



DESENVOLVIMENTO DE PESQUISAS EDUCACIONAIS: implicações teórico-metodológicas, propostas e desafios da gestão de dados científicos

Iandra Maria Weirich da Silva Coelho¹

RESUMO

O artigo problematiza o uso e a gestão de dados no âmbito educacional, evidenciando desdobramentos, implicações e desafios referentes à realização de pesquisas. Para tal, apresenta-se uma proposta metodológica que articula os pressupostos da Linguística de Corpus e da Abordagem Orientada por Dados, com o objetivo de auxiliar na coleta, gerenciamento e análise. As reflexões evidenciam a necessidade de uma nova cultura no fazer científico, com base no uso de recursos tecnológicos e estratégias que possam contribuir para a realização de pesquisas nas áreas de Ensino e Educação, nos âmbitos epistemológico e metodológico. Também demonstram a relevância da abertura dos dados educacionais, visando a fortalecer o acesso, o compartilhamento, a transparência, a redução do tempo de coleta, o impacto social, a reutilização e a potencial adaptação para o ensino, possibilitando a criação de produtos educacionais e novas oportunidades de aprendizagem, de maneira inclusiva e equitativa.

Palavras-chave: Pesquisa em Educação. Dados. Ensino-aprendizagem. Ciência Aberta.

DEVELOPMENT OF EDUCATIONAL RESEARCH: theoretical-methodological implications, proposals and challenges of scientific data management

ABSTRACT

This paper problematizes the use and data management in the educational field, highlighting developments, implications and challenges related to the conduct of research. To this end, it presents a methodological proposal that articulates the assumptions of Corpus Linguistics and the Data-Driven Approach, in order to assist in the collection, management and analysis. The reflections show the need for a new culture in scientific work, based on the use of technological resources and strategies that can contribute to the realization of research in the areas of Teaching and Education, in the epistemological and methodological scope. They also demonstrate the importance of opening educational data, aiming to strengthen access, sharing, transparency, reduction of collection time, social impact and reuse and potential

¹ Doutorado em Linguística. Professora do Instituto Federal do Amazonas, Manaus, Amazonas, Brasil. Orcid iD: <https://orcid.org/0000-0003-3513-962X>. E-mail: iandrawcoelho@gmail.com.

adaptation for teaching, which can enable the creation of educational products and new opportunities learning process, inclusively and equitably.

Keywords: Research in Education. Data. Teaching and learning. Open Science.

DESARROLLO DE LA INVESTIGACIÓN EDUCATIVA:

implicaciones teórico-metodológicas, propuestas y desafíos de la gestión de datos científicos

RESUMEN

El artículo aborda el uso y la gestión de datos en el ámbito educativo, destacando los avances, las implicaciones y los retos relacionados con la realización de investigaciones. Para ello, se presenta una propuesta metodológica que articula los supuestos de la Lingüística de Corpus y el Enfoque Basado en Datos, con el fin de ayudar en el plan de recogida, gestión y análisis. Las reflexiones ponen de manifiesto la necesidad de adoptar una nueva cultura en la práctica científica, a partir del uso de recursos tecnológicos y estrategias que puedan contribuir a la realización de investigaciones en el ámbito de la Enseñanza y la Educación, al alcance epistemológico y metodológico. También demuestran la relevancia de la apertura de datos educativos, con el objetivo de fortalecer el acceso, el intercambio, la transparencia, la reducción del tiempo de recolección, el impacto social y la reutilización y potencial adaptación para la enseñanza, lo que puede posibilitar la creación de productos educativos y nuevas oportunidades para el proceso de aprendizaje de manera inclusiva y equitativa.

Palabras clave: Investigación en Educación. Datos. Enseñanza y aprendizaje. Ciencia abierta.

CONSIDERAÇÕES INICIAIS

Este artigo problematiza o uso e o gerenciamento de dados no contexto de pesquisas educacionais, trazendo à baila algumas reflexões, implicações, e desafios enfrentados, principalmente, no âmbito da pós-graduação. Nesse cenário, de constante busca pela excelência na pesquisa e nas práticas pedagógicas, destaca-se a realização de estudos sustentados pela missão integradora que une pesquisa e ensino e seus potenciais movimentos, já que “[...] não há ensino sem pesquisa e pesquisa sem ensino. Esses quefazer se encontram um no corpo do outro” (FREIRE, 2002, p. 32).

