

15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

LEARNING ANALYTICS: UMA REVISÃO INTEGRATIVA DA LITERATURA

OTÁVIO H. PEREIRA¹, CAROLINA A. DE FREITAS-DIAS²

¹ Graduando em Administração, Bolsista PIBIT, IFSP, Campus São Roque, otavio.pereira@aluno.ifsp.edu.br.

² Professora Doutora em Administração, Orientadora do projeto, IFSP, Campus São Roque, carolina.dias@ifsp.edu.br.
Área de conhecimento (Tabela CNPq): 6.02.00.00-6 Administração

RESUMO: O *Learning Analytics* (LA), compreende a mensuração, coleta, análise e o reporte de dados sobre os alunos e seus contextos com o propósito de entender e otimizar a aprendizagem e o ensino. É um dos subcampos da ciência de dados e surgiu em função da massiva quantidade de dados complexos disponíveis, com a finalidade de informar e apoiar alunos, professores e suas instituições em uma melhor compreensão e previsão das necessidades pessoais de aprendizagem e desempenho. No Brasil, o tema ainda é praticamente inexplorado, visto que em uma busca realizada na base Scielo Brasil, por meio do termo “*learning analytics*” aplicado ao campo título, apenas um trabalho nacional foi localizado. Nesse contexto, o presente estudo teve o objetivo de criar uma ampla base de dados dos estudos anteriores sobre LA que sirva como ponto de partida para análise em um projeto futuro. Para alcançar o objetivo proposto, foi utilizado o método da Revisão Integrativa, que orientou o processo de busca, coleta e seleção dos artigos sobre o tema. Por fim, este trabalho resultou na construção de uma base de dados ampla e representativa, contendo 622 artigos para pesquisas futuras sobre *Learning Analytics*.

PALAVRAS-CHAVE: *learning analytics*; revisão da literatura; revisão integrativa.

LEARNING ANALYTICS: AN INTEGRATIVE LITERATURE REVIEW

ABSTRACT: Learning Analytics (LA) encompasses the measurement, collection, analysis, and reporting of data about learners and their contexts with the purpose of understanding and optimizing learning and teaching. It is one of the subfields of data science and emerged due to the massive amount of complex data available, aimed at informing and supporting students, teachers, and their institutions in better understanding and predicting personal learning and performance needs. In Brazil, the topic is still virtually unexplored, as a search conducted in the Scielo Brazil database using the term “learning analytics” in the title field found only one national study. In this context, the present study aimed to create a comprehensive database of previous studies on LA that would serve as a starting point for analysis in a future project. To achieve the proposed objective, the Integrative Review method was used, guiding the process of searching, collecting, and selecting studies on the topic. Finally, this work resulted in the construction of a comprehensive and representative database, containing 622 articles for future research on Learning Analytics.

KEYWORDS: learning analytics; literature review; integrative review.

INTRODUÇÃO

A ciência de dados é algo recente, surgiu formalmente em 2001 em função da massiva quantidade de dados complexos disponíveis e é composta por diversas perspectivas. Na educação, os subcampos *Educational Data Mining*, ou Mineração de Dados Educacionais; *Educational Analytics*, ou Analítica da Educação; e Inteligência Artificial na Educação compõe a chamada *Educational Data Science* ou Ciência de Dados Educacionais. É na área de *Educational Analytics* que se insere o *Learning Analytics* (LA) ou Analítica de Aprendizagem, foco do presente estudo, juntamente com o *Academic Analytics* e o *Learner Analytics* (Filatro, 2021). Na figura 1 são expostos os diferentes conceitos e as suas relações.

Learning Analytics compreende a mensuração, coleta, análise e o reporte de dados sobre os alunos e seus contextos com o propósito de entender e otimizar a aprendizagem e o ensino (Siemens, 2013). LA surgiu como campo em 2011, teve suas raízes na inteligência de negócios e na mineração de dados, os fatores que impulsionaram a sua criação e desenvolvimento foram a vasta disponibilidade de dados sobre ensino e aprendizagem, o crescimento da aprendizagem online e a preocupação nacional com o desempenho educacional (Ferguson *et al.*, 2019). Conforme Filatro (2021) o foco de LA é a análise, para informar e capacitar aprendizes, professores e gestores na tomada de decisão baseada em dados.

