Qualidade de dados no processo decisório de educação a distância

Lafaiete Henrique Rosa Leme[a], Felipe Stefani Correia dos Santos[b], Késsia Rita da Costa Marchi[c] e Leonardo Castro Botega[d]

1. Introdução

O campo educacional tem incorporado Tecnologias de Informação e Comunicação (TIC) em suas atividades referentes aos processos de ensino e aprendizagem. Por consequência disto, são produzidos cada vez mais dados que demandam de análises detalhadas e voltadas para um melhor planejamento e execução das ações educacionais.

Neste contexto, os sistemas de gestão de aprendizagem (em tradução livre de *Learning Management Systems* – LMS) geram muitos dados que podem ser analisados e utilizados para auxílio da tomada de decisão pedagógica, gerando informações úteis a partir dos dados coletados neste ambiente. Contudo, estes dados gerados nem sempre são dados de qualidade, podendo então

a Mestre em Ciência da Computação pela Universidade Estadual de Maringá – UEM. E-mail: lafaiete.leme@unesp.br.

b Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: felipe.stefani@unesp.br. ORCID: https://orcid.org/0000-0002-9359-9272.

c Doutoranda em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: kessia.marchi@unesp.br. ORCID: https://orcid.org/0000-0001-5504-5685.

d Doutor em Ciência da Computação pela Universidade Federal de São Carlos – UFSCar. Professor na Universidade Estadual Paulista – UNESP. E-mail: leonardo.botega@unesp.br. ORCID: https://orcid.org/0000-0003-1495-5935.

gerar informações as quais não tem aplicação ou não tem confiabilidade podendo culminar em uma má interpretação da realidade e gerar uma falsa impressão, o que atrapalha a tomada de decisão, inclusive no contexto pedagógico.

Diante do exposto, este trabalho apresenta uma revisão de literatura que tem como objetivo identificar dimensões de qualidade de dados abordadas, direta ou indiretamente, em artigos que tratam de tomadas de decisão pedagógica baseada em dados provenientes de LMS. Ao desenvolver este trabalho foram apresentados os conceitos de qualidade de dados, informações e as dimensões de qualidade, identificadas no contexto da educação, visando contribuir com uma análise crítica sobre estas abordagens. Ao considerar os trabalhos é possível observar uma escassez de artefatos que remetem à qualidade dos dados, o que levou a identificar lacunas que requerem atenção.

2. A qualidade de dados e informação

O conceito de qualidade de dados apresenta variações e pode ser interpretado de acordo com o contexto em que é aplicado. Floridi e Illari (2014) e Samitsch (2015) definem que uma informação é de qualidade quando atende ao seu propósito. Batini e Scannapieco (p. 5, 2016) apresentam algumas definições para qualidade que, de forma geral, compreendem a adequação do uso, a conformidade com os requisitos e a satisfação do usuário.

Desta forma, a qualidade é reconhecida como um problema multidimensional e impacta diretamente na performance de

processos operacionais, atividades de tomadas de decisão e de requisitos de cooperação inter-organizacionais. Samitsch (2015) afirma que tomadas de decisões baseadas em dados de alta qualidade proporcionam melhorias na produtividade organizacional.

Diante deste contexto, Wang e Strong (1996) e Wang (1998) abordaram que a qualidade dos dados não precisa ser absoluta, mas deve atender a critérios que possibilitam tomadas de decisões significativas e eficientes, considerando que uma vez que os stakeholders perdem confiança na qualidade dos dados, o sistema está relegado como não confiável, talvez para sempre (REDGATE, 2016).

Para alcançar a qualidade dos dados é necessária maior compreensão do ambiente organizacional e entender as necessidades dos stakeholders. Para isto, são propostas dimensões de qualidade que permitem avaliar os dados e caso atendam tais dimensões, podem ser classificados, então, como de qualidade.

Batini e Scannapieco (2016) afirmam que para avaliar a qualidade dos dados é necessário verificar as dimensões de qualidade, quando descreve seu aspecto de acordo com o contexto que este dado está inserido. São encontradas diversas dimensões de qualidade de dados na literatura. Diante desta variedade, Wang e Strong (1996), Wang (1998) e Batini e Scannapieco (2016) evidenciam que não existe um consenso

relativo ao conjunto de dimensões que devem ser considerados para realizar uma avaliação de qualidade.

Neste contexto e ao considerar a complexidade, o aumento da produção e a diversidade de dados em LMS, se faz necessário compreender o progresso e a performance dos alunos a partir destes dados, com a finalidade de descobrir relações não óbvias que apoiam tomadas de decisões pedagógicas. Para tanto, é necessária uma preocupação com a qualidade desses dados adquiridos, considerando que se percebe carência na atenção com a qualidade desses dados.

