

Mineração de Processo Educacional - uma Revisão Sistemática da Literatura

Francisco Genivan Silva^{1,2}, Thiago Reis da Silva^{1,3}, Eduardo H. S. Aranha¹

¹Programa de Pós-Graduação em Sistemas Computacionais – PPgSC
Universidade Federal do Rio Grande do Norte – UFRN
Laboratório Gamedu.net

²Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte – IFRN

³Instituto Federal de Educação, Ciência e Tecnologia do Maranhão – IFMA

genivan.silva@ifrn.edu.br, thiago.reis@ifma.edu.br,
eduardoaranha@dimap.ufrn.br

Abstract. Introduction: Educational Processes Mining (EPM), a research field of Educational Data Mining (EDM) has shown great potential to analyze, evaluate and improve processes models in educational contexts. **Objectives:** To establish solid evidence based on existing work related to use of MPE. **Method:** a systematic literature review was conducted to integrate the results from several previously published studies. **Results:** of the 463 primary studies, 14 were included in this study. **Conclusion:** 86% of studies reported benefits achieved through the use of mining processes, which indicates significant advancements brought into the field of education.

Resumo. Introdução: a Mineração de Processos Educacionais (MPE), campo de pesquisa da Mineração de Dados Educacionais (MDE), tem apresentado grande potencial em analisar, avaliar e aperfeiçoar modelos de processos em contextos educacionais. **Objetivos:** estabelecer evidências sólidas baseadas em trabalhos existentes relacionados a utilização de MPE. **Método:** foi realizada uma revisão sistemática da literatura para integrar os resultados provenientes de diversos estudos publicados anteriormente. **Resultados:** dos 463 estudos primários, 14 foram incluídos nesse estudo. **Conclusão:** 86% dos estudos relatam benefícios alcançados através da utilização de Mineração de Processos, o que indica avanços significativos trazidos para o campo da educação.

1. Introdução

A utilização cada vez mais frequente de sistemas de gestão acadêmicos, e o aumento na procura e oferta de cursos a distância e do uso de sistemas *e-learning* tem gerado uma quantidade significativa de dados educacionais. O aumento nos dados gera a necessidade de uma abordagem automática. Nos últimos anos, a utilização de metodologias de extração automática de informações úteis de grandes volumes de dados, em especial a Mineração de Dados, tem transformado diversos campos de conhecimento. O estudo e a aplicação dessas metodologias no campo da educação têm se mostrado promissores e deu origem a novas áreas de investigação científica, como por exemplo, Mineração de Dados Educacionais (MDE) e Mineração de Processos Educacionais (MPE).

Técnicas de MDE [Manhães *et. al.*, 2011] têm sido amplamente aplicadas para se encontrar padrões e se construir modelos descritivos e preditivos de grandes volumes de dados através da utilização de diferentes sistemas de informação e Ambientes Virtuais de Aprendizagens. Os resultados da MDE são usados para se obter uma melhor compreensão da aprendizagem, gerar recomendações e permitir a realização de melhorias [Romero e Ventura 2007]. Estes dados podem ser analisados a partir de vários níveis e perspectivas, mostrando diferentes aspectos do processo educacional [Trčka e Pechenizkiy 2009]. No entanto, a maioria das técnicas tradicionais de MDE não incide sobre o processo como um todo. Eles não visam explicitamente, por exemplo, a descoberta, análise, ou representação visual do processo educacional completo. Esse espaço, porém, é preenchido pela MPE, que visa extrair conhecimentos relacionados com o processo expresso por *logs* de eventos gravados por um sistema de informação.

De maneira geral, as técnicas de MPE visam: (i) a descoberta de modelos de processos educacionais completos e compactos, capazes de expressar o comportamento observado dos usuários; (ii) análise de conformidade, verificando se um comportamento previamente modelado corresponde ao comportamento observado; (iii) melhoria de processos, extraíndo conhecimento e facilitando uma melhor compreensão dos problemas e das oportunidades de otimização [Trcka e Pechenizkiy 2009].

