sustenta o que foi dito por P01, P18 e P20, que consideram o formato de vídeo *Khan academy* e *Talking-head* como os que mais favorecem o engajamento dos alunos.

A análise das entrevistas e dos artigos revela lacunas a serem preenchidas. Encontramos na literatura 20 ferramentas para auxiliar o professor na análise de *logs*. Porém, nenhuma delas preenche todos os requisitos levantados pelos professores, conforme descrito na Seção 2. Isso é evidente quando observamos os trabalhos de interações com vídeos em que não há nenhuma ferramenta de análise de *logs*. Há também uma lacuna em relação à análise de *logs* dos professores. O que observamos é que todos os trabalhos encontrados analisam apenas os dados dos alunos. Como trabalho futuro, buscamos investigar se há melhora no desempenho da turma quando os professores utilizam ferramentas de clusterização e predição de desempenho dos alunos.

A limitação do mapeamento sistemático executado nesse trabalho é similar a qualquer outro trabalho de revisão da literatura. Provavelmente existem trabalhos importantes que não foram incluídos no nosso mapeamento, tais como teses e dissertações, livros ou até mesmo artigos que não tenham sido encontrados com o uso da nossa *string* de busca nas bibliotecas digitais selecionadas. Para contornar esse problema, analisamos todos os trabalhos que foram referenciados e que fazem referência aos 109 artigos encontrados no nosso mapeamento, numa estratégia de *snowballing*. Aplicando os mesmos critérios de inclusão e exclusão sobre os títulos e *abstracts*, obtivemos mais 194 artigos, cuja leitura completa deixamos para trabalho futuro.

Referências

- [Dutt et al. 2017] Dutt, A., Ismail, M. A., and Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, 5:15991–16005.
- [Kitchenham and Charters 2007] Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
- [Na and Tasir 2017] Na, K. S. and Tasir, Z. (2017). A Systematic Review of Learning Analytics Intervention Contributing to Student Success in Online Learning. In *2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, pages 62–68. IEEE.
- [Romero and Ventura 2010] Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618.
- [Sergis and Sampson 2017] Sergis, S. and Sampson, D. G. (2017). Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review. pages 25–63.
- [Siemens and Baker 2012] Siemens, G. and Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge, LAK '12*, pages 252–254, New York, NY, USA. ACM.
- [Vieira et al. 2018] Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122:119–135.

Tabela 4: Problemas, objetivos, casos de uso e resultados gerais dos artigos.

Ano	N	DOI/ISBN	Problemas										Objetivos										Casos de Estudo										Resultados																																																																											
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106
2010	1	10.1016/j.eswa.2010.03.032																																																																																																										
2011	4	10.1145/2060116.2060128																																																																																																										
2012	8	10.1145/2469296.2469324																																																																																																										
2013	12	10.1145/2469296.2469330																																																																																																										
2014	26	10.1145/2566325.2566250																																																																																																										
2015	46	10.1145/2787922.2787731																																																																																																										
2016	76	10.1145/2883851.2883826																																																																																																										
2017	99	10.1145/3027385.3027448																																																																																																										
2018	109	10.1016/j.chb.2017.11.011																																																																																																										
Total			44	30	22	19	12	8	7	4	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2																								