Visando esta contribuição justifica-se avaliar como e quais dimensões de qualidade de dados vêm sendo abordadas em trabalhos que discutem tomada de decisão baseada em dados educacionais, o que possibilita reconhecer o panorama sobre estas dimensões e identificar lacunas que requerem maiores esforços.

3. Procedimentos metodológicos

Os procedimentos metodológicos constituem, do ponto de vista da sua natureza em uma pesquisa básica e pela natureza de seus objetivos, uma pesquisa exploratória. No que diz respeito aos procedimentos técnicos, este trabalho se compõe de uma revisão de literatura que incluiu artigos publicados em periódicos, em que as principais áreas que dialogam com a abordagem proposta foram as áreas de ciência da informação, ciência da computação e educação. A Figura 1 apresenta as etapas de desenvolvimento deste artigo.

Figura 1 – Etapas de desenvolvimento



Fonte: Autores.

O levantamento de artigos ocorreu no mês de maio de 2021 e foram inseridos os termos [“quality of data” in education* *“quality of data” in education management* *data quality lms management*], em bases de dados como a SCIELO, ACM, ScienceDirect, IBICT e Google Scholar. A partir dos resultados foram considerados 10 artigos internacionais, publicados nos anos de 2010 a 2020.

Para elencar os artigos candidatos deste estudo considerou-se pesquisas que trabalham com dados provenientes de ambientes educacionais à distância. Finalizada a coleta, procedeu-se a leitura classificatória em duas partes, a partir da perspectiva da interação autor-texto-leitor na qual, conforme afirmam Di Raimo e Petermann (2016), se faz necessário definir, inicialmente, os objetivos e estratégias de leitura, assim como ativar conhecimentos prévios.

Diante disso, no primeiro critério foram considerados trabalhos que versam sobre tomada de decisão baseada em dados, posteriormente a abordagem direta ou indireta das dimensões da qualidade de dados apresentadas por Wand e Wang (1996) e Wang e Strong (1996). Subsequentemente, as dimensões consideradas foram mapeadas e apontadas na próxima seção, bem como uma discussão dos resultados. Procedeu-se a

identificação quantitativa do uso de cada dimensão nos artigos, com o subsídio de conceitos científicos presentes na literatura.

4. Resultados e discussões

Nos trabalhos apreciados foram identificadas seis dimensões de qualidade. O Quadro 1 apresenta a categorização dos artigos de acordo com estas dimensões seguindo a ordem de colunas: 1) Granularidade, 2) Completude, 3) Acurácia, 4) Precisão, 5) Temporalidade e 6) Consistência. Foi observado que as dimensões granularidade e completude estão presentes em 70%, seguida pela dimensão acurácia, encontrada em 60% destes trabalhos.

Quadro 1 – Categorização dos artigos de acordo com os critérios de qualidade de dados

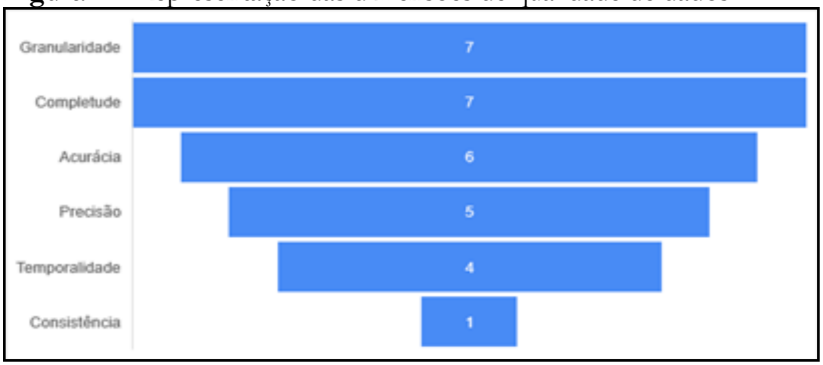
Artigos	Gran.	Comp.	Acur.	Prec.	Temp.	Con.
Ranjeeth (2020)		x	x			x
Villegas-ch (2020)	x	x	x			
Schildkamp (2019)	x	x	x		x	
Henrie (2018)	x	x	x	x	x	
Kadoic (2018)			x		x	
Cerezo (2016)	x	x		x		
You (2016)		x		x	x	
Dias (2015)	x			x		
Kazanidis (2012)	x			x		
Macfadyen (2010)	x	x	x			

Fonte: Autores.

A dimensão de precisão foi identificada em 50% dos trabalhos e as dimensões menos consideradas foram a temporalidade, observada em 40% dos trabalhos e a consistência, contemplada em 10% dos trabalhos. A Figura 1 exibe um gráfico contendo a representação das dimensões de qualidade de dados constatadas nos trabalhos.