Em suma, segundo Aalst (2012, p.11) “a Mineração de Processos é uma ferramenta importante para organizações modernas que necessitam gerenciar processos operacionais não triviais”. Neste contexto, o objetivo deste trabalho foi o desenvolvimento de uma Revisão Sistemática da Literatura (RSL), a fim de estabelecer evidências sólidas baseadas em trabalhos existentes relacionados à utilização de Mineração de Processos voltada para a educação. Uma RSL é uma pesquisa em profundidade de um fenômeno de interesse que produz resultados específicos e detalhados por meio da análise de conteúdo e qualidade do material pesquisado [Kitchenham *et. al.*, 2009]. A RSL foi utilizada visando extrair, catalogar, analisar e sintetizar dados de vários artigos publicados em congressos, jornais e periódicos internacionais, que abordem a temática da revisão [Silva, Medeiros e Aranha 2014].

Além desta seção introdutória, este artigo encontra-se organizada da seguinte forma. Na Seção 2 é apresentado o método utilizado para a RSL, as questões de pesquisa, o processo de busca e os critérios de inclusão e exclusão. Já a Seção 3 aborda os resultados gerais obtidos a partir da execução da RSL e, por fim, na Seção 4, são apresentadas as conclusões finais do trabalho.

2. Método Utilizado

Esta RSL está estruturada com base nas diretrizes originais propostas por Kitchenham *et. al.*, (2009). Seu objetivo principal é identificar e avaliar estudos relevantes, com o intuito de resumir e encontrar evidências existentes sobre pesquisas relacionadas a MPE. Como objetivo secundário, esta RSL visa esclarecer conceitos e aspectos importantes relacionados às publicações neste campo de pesquisa, especialmente acerca das técnicas de Mineração de Processos utilizadas, os algoritmos empregados e os principais benefícios do emprego dessas técnicas no contexto educacional. A partir dos objetivos

expostos e seguindo um protocolo planejadamente estruturado, a pesquisa segue o método detalhado nas próximas seções.

A RSL realizada neste trabalho visa a responder às Questões de Pesquisa (QP) apresentadas na Tabela 1. A primeira coluna desta tabela apresenta o código da questão de pesquisa, e a segunda, a descrição da mesma. O período de tempo escolhido para esta RSL compreende os estudos publicados nos últimos quatro anos (2012 a 2015), ou seja, artigos publicados entre 1º de janeiro de 2012 a 31 de dezembro de 2015.

Tabela 1: Questões de Pesquisa.

Questões	Descrição
QP1	Quais são os objetivos da utilização da Mineração de Processo Educacional?
QP2	Em quais das subáreas (principais) as pesquisas em Mineração de Processos para Educação se encaixam?
QP3	Quais os principais algoritmos estão sendo utilizados pelos pesquisadores para extrair informações dos processos?
QP4	Quais benefícios e limitações estão sendo reportados pelos estudos?

2.1 Processo de Busca e *String* de Pesquisa

O processo de busca de estudos primários consistiu em utilizar busca automática nos engenhos de busca, IEEEExplore, ACM *Digital Library* e ScienceDirect. As palavras chaves utilizadas na *string* de busca para obtenção dos estudos primários desta RSL foram *Process Mining*, representando o tema, e *Education*, *Scholar*, *Academic* e *Student*, representando a área de aplicação. Portanto, a *string* definida foi:

(“process mining” AND education) OR (“process mining” AND student) OR (“process mining” AND academic) OR (“process mining” AND scholar)

Também foi executada uma busca manual em anais de conferências reconhecidamente relevantes para a área em questão, como o *International Conference on Educational Data Mining* (EDM), o *International Conference Learning Analytics and Knowledge* (LAK) e o *International Conference on Advances in Information Mining and Management* (IMMM).

2.2 Critérios de Inclusão e Exclusão

Uma abordagem em quatro etapas foi utilizada na seleção dos trabalhos relevantes para esta revisão. Na primeira etapa, uma busca utilizando a *string* de busca previamente definida foi realizada nas bases de dados já mencionadas. Esta etapa restringiu os artigos pelo título e ano de publicação e inicialmente foi aplicado um único Critério de Inclusão (CI) que limitava os estudos a:

- **CI1:** Trabalhos que abordam a utilização de Mineração de Processo e técnicas correlatas aplicados a dados originados em contextos educacionais.