Nessa discussão, pesquisa e formação se caracterizam como elementos mutuamente dependentes, cabendo à pós-graduação

fornecer as ferramentas metodológicas e os embasamentos epistemológicos necessários para a constituição de quadros de pesquisadores-docentes de alto nível, comprometidos com a realidade do país e a produção de conhecimentos nacionais socialmente relevantes (FREITAS; SOUZA, 2018, p. 9).

Esses programas se constituem em lugares “de construção de aprendizagens, de exercício da investigação científica, de maturação quanto a teorias e tecnologias e de preparação para o exercício da docência [...]” (GATTI, 2001, p. 112). Evidenciam a busca constante da qualidade formal e do progresso do conhecimento científico, com devido interesse na avaliação e na comunicação adequada da produção acadêmica, assim como na otimização e geração de produtos que possam contribuir para melhorar a qualidade da vida coletiva e individual (LAROCCA; ROSSO; SOUZA, 2005).

Em se tratando da produção acadêmica, essas práticas epistêmicas são consideradas desafios a serem superados, tendo em vista o problema da qualidade das pesquisas na área da Educação, atrelado a “[...] aspectos como rigor, relevância, identidade, implicações e perspectivas da pesquisa educacional” (LAROCCA; ROSSO; SOUZA, 2005, p. 120), elementos importantes para a geração do conhecimento.

Deve-se levar em conta também as problemáticas apontadas pelos estudantes, relacionadas à dificuldade de escolha do tema de pesquisa, à definição do problema, à formulação, materialização e compreensão de objetivos, ao delineamento da metodologia, à coleta de dados e à definição de periódicos para submissão dos trabalhos (LAROCCA; ROSSO; SOUZA, 2005; AIELLO-SINDONI, 2009; PINTO; DE PAULA; ALVES, 2010).

Constatações dessa natureza justificam a realização deste estudo. Outros fatores levam em conta a importância da dimensão investigativa na formação profissional e a necessidade de práticas que possam auxiliar os estudantes na elaboração de projetos e artigos científicos (AIELLO-SINDONI, 2009).

Somam-se também outros argumentos, como o fato de que “a pesquisa é cada vez mais complexa, mais digital e mais baseada nos dados e no grande potencial da computação [...]” (MÉNDEZ, 2021, p. 2), evidenciando a necessidade de novas propostas que possam fomentar investigações no âmbito educacional de maneira significativa, confiável e com devido rigor metodológico.

Além das justificativas apresentadas, destaca-se a motivação acadêmica e profissional da autora, que leva em conta as dificuldades apresentadas pelos estudantes de graduação e pós-graduação, em disciplinas como Metodologia Científica e Uso de ferramentas e dados da web para a construção de pesquisas e recursos voltados para o ensino.

Dado o exposto, faz-se necessário consolidar novas ambiências de aprendizagem e pesquisa que possam estimular a autonomia investigativa dos estudantes para conduzir as etapas da investigação, permitir a exploração de diferentes problemas, otimizar o gerenciamento do tempo, possibilitar diferentes trajetórias de pesquisa, ampliar as possibilidades de construção de conhecimentos e de *insights*, bem como oferecer ferramentas e estratégias para auxiliar na construção do Estado da Arte e do processo analítico.

Com foco nessas contribuições, neste estudo, tem-se como objetivo apresentar uma proposta metodológica denominada DataCon, cuja finalidade é aprimorar e revisar o processo de coleta, gestão de dados e análise em pesquisas educacionais. Trata-se, portanto, de uma discussão teórico-metodológica que contempla os seguintes aspectos: as potencialidades dessa proposta para o âmbito acadêmico e pedagógico, assim como a importância e os desafios relacionados ao uso, reuso e gerenciamento dos dados gerados nas pesquisas.

Essa proposta envolve a delimitação conceitual de pressupostos da Linguística de Corpus (BERBER SARDINHA, 2004) e da Abordagem Orientada por Dados (JOHNS, 1991), evidenciando o potencial epistemológico e

metodológico dessas abordagens para o desenvolvimento do processo investigativo nas áreas de Ensino e Educação.