As dimensões de qualidade de dados identificadas foram mapeadas e categorizadas (Quadro 1) conforme suas características intrínsecas. Inicialmente, foram detectadas seis dimensões (Figura 2), e se observa granularidade e completude abordadas em sete trabalhos.

Figura 2 – Representação das dimensões de qualidade de dados



Fonte: Autores.

A dimensão granularidade se refere a sua capacidade de distinguir objetos (FURBER, 2016). Para Santos e Santana (2013) a granularidade está relacionada ao conteúdo disponível no conjunto de dados e se refere ao número de atributos que

compõem este conjunto e a sua diversidade de conteúdos. Os autores complementam que a granularidade impacta diretamente nos processos de acesso e de tratamento dos dados.

A completude é definida por Wand e Wang (1996) como o grau no qual uma coleção de dados inclui referências diante da descrição de um conjunto de objetos que se relaciona com o mundo real. Batini e Scannapieco (p. 70, 2016) comentam que são propostas diferentes definições para completude, todas coerentes e complementares. Entretanto, os autores ressaltam que muitas vezes estas definições se referem a diferentes níveis de granularidade ou de modelos de dados.

A acurácia é citada em seis trabalhos e, segundo Wang e Strong (1996), é obtida quando um dado é considerado correto, confiável e certificado. Esta dimensão permite observar a sua relação com a confiabilidade, o que ressalta a sua importância, uma vez que avaliar a confiabilidade, bem como determinar se um dado é correto e certificado é de extrema relevância para o contexto da educação.

A precisão, considerada em cinco trabalhos, pode variar de acordo com suas medições sobre um determinado valor. Wang e Strong (1996) afirmam que a precisão é obtida a partir da medida em que os conteúdos estejam corretos, confiáveis e livres de erros.

A temporalidade, identificada em quatro trabalhos, é definida por Wand e Wang (1996) como o limite para o uso apropriado do dado. Bovee et al (2003) abordam dois componentes para a

dimensão de temporalidade. Primeiramente a idade, quão velho um determinado dado é, e depois, a volatilidade, que aborda sobre a instabilidade e frequência de mudança de um atributo pertencente a uma entidade. Para Liu e Chi (2002), esta dimensão é utilizada de forma a analisar se os dados estão atualizados suficientemente para realizar uma tarefa.

Por fim, a dimensão consistência, identificada em apenas um trabalho, tem sua relevância no sentido de determinar conflitos entre os dados, verificando se sua função e a posição são válidas a partir de dados anteriores (WANG; STRONG, 1996). Para alcançá-la é necessário considerar todo o contexto e avaliar a sua utilização diante do cenário. No contexto de qualidade é uma dimensão que pode impactar diretamente o resultado, uma vez que sua abordagem se baseia mediante regras semânticas sobre um determinado conjunto de dados.

As dimensões abordadas nos trabalhos demonstram a existência de lacunas no que diz respeito a pesquisas relacionadas à qualidade de dados em ambientes educacionais. O contexto educacional, em especial a modalidade à distância, apresenta especificidades próprias, especialmente por requerer uma comunicação bidirecional. Logo, o apoio tecnológico referente a obtenção de dados para subsidiar tomadas de decisões pedagógicas é extremamente importante e consequentemente a validação da qualidade dos dados também é fator decisivo para a eficiência destas decisões. Redgate (2016) afirma que, em educação, não pode haver tolerância para dados defeituosos;

quanto maior a acurácia, traz-se mais conforto a estudantes, responsáveis, professores e gestores.

Diante disso, questões relacionadas à atualidade, acessibilidade, corretude, rastreabilidade e recuperabilidade, dimensões estas sugeridas pela ISO/IEC 25012:2008, devem ser consideradas no contexto de sistemas informacionais educacionais na contribuição para uma decisão mais assertiva e que poderá acarretar melhores resultados na performance dos alunos.

5. Considerações finais

Há imensidão de dados educacionais disponíveis e gerados continuamente, traz consigo uma dificuldade de processamento e geração de informações úteis, devido à alta granularidade dos dados e por se tratar de um ambiente que comporta inúmeros contextos. Perante este desafio, é necessária a preocupação com a qualidade desses dados para gerar informações que possibilitem um melhor suporte às tomadas de decisões pedagógicas.

No atual cenário, pesquisas sobre a qualidade dos dados são um importante instrumento para a adequação dos dados e a melhoria dos ambientes educacionais. Por meio delas é possível ampliar os benefícios e o potencial da utilização dos dados. Neste contexto, as evidências encontradas neste estudo permitem concluir que existem poucos trabalhos que abordam a qualidade dos dados, além de poucas dimensões encontradas nestes trabalhos, o que possibilita a conclusão de que ainda há muitas

relações inexploradas que podem trazer maior qualidade informacional no suporte à tomada de decisão pedagógica.