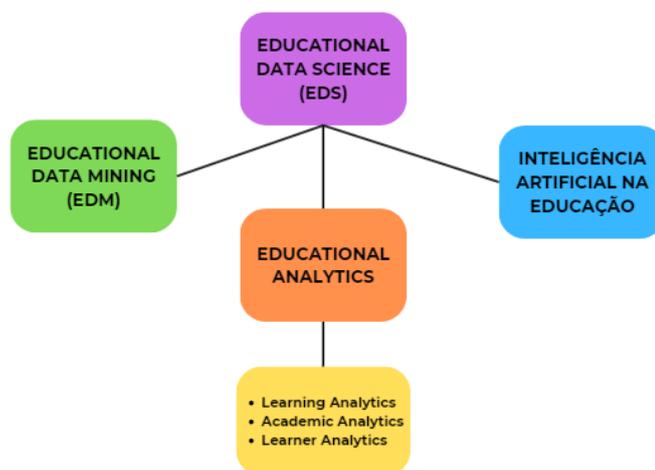


FIGURA 1. Campo da ciência de dados educacionais e seus subcampos.
Fonte: criado a partir de Filatro (2021).

Nesse cenário, fica evidente que se trata de um tema amplo e recente, com uma produção internacional numerosa, pois, em 2022, na base de dados Web of Science, a busca do termo “*learning analytics*” resultou em 2.236 estudos enquanto na base Scielo Brasil, apenas um estudo nacional foi localizado. Assim, estudos nacionais sobre LA são fundamentais para apontar o estado da arte, organizar a literatura e orientar estudos futuros. Por fim, o presente trabalho teve o objetivo de criar uma base de dados robusta e representativa sobre LA que sirva como ponto de partida para pesquisas futuras.

MATERIAL E MÉTODOS

Para alcançar o objetivo principal deste trabalho, criar uma base de dados ampla e representativa sobre o tema LA, foram estabelecidos os seguintes objetivos específicos: (1) compreender o método da revisão integrativa; (2) estabelecer o protocolo de pesquisa para busca, coleta e seleção dos estudos sobre o tema; e (3) guiado pelo protocolo estabelecido, realizar as operações para construção da base de dados.

Revisão integrativa

O primeiro objetivo específico do presente estudo foi compreender o método da revisão integrativa (RI) buscando artigos teóricos e empíricos sobre o tema. Assim, foram revisados os artigos teóricos de Ganong (1987), Broome (2000), Whittemore e Knafl (2005), Crossan e Apaydin (2010) e Botelho, Cunha e Macedo (2011).

O método da revisão sistemática (RS) é composto por estratégias que permitem ao pesquisador mitigar os vieses da seleção de artigos, avaliá-los de maneira crítica e sintetizar as pesquisas relevantes sobre um tópico específico (Botelho; Cunha; Macedo, 2011). Em seu estudo, Whittemore e Knafl, (2005) apresentaram a RI como um dos métodos que compõe a RS e evidenciaram o seu potencial de fornecer uma visão holística sobre o tópico de interesse por incluir diversas fontes de dados que aproximam o pesquisador da compreensão do tema. Botelho, Cunha e Macedo (2011) preconizam que através da RI o pesquisador pode conhecer o desenvolvimento do tema, identificar lacunas e oportunidades para pesquisas futuras. Desse modo, os autores apresentaram uma estrutura composta por seis etapas predefinidas que devem ser seguidas para a realização da RI, as quais foram listadas e brevemente descritas a seguir:

1. Identificação do tema e seleção da questão de pesquisa: processo composto pela identificação do problema e elaboração da pergunta de pesquisa. Em seguida, define-se as palavras-chave e estratégia de busca para obtenção de materiais na base de dados.
2. Estabelecimento dos critérios de inclusão e exclusão: utilizam-se as bases de dados para a realização das buscas por materiais, nessa etapa o pesquisador poderá se deparar com um retorno volumoso e amplo, o que exigirá a análise dos materiais de maneira crítica. As definições da etapa anterior são essenciais para o sucesso dessa. Pode-se concluir que o pesquisador deve definir critérios objetivos de inclusão e exclusão.
3. Identificação dos estudos pré-selecionados e selecionados: análise preliminar dos artigos, composta pela leitura do título, resumo e palavras-chave. As publicações selecionadas e pré-selecionadas devem ser organizadas através de uma tabela ou matriz. Se não for possível decidir após a análise preliminar, deve-se ler o documento na íntegra para verificar a sua adequação à pesquisa e decidir por sua inclusão ou exclusão.
4. Categorização dos estudos selecionados: elabora-se então a matriz de síntese para categorização e análise crítica das informações obtidas nos estudos selecionados.
5. Análise e interpretação dos resultados: discussão dos resultados obtidos após a análise dos artigos selecionados.
6. Apresentação da revisão/síntese do conhecimento: nesta etapa o pesquisador deve criar o documento em que sintetiza os principais resultados encontrados, assim como descrever como realizou a pesquisa, de maneira criteriosa, apresentando todas as fases percorridas uma vez que a revisão deve disponibilizar as informações necessárias que possibilitem a reprodução da pesquisa.

Protocolo de pesquisa

O segundo objetivo específico deste trabalho, foi estabelecer um protocolo de pesquisa para guiar a busca, coleta e seleção de estudos sobre o tema *Learning Analytics*, desse modo, neste estudo foram conduzidas as etapas 1, 2 e 3 da estrutura recomendada por Botelho, Cunha e Macedo (2011) para a RI e as especificações do protocolo foram descritas a seguir:

1. Identificação do tema e seleção da questão de pesquisa

A questão de pesquisa que orientou o estudo foi “Qual é o estado da arte sobre o tema *Learning Analytics*?”. Essa questão direcionou a criação dos objetivos do estudo, os quais procuraram guiar a criação de uma base de dados para projetos futuros sobre o tema.

2. Estabelecimento dos critérios de inclusão e exclusão

Neste estudo, foram utilizados artigos revisados por pares já que esses têm o seu conhecimento considerado validado e é provável que tenham maior impacto (Podsakoff *et al.*, 2005 *apud* Crossan; Apaydın, 2010) e nesse contexto, a base de dados Clarivate Web of Science (WoS) foi escolhida por ser uma das bases mais abrangentes no quesito artigos revisados por pares (Crossan; Apaydın, 2010). O acesso foi realizado através do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), com acesso pelo Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), por meio da Comunidade Acadêmica Federada (CAFe).

Os artigos extraídos da base de dados foram organizados em três grupos, o primeiro grupo foi composto por revisões e meta-análises, o segundo grupo por artigos altamente citados e o terceiro grupo por publicações recentes (2018 a 2023), com o objetivo de mitigar defasagens de citação e outros vieses. É importante destacar que antes de iniciar a análise preliminar, as sobreposições foram removidas, de

modo que os artigos fossem retidos no grupo em que apareceram pela primeira vez. Os critérios utilizados para a criação dos três grupos foram descritos a seguir.

Grupo 1 – revisões e meta-análises: apresentar o estado da arte de um tema é uma das características de uma boa revisão (Botelho; Cunha; Macedo, 2011) e para compreender seus fundamentos teóricos, artigos de revisão e meta-análises (Crossan; Apaydin, 2010) foram utilizados para composição deste grupo.

Grupo 2 – artigos altamente citados: a análise baseada em citações foi o critério utilizado para a formação desse grupo por ser amplamente aplicada para avaliar a qualidade dos artigos. O alto número de citações pode ser considerado como um voto da sua contribuição para o desenvolvimento e acúmulo do conhecimento (Saha *et al.*, 2003 *apud* Crossan; Apaydin, 2010). Neste estudo, artigos com cinco citações ou mais foram considerados elegíveis para a etapa de identificação de estudos selecionados e pré-selecionados.