Nas etapas seguintes, sendo estas a leitura e análise do resumo (etapa 2), seguida da introdução e conclusão (etapa 3) e por fim a leitura completa dos artigos (etapa 4), foram excluídos trabalhos que atendessem aos seguintes Critérios de Exclusão (CE):

- **CE1:** Artigos que não apresentem dados empíricos (ou seja, artigos apenas conceituais, ou com demonstração de ferramentas, etc.);
- **CE2:** Artigos não completos (*short papers*);

- **CE3:** Capítulos de livros;
- **CE4:** Estudos terciários e outras revisões ou mapeamentos sistemáticos.

O critério de inclusão CII foi utilizado em todas as etapas do processo para garantir a manutenção do escopo e relevância dos estudos para a revisão. É importante também ressaltar que, em alguns casos, a leitura de outras partes dos artigos foi realizada de maneira antecipada para aumentar o grau de confiança da seleção. Além disso, estudos semelhantes que apresentavam a evolução de uma mesma pesquisa e que continham os mesmos autores tiveram suas versões mais antigas removidas para se evitar duplicações.

2.3 Processo de Extração de Dados

Após a leitura completa de um artigo, foram extraídos os seguintes dados: objetivos do estudo; método do estudo (por exemplo, experimento controlado, relato de experiência, estudo de caso, entre outros); como os dados foram coletados e analisados durante o estudo; técnicas de EPM utilizadas, *plugins* ou ferramentas desenvolvidas; os benefícios, conclusões e dificuldades encontradas. Todos os artigos foram lidos e tiveram seus dados extraídos por dois revisores, enquanto um terceiro revisor supervisionou o trabalho. Os resultados dos revisores foram comparados e não apresentaram diferenças significativas, evidenciando que a estratégia de extração de dados foi apropriada.

2.4 Avaliação da Qualidade

Cada publicação foi avaliada pela sua qualidade, já que a qualidade dos resultados desta RSL depende da qualidade dos artigos encontrados. Os critérios de Avaliação de Qualidade (AQ) utilizados (Tabela 2) são baseados no trabalho de Dyba e Dingsøyr (2008).

Tabela 2: Critérios de Avaliação da Qualidade.

Critérios	Descrição
QA1	O estudo está baseado em pesquisas empíricas ou em relatos de experiência com base em relatórios ou na opinião de especialistas?
QA2	Existe uma definição clara dos objetivos da pesquisa?
QA3	Existe uma descrição adequada do contexto em que a pesquisa foi realizada?
QA4	O planejamento da pesquisa foi adequado para abordar os objetivos da pesquisa?
QA5	A estratégia de extração de dados foi adequada aos objetivos da pesquisa?
QA6	Os dados foram coletados de forma que abordasse as questões de pesquisa?
QA7	A análise dos dados foi suficientemente rigorosa?
QA8	Existe uma indicação clara dos resultados?

Os dois primeiros critérios foram utilizados para excluir estudos que não indicassem os objetivos da pesquisa. Isto representa o limiar mínimo de qualidade que foi observada durante esta RSL. Os critérios restantes visam determinar o rigor, análise e credibilidade dos métodos de pesquisa empregados, bem como a relevância e a qualidade de cada estudo para esta revisão. Para cada questão é atribuído o valor 1 (“Sim”), 0,5 (“Parcialmente”) ou 0 (“Não”), de acordo com o seu grau de atendimento.

3. Resultados

A execução foi realizada conforme o planejamento descrito nas seções anteriores. A Figura 1 apresenta um resumo dos resultados da aplicação dos critérios de inclusão e exclusão em cada etapa. A busca automática, realizada em quatro etapas, saiu de 463 artigos para apenas 12 artigos de acordo com os critérios de busca utilizados. Já a busca manual tem na Figura 1 apenas seu resultado final apresentado: somente três artigos foram selecionados, sendo um destes eliminado por se configurar uma síntese de outro artigo já incluído na busca automática.