ENQUADRAMENTO TEÓRICO

O processo gradualmente complexo e recursivo entre a teoria e a experiência de conduzir um projeto de pesquisa no âmbito educativo evidencia o contínuo enfrentamento de questões epistemológicas e metodológicas e o “debate sempre aberto e problemático, especialmente no caso das ciências humanas” (BRAVIN; PIEVI, 2008, p. 117).

Nesse sentido, faz-se necessário compreender que, “por sua própria natureza, a realidade educacional é diversa, complexa e mutante” (BRAVIN; PIEVI, 2008, p. 63) e são vários os fatores e dificuldades que devem ser levados em conta durante o processo de pesquisa.

Aiello-Sindoni (2009) destaca algumas dessas dificuldades agrupadas em três dimensões: epistemológica, estratégia geral da investigação e técnicas de coleta e análise de informação empírica. Esse conjunto de dimensões foi levado em conta para a construção do DataCon.

O desenvolvimento desse *software* tem amparo teórico-metodológico nos pressupostos da Linguística de Corpus (BERBER SARDINHA, 2004), na Abordagem Orientada por Dados (JOHNS, 1991) e nos estudos de Sirvent (1995, 1999, 2007), com foco na teoria e informação empírica.

As características marcantes que compõem a gênese dessa ferramenta são as descobertas e a verificação. Com base nesses dois contextos, busca-se contribuir para: i) observar informações que justifiquem novas pesquisas; ii) constituir amostras que tenham representatividade, coerência dos dados e validade; iii) ampliar a visão do contexto da pesquisa e das variáveis de análise; iv) fomentar a credibilidade e a validade; v) potencializar a organização de amostras orientadas aos objetos específicos de investigação; vi) permitir a descrição dos contextos da pesquisa e caracterização dos ambientes; vii) servir como fonte de dados; viii) identificar lacunas com base nos estudos já publicados; ix) proporcionar a busca de dimensões e

estabelecimento de categorias analíticas; x) migrar das opiniões para o contexto da observação, em busca de “achados”; xi) otimizar a busca de informações relevantes sobre o objeto de estudo; xii) promover maior interesse e autonomia no percurso investigativo e; xiii) reduzir o tempo de coleta e ampliar o acesso aos dados, de forma dinâmica e motivadora.

A fim de fomentar as descobertas e a verificação, toma-se o *corpus* como mobilizador dos processos de coleta, análise e síntese dos resultados. Nesse sentido, a Linguística de Corpus se apresenta como uma metodologia robusta para auxiliar nas demandas das investigações científicas, cuja representatividade e pontos de interface podem dialogar com outras áreas, como a do Ensino, para além dos estudos da linguagem (Linguística Aplicada, Processamento da Linguagem Natural).

Tanto a compilação como a observação desse *corpus* vão exigir dos estudantes uma atitude autônoma e reflexiva durante o percurso investigativo, a fim de compreender o funcionamento de determinado fenômeno, considerando, especialmente que, “um *corpus* de texto oferece diferentes leituras, dependendo dos vieses que ele contém” (GASKELL, 2008, p. 191 apud MOZZATO; GRZYBOVSKI, 2011, p. 742).

Dessa forma, sob uma perspectiva heurística, a investigação orientada por dados busca auxiliar na construção do Estado da Arte referente à problemática de pesquisa e desenvolvimento de estratégias automatizadas para o gerenciamento e análises.

Nesse cenário, considerando que a construção do Estado da Arte é uma das primeiras etapas a serem desenvolvidas na investigação, tanto a compilação de um *corpus* como seu gerenciamento serão elementos fundamentais durante o processo que está diretamente vinculado às próprias inquietudes e questões de pesquisa, que surgem do envolvimento em um contexto concreto de práticas e significados aproximados à realidade educativa (BRAVIN; PIEVI, 2008, p. 121).

Por meio de ferramentas tecnológicas e estratégias adequadas é possível potencializar esse processo, a fim de obter novos conhecimentos a

partir dos questionamentos, para os quais não se admitem “[...] resultados definitivos, estabelecendo a provisoriedade metódica como fonte principal da renovação científica” (FREITAS; SOUZA, 2018, p. 74).