Grupo 3 – artigos recentes: o *Journal Impact Factor* (JIF) ou Fator de Impacto do Periódico, foi utilizado como critério de seleção de estudos nesse grupo com o objetivo de reduzir as defasagens e vieses de citação das publicações mais recentes (2018-2023) ao reconhecer que o método baseado na quantidade de citações pode prejudicar a pesquisa através da exclusão de artigos recentes que ainda não tiveram tempo de acumular citações (Crossan; Apaydin, 2010). Para Strehl (2005), periódicos com JIF superior 0,5 podem ser considerados de alto impacto na área das Ciências Sociais e, sendo assim, esse valor foi utilizado como base para a pré-seleção de artigos desse grupo, para que trabalhos da área organizacional fossem considerados.

3. Identificação dos estudos pré-selecionados e selecionados

A análise para seleção e pré-seleção de artigos foi realizada a partir da leitura do título, resumo e palavras-chave dos estudos (Botelho; Cunha; Macedo, 2011), durante a leitura buscava-se por artigos que abordaram o tema *Learning Analytics* em qualquer nível (individual, grupal e organizacional), foco (teórico e prático), tipo de aplicação (desenvolvimento de aplicativos, *dashboards*, modelos etc.), avaliação (críticas, questões éticas ou de privacidade), bem como aqueles que indicaram o uso de LA em organizações.

Conforme Ganong (1987), os critérios estabelecidos devem ser localizados nos trabalhos analisados e sendo assim, os artigos que atenderam aos critérios citados na seção anterior foram selecionados, extraídos da base de dados, divididos e organizados nos três grupos descritos anteriormente. Ainda com relação aos artigos selecionados, foram considerados apenas aqueles em que LA foi o tema central da pesquisa, ou seja, quando o tema foi tangencial o estudo não foi selecionado para a base de dados, objeto do presente estudo. Assim, pesquisas que abordavam temas como *Machine Learning*, *Deep Learning*, *Big Data Analytics*, Inteligência Artificial ou aqueles em que o tema *Analytics* teve o seu foco direcionado a outras disciplinas, por exemplo análise de multimídia, de tráfego, saúde entre outros, também não foram selecionados.

RESULTADOS E DISCUSSÃO

O terceiro objetivo específico deste estudo foi realizar as operações do protocolo estabelecido e sendo assim, em 24/05/2023 a base de dados Clarivate Web of Science foi acessada via Portal de Periódicos da CAPES, procedeu-se uma busca que iniciou com o termo “*Learning Analytics*” aplicado ao campo de título das publicações que resultou em 2.717 artigos. Em seguida, os resultados foram refinados através da aplicação dos filtros a seguir:

TABELA 1. Filtros utilizados para refinar os resultados da busca inicial.

Seção	Filtro
Tipo de documento	Artigo; Artigo de revisão
Idioma	Inglês; Português

Fonte: elaborado pelos autores.

Não foram condicionadas restrições cronológicas, de categoria ou qualquer outra resultando, após a aplicação dos filtros, em uma amostra inicial de 1.334 artigos, os quais foram utilizados como base para a identificação dos estudos selecionados e pré-selecionados, conforme a seguir.

Grupo 1 – revisões e meta-análises: após a busca inicial, foram realizados dois refinamentos através de um campo dedicado a busca nos resultados obtidos inicialmente. O objetivo foi localizar os artigos desejados para este grupo, assim, o primeiro refinamento ocorreu com o acréscimo do termo “*meta*” e separadamente, um segundo refinamento foi realizado através da aplicação do termo “*review*” ao mesmo campo, ambas operações sobre a amostra inicial de 1.334 artigos. A busca com a aplicação do termo *meta* resultou em um conjunto de 15 estudos e para o termo *review* em 231, totalizando 246 itens para o grupo 1 que, após a remoção das duplicidades resultou em 241 artigos. Por fim, após a análise preliminar dos artigos, composta pela leitura do título, resumo e palavras-chave dos trabalhos, este grupo foi composto por 160 estudos sendo 12 resultantes da busca com o termo *meta* e 148 da busca com o termo *review*.