Para realização deste, se faz necessário explorar as maneiras pelas quais a qualidade dos dados nos ambientes educacionais são abordadas. Para tal, a ciência da informação traz uma relevante contribuição, por ser uma área que se caracteriza pela interdisciplinaridade e que é dedicada a estudos relacionados à informação, com foco em atender as necessidades dos usuários da melhor forma possível.

Ao se levar em conta o objetivo deste trabalho, a natureza interdisciplinar e colaborativa da Ciência da Informação e os interesses relacionados aos professores e gestores, a exploração de estudos que se remetem à qualidade devem contemplar diversas óticas em múltiplas dimensões, essas ainda carentes e que podem ser consideradas em trabalhos futuros.

Com isto, acredita-se que a presença da Ciência da Informação em estudos referentes à qualidade de dados pode possibilitar a identificação, o reconhecimento de fatores e características que propiciem uma ampliação da harmonia entre os profissionais envolvidos no processo e na performance da utilização desses dados, o que traz maior desenvolvimento à área.

Referências

BATINI, C.; SCANNAPIECO, M. **Data and information quality: Dimensions, Principles and Techniques**. Cham, Switzerland: Springer International Publishing, 2016.

BOVEE, M.; SRIVASTAVA, R. P.; MAK, B. A conceptual framework and belief-function approach to assessing overall information quality. **International journal of intelligent systems**, v. 18, n. 1, p. 51-74, 2003.

CEREZO, R. et al. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. **Computers & Education**, v. 96, p. 42-54, 2016.

DI RAIMO, L. C. F.; PETERMANN, R. Perguntas de leitura: uma abordagem na perspectiva da interação. **Entretextos**, v. 16, n. 1, p. 145-169, 2016.

DIAS, S. B. et al. Fuzzy cognitive mapping of LMS users' quality of interaction within higher education blended-learning environment. **Expert systems with Applications**, v. 42, n. 21, p. 7399-7423, 2015.

FLORIDI, L.; ILLARI, P. (Ed.). **The philosophy of information quality**. Cham, CH: Springer, 2014.

FÜRBER, C. Data Quality in the Semantic Web. In: **Data Quality Management with Semantic Technologies**. Springer Gabler, Wiesbaden, 2016.

HENRIE, C. R. et al. Exploring the potential of LMS log data as a proxy measure of student engagement. **Journal of Computing in Higher Education**, v. 30, n. 2, p. 344-362, 2018.

KADOIĆ, N.; OREŠKI, D. Analysis of student behavior and success based on logs in Moodle. In: 2018 41st **International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)**. IEEE, 2018. p. 0654-0659.

KAZANIDIS, I. et al. Measuring and mining LMS data. In: 2012 16th Panhellenic Conference on Informatics. **IEEE**, 2012. p. 296-301.

LIU, L.; CHI, L. Evolutional Data Quality: A Theory-Specific View. In: **ICIQ**. 2002. p. 292-304.

MACFADYEN, L. P.; DAWSON, S. Mining LMS data to develop an "early warning system" for educators: A proof of concept. **Computers & education**, v. 54, n. 2, p. 588-599, 2010.

RANJEETH, S.; LATCHOUMI, T. P.; PAUL, P. V. A survey on predictive models of learning analytics. **Procedia Computer Science**, v. 167, p. 37-46, 2020. REDGATE, Russ. **What Matters the Most with Education Data.** eScholar, 2016. Disponível em: https://escholar.com/2016/03/why-data-quality-matters/. Acesso em 27 de maio de 2021.

SAMITSCH, C. **Data Quality and its Impacts on Decision-Making: How Managers can benefit from Good Data.** Germany: Gabler Verlag. DOI 10.1007/978-3-658-08200-0, 2015.

SANTOS, P. L. V. A. da. C.; SANTANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, 2013.

SCHILDKAMP, K. Data-based decision-making for school improvement: Research insights and gaps. **Educational research**, v. 61, n. 3, p. 257-273, 2019.

VILLEGAS-CH, W.; ROMÁN-CAÑIZARES, M.; PALACIOS-PACHECO, X. Improvement of an online education model with the integration of machine learning and data analysis in an LMS. **Applied Sciences**, v. 10, n. 15, p. 5371, 2020.

WAND, Y.; WANG, R. Y. Anchoring data quality dimensions in ontological foundations. **Communications of the ACM**, v. 39, n. 11, p. 86-95, 1996.

WANG, R. Y. A product perspective on total data quality management. **Communications of the ACM**, v. 41, n. 2, p. 58-65, 1998.

WANG, R. E.; STRONG, D. Beyond accuracy: What Data quality means to data consumers. **Journal of Management Information System**, v .12, n .4, p.5-33. 1996.

YOU, J. W. Identifying significant indicators using LMS data to predict course achievement in online learning. **The Internet and Higher Education**, v. 29, p. 23-30, 2016.