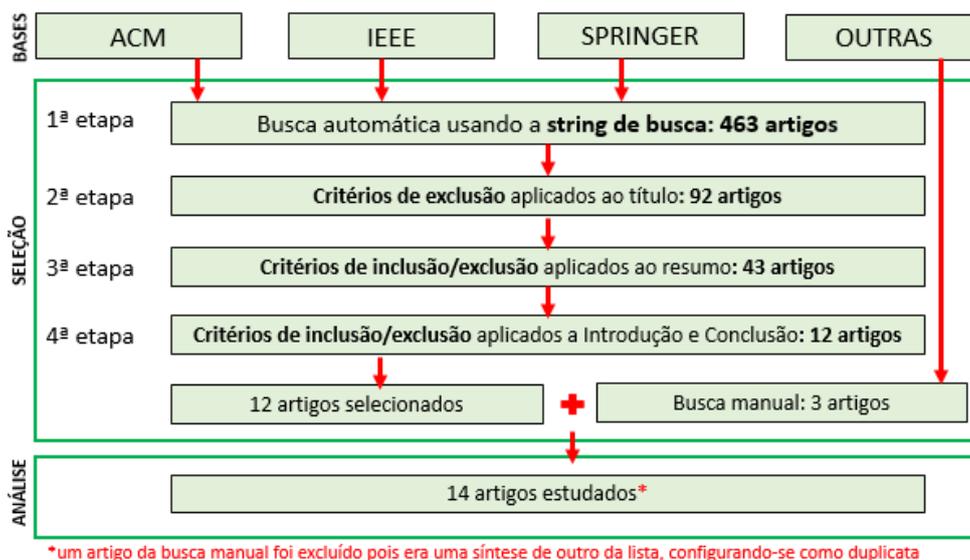


Figura 1: Estratégia de pesquisa e seleção.

As referências bibliográficas dos 14 artigos incluídos na RSL podem ser visualizadas em: <<https://goo.gl/VH7Yw9>>. Nas seções seguintes, serão apresentadas a avaliação da qualidade dos estudos primários, as respostas às questões de pesquisa e, finalmente, as limitações e ameaças do estudo.

3.1 Avaliação da Qualidade dos Estudos Reportados

Os resultados da avaliação da qualidade dos 14 artigos selecionados são apresentados na Tabela 3. Todos os artigos incluídos na revisão foram baseados em pesquisas empíricas ou apresentam relatos de experiências ou aplicações práticas (QA1), com objetivos claramente definidos (QA2). Dos 14 estudos, sete ofereceram alguma descrição do contexto na qual a pesquisa foi realizada (QA3), enquanto oito foram considerados como tendo um projeto de pesquisa apropriado (QA4).

A análise da Tabela 3 também mostra que os estudos incluídos tinham uma estratégia de extração adequada ao contexto da pesquisa (QA5). Em apenas um estudo os dados foram coletados de forma que abordaram as questões de pesquisa (QA6). Entretanto, a análise dos dados foi considerada suficientemente rigorosa em nove estudos (QA7). Em contrapartida, cinco estudos apresentaram uma indicação clara dos resultados (QA8). A pontuação média da qualidade dos trabalhos incluídos na revisão foi de 6 sendo de 4,0 o trabalho com menor nota.

Tabela 3: Resultados da Avaliação da Qualidade.

Estudos	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	Total
E1	1	1	0.5	0.5	1	0.5	0.5	0.5	5.5
E2	1	0.5	1	1	1	1	1	0.5	7
E3	0.5	1	1	1	0.5	0.5	0.5	1	6.5
E4	1	1	1	0.5	0.5	0.5	1	1	6.5
E5	0.5	0.5	1	1	1	0.5	0.5	0.5	5.5
E6	1	0.5	0.5	0.5	0.5	0.5	1	0.5	5
E7	0.5	1	0	1	1	0	1	1	5.5
E8	1	1	1	1	1	0	1	0.5	6.5
E9	0.5	1	1	1	1	0	1	0.5	6
E10	1	0.5	0.5	0.5	0.5	0	1	0.5	5
E11	0.5	1	0.5	1	1	0	1	0.5	5.5
E12	1	1	1	1	0	0.5	1	1	6.5
E13	0.5	1	0.5	0.5	1	0	0.5	1	5
E14	0.5	1	0.5	0.5	0.5	0	0.5	0.5	4
Total	10.5	12	10	11	10.5	4	11.5	9.5	/

3.2 Respostas às Questões de Pesquisa

As respostas às questões pesquisa descritas na Subseção 2.1 são discutidas a seguir.