Grupo 2 – artigos altamente citados: foram analisados os 1.334 artigos e aqueles com 5 citações ou mais foram mantidos, resultando em um conjunto de 739 artigos de alto impacto. Após a análise de duplicidades comparativamente aos estudos elegíveis do grupo 1, o grupo 2 foi formado por 549 estudos, os quais foram submetidos à análise preliminar que resultou em um conjunto total de 398 artigos de alto impacto.

Grupo 3 – artigos recentes: do conjunto inicial de 1.334 artigos, um total de 1.096 foram publicados nos últimos 5 anos (2018-2022) e em parte de 2023 (até 24/05), já que a busca foi realizada na base de dados em maio de 2023. A aplicação do critério alternativo, considerando o JIF superior a 0,5, formou um conjunto composto por 775 estudos que foram submetidos à análise de sobreposição com os grupos 1 e 2, resultando em 258 artigos para análise preliminar. Finalmente, após a análise, um conjunto de 93 artigos foi selecionado para a base de dados.

Em resumo, após a formação dos grupos, foi realizada a análise preliminar dos artigos em que foram selecionados para a base de dados todos os estudos que atenderam aos critérios de seleção descritos anteriormente e essa operação resultou em 651 artigos, amostra robusta, composta por diversos estudos em que *Learning Analytics* foi o foco das pesquisas, oferecendo assim maior relevância e representatividade a amostra do presente estudo. A tabela 2 apresenta a distribuição de artigos selecionados para a base de dados por grupo.

TABELA 2. Artigos selecionados para a base de dados por grupo.

Grupo	Quantidade de artigos	Observação
1	160	12 artigos para o filtro <i>meta</i> 148 artigos para o filtro <i>review</i>
2	398	
3	93	
Total	651	

Fonte: elaborado pelos autores.

Em atendimento ao terceiro objetivo específico desse estudo, após a seleção de estudos, procedeu-se a criação da base de dados a partir do *download* dos artigos selecionados e sua respectiva importação para o software de gestão de referências Mendeley, onde os trabalhos foram classificados nos três grupos inicialmente definidos nessa seção. Devido a impossibilidade de *download* de 29 estudos que não são de livre acesso, a base final foi composta por 622 artigos sendo, 153 no grupo 1 (12 para o filtro *meta* e 141 para o filtro *review*), 378 no grupo 2 e 91 no grupo 3. A figura 2 apresenta o fluxo do processo de operacionalização realizado para obtenção da amostra final.

CONCLUSÕES

Este estudo teve o objetivo principal de criar uma base de dados abrangente sobre *Learning Analytics* que sirva como ponto de partida para projetos futuros e foi guiado pelo método da revisão integrativa que se mostrou adequado para a tarefa. Dentre as limitações do método da RI, destaca-se a dificuldade de interpretação de dados de fontes primárias (Whittemore; Knafl, 2005).

O primeiro objetivo específico do estudo foi compreender o método da RI e ficou evidente que este pode ser aplicado à pesquisas do campo organizacional, assim como sugerido no estudo de Botelho, Cunha e Macedo (2011). O segundo objetivo específico do estudo foi criar um protocolo de pesquisa para busca, coleta e seleção de artigos para a base de dados em que as etapas da estrutura da RI provaram-se pertinentes para guiar as operações. O terceiro objetivo específico foi realizar as operações do protocolo estabelecido para criação da base de dados, tarefa que foi realizada de maneira eficaz e resultou em uma amostra robusta e abrangente sobre LA, composta por 651 estudos de 2011 a 2023.

Devido a impossibilidade de *download* de 29 trabalhos, os quais não dispõem de acesso aberto ou não estavam acessíveis no Portal de Periódicos CAPES, este estudo resultou em uma base de dados final composta por 622 artigos, ainda robusta e representativa, com o potencial de contribuir com os avanços dos estudos sobre o *Learning Analytics*.

