

# Design de uma Escala para Avaliação de Literacia de Dados

Luciana S. Brito<sup>1</sup>, Juliana B. S. França<sup>1</sup>, Angélica F. S. Dias<sup>2</sup>, Adriana S. Vivacqua<sup>1</sup>

<sup>1</sup>Instituto de Computação – Universidade Federal de Janeiro (UFRJ)

<sup>2</sup>Núcleo de Computação Eletrônica – Universidade Federal do Rio de Janeiro (UFRJ)

**Abstract.** Large amounts of data have been widely created and consumed since the first two decades of the twenty century. The ability to read, work, analyze and argue with data is called data literacy. These skills are not always well developed in schools, even though related aspects are mentioned in Brazil's National Curriculum. Holes in basic education will eventually lead to citizens who are not able to critically read a data-based piece and might more easily fall prey to fake news or misleading news stories. We designed a data literacy scale through Design Science Research. The artifact was evaluated positively by mathematicians, statisticians, and a data scientist from public institutions, who contributed with suggestions to its final version. Our contributions include a qualitative data literacy scale, that assesses students on sixty data literacy abilities organized into seven categories that go from "data collection" to "make informed decisions"; an actualized data literacy definition, and a framework generalizing the process of scale development.

***Resumo.** Grandes quantidades de dados foram amplamente criadas e consumidas desde as duas primeiras décadas do século XX. A capacidade de ler, trabalhar, analisar e argumentar com dados é chamada de literacia de dados. Essas habilidades nem sempre são bem desenvolvidas nas escolas, embora aspectos relacionados sejam mencionados na Base Nacional Comum Curricular brasileira. Lacunas na educação básica podem levar cidadãos a serem vítimas de notícias enganosas ou desinformação. Projetamos uma escala de literacia de dados por meio da Design Science Research. O artefato foi avaliado positivamente por matemáticos, estatísticos e um cientista de dados de instituições públicas que contribuíram com sugestões para a sua versão final. Nossas contribuições incluem uma escala qualitativa de alfabetização de dados, que avalia os alunos em sessenta habilidades organizadas em sete categorias que vão desde a “coleta de dados” até a “tomada de decisões informadas”. Foi gerada também uma definição de literacia de dados e uma estrutura que generaliza o processo de desenvolvimento da escala produzida.*

## 1. Introdução

O crescimento sem precedentes históricos da quantidade de dados que são gerados e consumidos tem impacto em vários aspectos da sociedade contemporânea, incluindo negócios, marketing, agências governamentais, desenvolvimento científico e tecnológico, área da saúde, direito, meio ambiente, pesquisas em geral e a educação

(Khan et al., 2018). Segundo Tableau (2019), os dados digitais aumentariam a uma taxa de crescimento composta anual de 42% até 2020.

Por outro lado, vêm surgindo oportunidades de exercício da cidadania a partir do uso e reuso de dados abertos gerados tanto por organizações não governamentais, quanto por agências de governo. Segundo Locke e Heppler (2018), “nota-se que há um problema muito claro no que se refere à capacidade dos cidadãos em acessar e formatar dados para reuso, pois isso requer um grau considerável de conhecimento técnico”. O problema da falta de conhecimento técnico agrava-se quando se leva em conta que os governos muitas vezes disponibilizam dados abertos sem nenhum tratamento para facilitar seu uso pela população. Desta forma, é importante capacitar cidadãos para entender o mundo através de dados e, mais do que isso, construir narrativas através deles, uma vez que uma parte importante da história deste mundo em que se vive hoje só é possível ser escrita por meio de dados (Knafllic, 2019).

Esta pesquisa se propõe a ajudar na construção de um caminho para a educação popular através do design de uma escala qualitativa que permitirá a avaliação de conhecimentos em literacia de dados e conseqüentemente a contribuição com o debate curricular nacional sobre habilidades e competências de Literacia de Dados abordadas no ensino básico.

## **2. Literacia de Dados**

Existem diferentes definições para Literacia de dados, e cada uma delas revela as necessidades e aspirações de cada contexto de pesquisa onde têm sido desenvolvidas e empregadas.

O relatório *Strategies and Best Practices for Data Literacy Education*, da Universidade de Dalhousie define alfabetização de dados como a capacidade de coletar, gerenciar, avaliar e aplicar dados de uma forma crítica (Ridsdale et al., 2015). No blog da curadora de dados ChartMogul, Shelley (2019) define *Data Literacy* como a habilidade de ler, entender, criar e comunicar dados como informação, sendo ainda a habilidade de trabalhar de forma efetiva com dados. Para Gibson e Mourad (2018), a Literacia de Dados é muito mais que numeracia e representação gráfica, sendo um conceito mais amplo que também inclui o entendimento de quais ferramentas analíticas usar, como aplicar essas ferramentas como um componente integrado de um experimento, como interpretar os dados analisados à luz de uma pergunta específica ou de uma hipótese e como comunicar resultados efetivamente em diferentes plataformas.