QP1: Quais são os objetivos da utilização da Mineração de Processo Educacional?

Esta questão de pesquisa buscou esclarecer quais os objetivos educacionais que orientaram os estudos selecionados e que motivaram a utilização de Mineração de Processos. Devido à variedade de respostas encontradas nos estudos, esta questão de pesquisa exigiu uma avaliação e interpretação mais minuciosa dos dados. A análise indicou, por exemplo, que não há predominância na resolução de problemas de uma área específica no campo da educação. Ou seja, não há evidências que apontem que determinados objetivos são mais comuns. Há, na verdade, diversas questões no campo educacional que são tratadas através da Mineração de Processos. Ainda assim, é importante destacar a temática voltada para a aprendizagem autorregulada, presente em três [E1, E3 e E12] dentre os catorze artigos avaliados, o que corresponde a 21%.

De forma resumida, os objetivos de cada estudo reportado na RSL são: [E1]: examinar a relação entre atitudes de estudantes e aprendizagem autorregulada; [E2]: entender o processo de desenvolvimento de software e direcionar orientações e *feedback* de professores; [E3]: analisar a ordem temporal de atividades de aprendizagem autorregulada; [E4]: identificar fatores que influenciam a performance e diferenciar grupos de estudantes acerca dos modelos de processos; [E5]: analisar processos de formação e conformidade com restrições curriculares; [E6]: demonstrar técnicas de análise de redes sociais e avaliar interações entre unidades de cursos e provedores de cursos e checagem de caminhos de formação; [E7]: descobrir padrões de interação de usuários; [E8]: detectar discrepâncias e violações entre os fluxos previstos no modelo de registro de estudantes (currículo) e analisar diferenças entre os modelos de processos (propostos e diferenciados); [E9]: análise de desempenho de alunos para melhorias no processo de avaliação do TOEFL; [E10]: não especificados ou não identificados; [E11]: permitir através da criação de uma plataforma de computação distribuída que centros de ensino possam analisar e reestruturar seus processos; [E12]: avaliar atividades de aprendizagem autorregulada; [E13]: melhorar o processo de registro de estudantes através da recomendação de um modelo de processos; [E14]: facilitar a geração de um

mapa de navegação para unidades de aprendizagem e reorganizar o processo de aprendizagem adaptado para a situação de cada usuário.

QP2: Em quais das subáreas (principais) as pesquisas em Mineração de Processos para Educação se encaixam?

Mineração de Processos pode ser subdividida em três grandes subáreas ou conjuntos de técnicas [Trcka et al., 2010]:

- *Process Model Discovery*: criação de um modelo de processos a partir da realidade observada e armazenada, usualmente, em *logs*;
- *Conformance Checking*: checagem de um modelo esperado ou especificado e verificação de sua equivalência à realidade observada;
- *Process Model Extension*: projeção das informações extraídas dos *logs* ou adição de uma nova perspectiva e características adicionais a um modelo de processos existente.