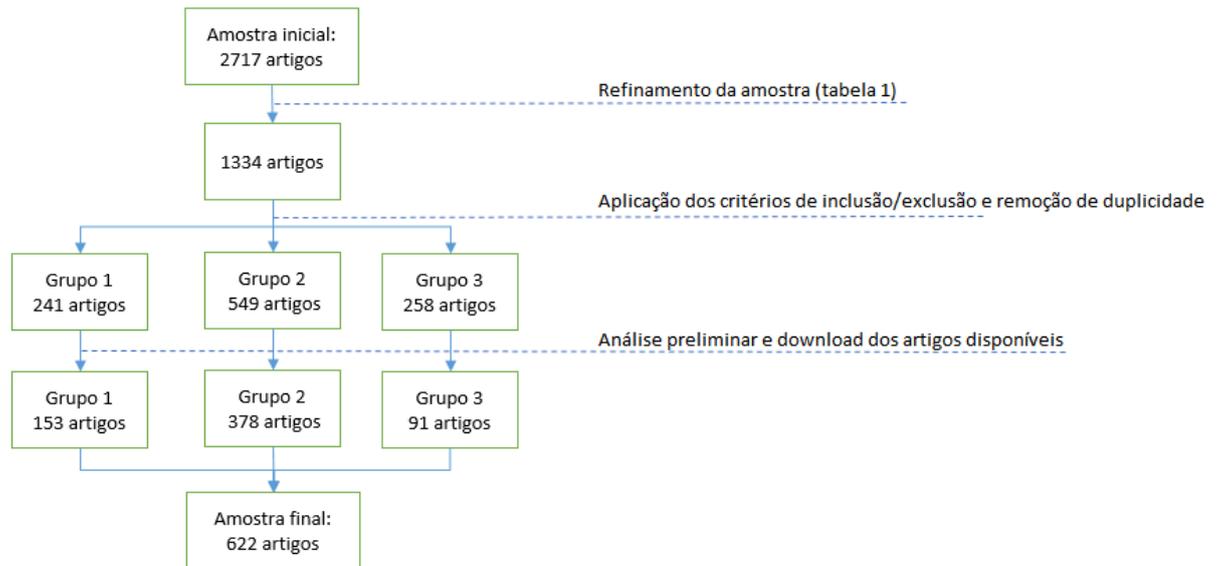


FIGURA 2. Processo de busca e obtenção de artigos para a amostra.
Fonte: criado pelos autores.

CONTRIBUIÇÕES DOS AUTORES

O.H.P desenvolveu a metodologia do estudo, realizou a coleta dos dados, organizou a base de dados e elaborou a redação do manuscrito principal.

C.A.F.D. contribuiu com a concepção, obtenção de financiamento, administração do projeto, supervisão, curadoria dos dados, redação e revisão do relatório.

AGRADECIMENTOS

Este projeto foi totalmente financiado através do Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), edital 218/2022.

REFERÊNCIAS

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. O método da revisão integrativa nos estudos organizacionais. **Revista Eletrônica Gestão e Sociedade**, v. 5, n. 11, p. 121–136, 2011.

BROOME, M. E. Integrative literature reviews for the development of concepts. In: **Concept development in nursing: foundations, techniques and applications**. 2^o ed. [s.l.: s.n.]. p. 231–250.

CROSSAN, M. M.; APAYDIN, M. A multi-dimensional framework of organizational innovation: a systematic review of the literature. **Journal of Management Studies**, v. 47, n. 6, p. 1154–1191, 30 set. 2010.

FERGUSON, R.; CLOW, D.; GRIFFITHS, D.; BRASHER, A. Moving forward with learning analytics: expert views. **Journal of Learning Analytics**, v. 6, n. 3, p. 43–59, 2019.

FILATRO, A. **Data Science na educação: presencial, a distância e corporativa**. 1ª ed. São Paulo - SP: Saraiva, 2021.

GANONG, L. H. Integrative reviews of nursing research. **Research in Nursing & Health**, v. 10, n. 1, 1987.

SIEMENS, G. Learning analytics: the emergence of a discipline. **American Behavioral Scientist**, v. 57, n. 10, p. 1380–1400, out. 2013.

STREHL, L. O fator de impacto do ISI e a avaliação da produção científica: aspectos conceituais e metodológicos. **Ciência da Informação**, v. 34, n. 1, p. 19–27, 2005.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. **Journal of Advanced Nursing**, v. 52, n. 5, p. 546–553, 2005.