Para os propósitos desta pesquisa utilizamos a definição de Literacia de dados de D’Ignazio e Bhargava (2015), em que “a Literacia de Dados inclui a capacidade de ler, trabalhar, analisar e argumentar com dados. Ler dados envolve a compreensão de quais são os dados necessários para representar um universo e que aspectos desse universo esses dados representam. Trabalhar com dados envolve criação, aquisição, limpeza e gerenciamento de dados. Analisar dados envolve filtrar, classificar, agregar, comparar e realizar outras operações analíticas neles. Argumentar com dados envolve utilizar dados como suporte para uma narrativa destinada a comunicar alguma mensagem para um público específico”.

### 3. Base Nacional Comum Curricular (BNCC)

No contexto brasileiro, o debate curricular tem sido abordado pelo Ministério da Educação segundo as normas da Base Nacional Comum Curricular (BNCC) (Brasil, 2017), documento que foi homologado no ano de 2018 e que normatiza oficialmente o conjunto de aprendizagens essenciais que todo o aluno deve desenvolver durante a Educação Básica.

No quadro 1 foi proposta a organização das habilidades da BNCC que percebeu-se conter semelhança epistemológica com o escopo da Literacia de Dados oferecido pela definição de D’Ignazio e Bhargava (2015), organizadas segundo a divisão dos anos escolares vigente para o ensino médio.

**Quadro 1. Habilidades listadas pela BNCC para o Ensino Médio na área da Matemática e suas Tecnologias**

(EM13MAT102) Analisar gráficos e métodos de amostragem e pesquisas estatísticas apresentadas em relatórios divulgados por diferentes meios de comunicação, identificando, quando for o caso, inadequações que possam induzir a erros de interpretação, como escalas e amostras não apropriadas.
(EM13MAT202) Planejar e executar pesquisa amostral usando dados coletados ou de diferentes fontes sobre questões relevantes atuais, incluindo ou não, apoio de recursos tecnológicos, e comunicar os resultados por meio de relatório contendo gráficos e interpretação das medidas de tendência central e das de dispersão.
(EM13MAT408) Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.
(EM13MAT409) Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos, como o histograma, o de caixa (box-plot), o de ramos e folhas, reconhecendo os mais eficientes para sua análise.
(EM13MAT510) Investigar conjuntos de dados relativos ao comportamento de duas variáveis numéricas, usando tecnologias da informação, e, se apropriado, levar em conta a variação e utilizar uma reta para descrever a relação observada.

Este levantamento teve como objetivo identificar pontos de convergência entre o currículo de Matemática nacional e a definição de Literacia de Dados que está sendo adotada nesta pesquisa, para que, a partir dessa identificação seja possível entender melhor o cenário da Literacia de Dados no nosso país e tomar partido das reais necessidades dos nossos estudantes, docentes e sociedade.

### 4. Materiais e métodos

A solução proposta para o problema de pesquisa foi desenvolvida a partir da definição de Literacia de Dados de D’Ignazio e Bhargava (2015) utilizando a abordagem da *Design Science Research*. Primeiramente lançou-se mão da definição de Literacia de Dados e identificou-se os verbos que indicavam competências próprias da Literacia de Dados. Depois, analisou-se qual seria a relação entre as partes constituintes resultantes da decomposição da definição de Literacia de Dados de D’Ignazio e Bhargava (2015) e percebeu-se que havia grande semelhança com o processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Datasets*) - (KDD), dado que suas etapas, apesar de diferentemente nomeadas por D’Ignazio e Bhargava (2015) transitam entre as fases retiradas, como exemplo, de Roiger e Geatz (2003) (Figura 1).



**Figura 1. Fluxo de análise de dados com base na definição de literacia de dados de D'Ignazio e Bhargava (2015).**

A concepção do artefato surgiu através da reflexão derivada do desenho do fluxo de análise de dados (Figura 1), que serviu como estrutura para a percepção de poderíamos construir uma atividade em que os estudantes deveriam passar de maneira necessariamente sequencial pelas etapas descritas na Figura 1, para que ao final entregassem como resultado da atividade um texto detalhado que contasse para a audiência de pesquisadores e especialistas envolvidos uma história a partir de um dataset oferecido – uma contação de história através de dados.

O modo encontrado para articular essas aspirações foi criando um conjunto de cartas com desafios que estimulassem os estudantes a passar pelas fases do pipeline de forma reflexiva e crítica e que também oferecessem um gancho para a facilitar as respostas das solicitações propostas nas fases subsequentes. As solicitações levantadas ajudaram a projetar o conjunto de cartas de desafio da Figura 2.



**Figura 2: Conjunto de cartas de desafio denominado “Cartas para a Literacia”.**

Para avaliar o artefato, foram realizados 2 ciclos de estudo de caso. O primeiro com estudantes de uma turma de graduação em Ciência da Computação e o segundo com estudantes de uma turma de pós-graduação em Informática de uma universidade pública brasileira.

O estudo de caso foi proposto por meio de uma atividade em sala de aula, utilizando a dupla de artefatos “Cartas para a Literacia” mais um dataset em uma planilha Google Sheets. Com o objetivo de fornecer insumos para que os estudantes utilizassem a criatividade para a expressão da sua literacia de dados na resolução dos desafios propostos na atividade, foi disponibilizado para uso farto material de papelaria.