Esta questão buscou avaliar a adequação dos estudos à essas subáreas. A partir da análise efetuada, foi possível constatar que prevalecem os estudos que se enquadram na área de *Process Model Discovery* [E1, E2, E3, E4, E6, E7, E11, E12 e E13] e *Conformance Checking* [E2, E3, E5, E7, E8, E9, E10, E11]. É importante também destacar que 36% dos estudos ingressam concomitantemente em mais de uma subárea, o que demonstra que, embora as técnicas empregadas em cada subárea sejam distintas, elas podem ser combinadas para se alcançar um objetivo de pesquisa maior. Os estudos [E2, E3, E7, E11] adotaram conjuntamente as técnicas *Process Model Discovery* e *Conformance Checking*, enquanto o estudo [E5] utilizou as técnicas *Conformance Checking* e *Process Model Extension*. Conforme pode se observar, as técnicas das áreas *Process Model Discovery* e *Conformance Checking* são utilizadas em praticamente todos os trabalhos, seja de forma isolada ou em combinação e há apenas um trabalho que se enquadra exclusivamente na subárea de *Process Model Extension* [E14]. O gráfico da Figura 2 ilustra a (a) porcentagem de trabalhos em cada subárea e a (b) ocorrência dos tipos de MPE.

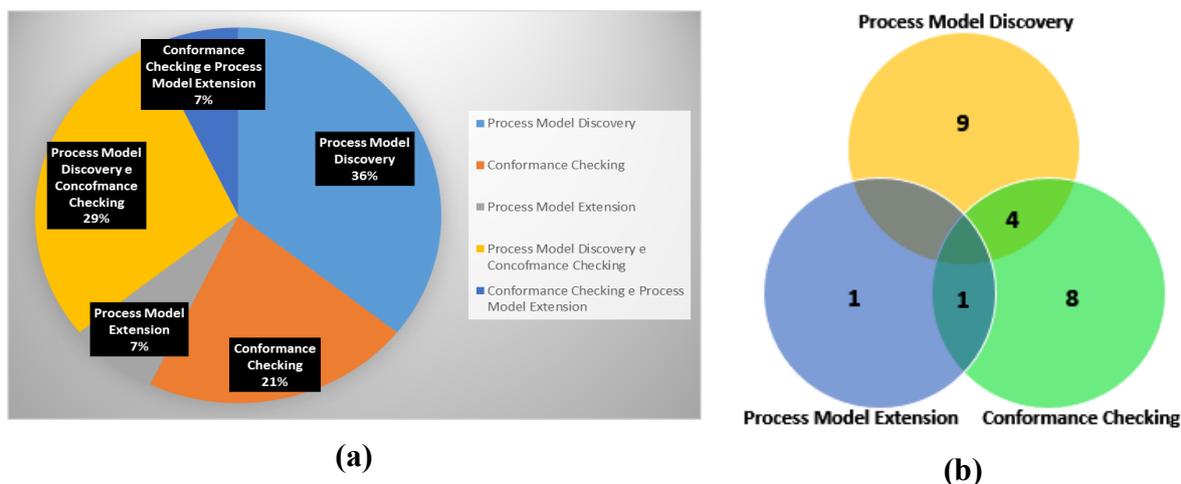


Figura 2: (a) Técnicas de MPE utilizadas nos estudos e (b) Diagrama ilustrando a ocorrência dos tipos de MPE.

QP3: Quais as principais técnicas (plugins) e algoritmos estão sendo utilizados pelos pesquisadores para extrair informações dos processos?

A análise dos estudos indicou uma alta variedade de algoritmos que são utilizados em MPE, sendo possível destacar o *Fuzzy Miner*, presente em seis estudos [E1, E2, E3, E4, E7 e E13], equivalente a 43% do total. Além do *Fuzzy Miner*, não é possível evidenciar a predominância de nenhum outro algoritmo, havendo assim uma distribuição de diversos algoritmos pelos estudos analisados. Ainda assim, alguns algoritmos aparecem em mais de um estudo, são eles o *Alpha Algorithm* [E7, E10 e E13], o *Heuristic Miner* [E6 e E9] e o *Dotted Chart* [E5 e E6]. Essa diversidade de algoritmos se deve, muito provavelmente, à variedade de objetivos definidos nos estudos, conforme destacado nas respostas a QP1, e também ao número elevado de soluções para um mesmo problema, ficando a critério do pesquisador a escolha do algoritmo a ser utilizado.

Além do emprego dos algoritmos, outra questão importante a ser esclarecida diz respeito às ferramentas que dão suporte as atividades de MPE. Neste sentido, a análise evidenciou o uso intenso do *framework* ProM (abreviação de *Process Mining*), presente em dez estudos [E3, E4, E5, E6, E8, E9, E10, E11 e E13]. O ProM é um *framework open source* extensível que dá suporte a uma ampla variedade de técnicas e algoritmos de Mineração de Processos sob a forma de *plug-ins* [Romero et al, 2010]. Além desse *framework*, outra ferramenta utilizada foi a Disco, mencionada nos estudos [E2, E4 e E7]. Isso evidencia o domínio dessas duas ferramentas neste campo de pesquisa. Há ainda a utilização de outras ferramentas com propósitos distintos da Mineração de Processos, estas dão apoio às atividades de coleta de dados e métricas (Bugzila [E2], Mercurial, Git [E1] e Bitbucket [E12]) e nas atividades voltadas ao ensino e aprendizagem (nStudy e SoftLearn).

QP4: Quais os benefícios e limitações/dificuldades estão sendo relatados nos estudos?

Ao analisar os estudos, foi possível verificar que 86% deles apontavam claramente uma lista de benefícios alcançados ao se utilizar MP em contextos educacionais. A maioria dos estudos aponta como benefícios diretos a avaliação ou criação de modelos de processos de comportamento, interação social, dentre outros. Porém, em termos de contribuições no campo da educação é possível destacar os estudos [E1, E2, E5, E7 e E12], cujos resultados referenciam e favorecem diretamente o processo de aprendizagem de estudantes por meio de análise comportamental ou das estratégias de aprendizagem adotadas. Há ainda benefícios direcionados para análise de currículos e criação de itinerários de formação [E5, E6 e E11] e para o desenvolvimento de *frameworks* e plataformas de serviços educacionais [E2 e E14]. Dentre as dificuldades relatadas nos estudos pode se destacar problemas de ordem estatística e/ou metodológica, relacionados principalmente ao tamanho das amostras utilizadas [E1, E3 e E4]. Além destas, uma limitação pertinente identificada nos estudos foi falta de adequação dos algoritmos utilizados aos dados disponíveis [E4, E5 e E10].

3.3. Ameaças a Validade da RSL

No geral, as principais ameaças à validade da RSL são em relação ao viés de seleção de publicações e aos dados imprecisos extraídos. As pesquisas foram concebidas empregando principalmente recursos eletrônicos. Estes foram pesquisados após a

implementação de pesquisas experimentais, *strings* de busca e utilização de dicionários de sinônimos. Apesar disso, não é possível garantir que todos os estudos relevantes foram coletados e há um pequeno risco de que alguns estudos podem ter sido omitidos, devido aos termos de busca utilizados. O processo de extração de dados também pode resultar em dados irrelevantes. Isto pode ocorrer, porque o processo de extração de dados foi realizado somente por dois pesquisadores. O desenvolvimento de um protocolo da RSL e o uso de uma estratégia de verificação da qualidade da revisão (por um terceiro revisor) ajuda a garantir que a ameaça foi devidamente controlada. Finalmente, é possível que os critérios de inclusão possam ter ocasionalmente excluídos alguns estudos relevantes, caso os estudos não contivessem palavras-chaves apresentadas nos títulos ou nos resumos. Não houve, porém, nenhum indicio de que essa ameaça possa ter se concretizada.

4. Conclusões

Esta RSL buscou traçar um panorama das pesquisas em MPE, avaliando e evidenciando os trabalhos mais relevantes na área. A busca inicial retornou a soma de 463 artigos que foram avaliados segundo critérios de inclusão e exclusão estabelecidos no protocolo de seleção, restando 14 trabalhos que foram incluídos na extração dos dados e estruturaram a análise final.