A atividade levou o tempo total de 60 minutos para ser concluída e foi realizada com a turma dividida em grupos de cerca de 8 estudantes, que responderam os desafios colaborativamente. Depois de realizada com uma turma graduação, a atividade foi realizada com uma turma de pós-graduação para verificar se haveria a possibilidade de refinamento dos resultados.

Os relatórios resultantes da realização das atividades foram analisados qualitativamente, utilizando a Teoria Fundamentada em dados. Como produto, obteve-se uma escala de literacia de dados e uma nova definição de literacia de dados. A escala na íntegra encontra-se no link do rodapé<sup>1</sup>. E a nova definição de literacia de dados diz que **Literacia de Dados “é a capacidade de coletar dados, ler dados, gerenciar dados, analisar dados, informar resultados, argumentar com base em dados e tomar decisões com base em dados”**

Como subproduto, foi gerado um framework (Figura 4) que representa a generalização do processo de desenvolvimento da escala de Literacia de Dados.



Figura 4: Estrutura que generaliza o processo de construção da escala de Literacia de Dados.

<sup>1</sup> <https://lu-brito.github.io/towardsdataliteracy/>.

#### 4. Conclusão e próximos passos

Apesar de haver na literatura alguns esforços institucionais direcionados para a avaliação de literacia de dados de estudantes, o tema é de preocupação mundial tanto de da parte de gestores quanto de educadores, devido à falta de políticas públicas e estratégias de gestão educacional de suporte à promoção e garantia da cidadania através da Literacia de Dados.

Esta pesquisa apresentou o desenvolvimento de uma escala que permite avaliar a existência de competências em Literacia de Dados em estudantes de ensino médio. Porém, esta pesquisa representa uma etapa preliminar em investigações sobre a gestão e representação de habilidades e competências e atitudes no campo de estudo. Indicamos a necessidade de expandir a aplicação destes conhecimentos considerando estudantes com necessidades educacionais especiais, bem como a produção de novos ciclos de estudo de caso que testem a escala fruto deste trabalho de pesquisa. Além disto, indicamos a criação de um sistema de gestão de conhecimento que realize a avaliação da Literacia de Dados de indivíduos de forma automatizada, considerando o atendimento às distintas competências listadas na escala produzida neste estudo.

#### Referências

- BRASIL. MINISTÉRIO DA EDUCAÇÃO. (MEC/SEED). Base Nacional Comum Curricular. 2017. Disponível em:  
[http://basenacionalcomum.mec.gov.br/images/BNCC\\_EI\\_EF\\_110518\\_versaofinal\\_site.pdf](http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_site.pdf). Acesso em: 02 dez. 2019.
- D'IGNAZIO, Catherine; BHARGAVA, Rahul. Approaches to building big data literacy. In: BLOOMBERG DATA FOR GOOD EXCHANGE CONFERENCE, 2015, New York. Proceedings [...]. [S. l.: s.n.], 2015.
- GIBSON, J. Phil; MOURAD, Teresa. The growing importance of data literacy in life science education. *American Journal of Botany*, Lancaster, v. 105, n. 12, p. 1953-1956, 2018.
- KHAN, Hammad R.; KIM, Jeonghyun; CHANG, Hsia-Ching. Toward an Understanding of Data Literacy. 2018. Disponível em: [https://www.ideals.illinois.edu/bitstream/handle/2142/100243/Khan-Hammad\\_20180417\\_V01.pdf?isAllowed=y&sequence=1](https://www.ideals.illinois.edu/bitstream/handle/2142/100243/Khan-Hammad_20180417_V01.pdf?isAllowed=y&sequence=1). Acesso em: 18 set. 2020.
- KNAFLIC, Cole Nussbaumer. *Storytelling com dados: um guia sobre visualização de dados para profissionais de negócios*. Rio de Janeiro: Alta Books, 2019.
- LOCKE, Brandon T.; HEPPLER, Jason A. Teaching Data Literacy for Civic Engagement: Resources for Data Capture and Organization. *KULA: knowledge creation, dissemination, and preservation studies*, [S. l.], v. 2, n. 1, p. 23, 2018.
- RIDSDALE, Chantel et al. Strategies and best practices for data literacy education: knowledge synthesis report. [S. l.]: Dalhousie University, 2015.
- ROIGER, Richard J.; GEATZ, M. W. *Data mining: a tutorial-based primer*. Boston: Addison Wesley, 2003.
- SHELLEY, Ed. *SaaS Metrics Refresher #8: Data Literacy*. 2019. Disponível em:  
<https://blog.chartmogul.com/saas-metrics-refresher-data-literacy/>. Acesso em: 01 jun. 2019.
- TABLEAU. 2019 Tendências de Business Intelligence. 2019. Disponível em:  
<https://www.tableau.com/pt-br/reports/business-intelligence-trends>. Acesso em: 18 set. 2021.