O objetivo da Mineração de Processos é auxiliar no entendimento e aperfeiçoamento do modelo de negócios como um todo [Trcka *et. al.*, 2010]. Desta forma, a mineração trazida para o âmbito educacional ajuda a formar uma representação mais clara do processo educacional, tratando-o como uma unidade completa. Nesse sentido, uma das principais motivações para a revisão e que ajudou a conduzir as conclusões do artigo foi buscar evidências dos benefícios da utilização de Mineração de Processos sobre dados educacionais. Questões de pesquisa foram criadas e respondidas com esse propósito e, para além, buscou-se traçar um perfil metodológico e tecnológico acerca dos trabalhos em MPE. A revisão apresentou resultados que dão uma visão mais precisa a respeito dos estudos e aponta ainda tendências para pesquisas em MPE, ajudando no direcionamento de trabalhos futuros.

Com relação a esses objetivos e com base na análise efetuada, foi possível chegar a algumas observações pertinentes: (i) a utilização de Mineração de Processo efetivamente traz benefícios no campo educacional, comprovado pelo número de 86% de trabalhos que mencionaram contribuições para área; (ii) a Mineração de Processos não se detém em resolver problemas educacionais pontuais, havendo uma gama de questões em que a MPE está inserida; (iii) ainda não há um padrão metodológico na escolha das técnicas e algoritmos empregados nos problemas investigados; e (iv) é notório o domínio do framework ProM como principal ferramenta utilizada na MPE.

Além destas observações é fundamental destacar ainda dois aspectos importantes sobre a pesquisa em MPE. Primeiro: este campo de investigação ainda é recente e se encontra em fase de consolidação. Esta afirmação é corroborada pela diversidade de áreas dos eventos e periódicos onde os estudos foram publicados. Este fato sugere que a MPE ainda está em busca de uma maturidade científica, dificultada por sua origem histórica interdisciplinar. Segundo: embora ainda esteja em busca de espaço, houve um aumento significativo no número de publicações relevantes nos últimos anos. O ano de

2013 contou com uma única publicação, enquanto que em 2015 este número chegou a sete publicações, o que representa um interesse crescente por pesquisas relacionadas a MPE.

Estas observações só foram possíveis graças a análise cautelosa dos estudos reportados, o que nos permite também fazer prospecções quanto as pesquisas futuras em MPE. Dentre as oportunidades de pesquisa identificadas destacamos a utilização de dados multimídia (voz e vídeo) como base para MPE, o estudo da efetividade de algoritmos de Mineração de Processo sobre dados educacionais e a investigação de modelos de processos voltados para ações e comportamentos de professores.

Agradecimentos

Os autores agradecem a CAPES pelo suporte parcial a esta pesquisa, ao Instituto Nacional de Engenharia de Software (INES), ao CNPq, aos integrantes do Laboratório Gamedu.net e, em especial, ao Programa de Pós-Graduação em Sistemas Computacionais – PPgSC/UFRN, por toda infraestrutura oferecida.

Referências

- Aalst, W. V. D. (2012) “Process Mining: Overview and Opportunities”. ACM Transactions on Management Information Systems – TMIS, v.3, n.2.
- Kitchenham, B.; Brereton, P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. (2009) “Systematic literature reviews in software engineering – A systematic literature review”. In: Information and Software Technology, v.51, p.7-15.
- Manhães, L. M. B.; Cruz, S. M. S.; Costa, R. J. M.; Zavaleta, J.; Zimbrão, G. (2011) “Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados”. In: XXII Simpósio Brasileiro de Informática na Educação – SBIE, p.150-159.
- Romero, C.; Ventura S.; Pechenizkiy, M.; Baker, R. (2010) “Handbook of Educational Data Mining”. Boca Raton, FL: CRC Press, Taylor and Francis.
- Romero, C.; Ventura, S. (2007) “Educational Data Mining: A survey from 1995 to 2005”. Expert Systems with Applications, 33(1):135–146.
- Silva, T. R.; Medeiros, T. J.; Aranha, E. H. S. (2014) “Jogos Digitais para Ensino e Aprendizagem de Programação: uma Revisão Sistemática da Literatura”. In: XXV Simpósio Brasileiro de Informática na Educação - SBIE, p.692-701.
- Trčka, N.; Pechenizkiy, M. (2009) “From Local Patterns to Global Models: Towards Domain Driven Educational Process Mining”. In: Ninth International Conference on Intelligent Systems Design and Applications – ISDA, Pisa, pp.1114-1119.