Diante desse cenário, destaca-se também a importância do compartilhamento e divulgação dos resultados de pesquisa, com o intuito de fomentar uma cultura diferenciada de uso e reuso dos dados oriundos do contexto educativo, evidenciando maior transparência e confiabilidade.

De acordo com Sales e Sayão (2019, p. 31), “[...] no contexto da ciência contemporânea, os dados de pesquisa deixam de ser meros subprodutos das atividades de pesquisa e ressurgem como protagonistas na busca por novos conhecimentos”. Nesse sentido, “[...] precisamos ter os dados que sustentam os principais resultados da pesquisa, disponíveis e compreensíveis, tanto para reprodutibilidade quanto para prevenir fraude/ uso indevido” (CALLAGHAN, 2019, p. 2).

Com base nesses argumentos, destaca-se a relevância de um processo, denominado neste estudo, de imersão nos dados, provenientes de uma compilação sistemática, constituídas a partir de técnicas de coleta (observação, questionário, entrevista), da web e/ou de *corpus* de aprendizes.

O processo de imersão busca estimular as análises, divulgação e socialização dos dados, já que são considerados “[...] a força vital na tomada de decisões e a matéria-prima para a responsabilidade” (THE UNITED NATIONS, 2014. p. 2). Essa divulgação permite torná-los utilizáveis por terceiros, procedimento que exige esforço e tempo e “[...] geralmente não é recompensado pelo sistema atual de obtenção de crédito acadêmico” (CALLAGHAN, 2019, p. 2).

Esse movimento histórico e progressista tem por base o desenvolvimento e a evolução de recursos educativos abertos e está alinhado aos pressupostos da Ciência Aberta, que enfatiza o livre acesso, os *softwares* livres, os dados, tecnologias e formatos abertos, cursos e outros materiais que possam ser reutilizados ou adaptados. Nesse sentido, a Ciência Aberta

não apenas nos fornece ferramentas e nos desafia, mas também representa uma rara oportunidade para revisar o atual sistema

científico, refletir sobre seu anacronismo e adaptar a ciência ao contexto atual: digital, colaborativo e orientado por dados (MÉNDEZ, 2021, p. 16).

Esses dados são cruciais para a investigação, tomada de decisões e desenvolvimento de processos e recursos educacionais, permitindo a geração de novas práticas pedagógicas e investigativas, com ênfase na produção colaborativa e no conhecimento compartilhado.

O fomento ao acesso e reuso dos dados, sobretudo os financiados por recursos públicos, evidenciam os princípios de uma filosofia aberta que apresenta um novo *modus operandi* de projetar, executar e divulgar as pesquisas desenvolvidas. Isso implica o estímulo à ampla divulgação dos dados primários que dão origem aos artigos científicos, mediante a publicação das amostras compiladas em repositórios certificados, ações que podem fomentar mudanças em diferentes âmbitos, incluindo a editoria científica.

Nessa direção, Callaghan (2019, p. 4) ressalta que, constantemente, os artigos são tomados como “principal (e único) objeto de autoridade e o conjunto de dados pode ser rebaixado como meramente ‘informações de suporte’, quando, na verdade, é o artigo que fornece as informações de suporte sobre os dados”. Contudo, “o *corpus* não é a ferramenta da pesquisa, mas o objeto da pesquisa, pelo menos o objeto que foi construído num *corpus* em função de pressupostos teóricos e metodológicos” (CHARAUDEAU, 2011, p. 16).

Nesse contexto, vale afirmar que “Sem dados, não existe ciência. Sem dados, temos opinião, “achismo”, mas não evidência. Essa é a base das ciências empíricas, aquelas que trabalham com base em evidências” (BERBER SARDINHA; SHEPHERD, 2012, p. 9). Essa constatação é corroborada por uma frase alcunhada pelo estatístico e professor universitário, William Edwards Deming (1900-1993) de que “sem dados você é apenas uma pessoa com uma opinião”.

Tais argumentos demonstram a importância dos dados, especialmente, devido à ampla disseminação e acesso a um grande quantitativo de

informações, favorecidas pelas tecnologias de informação e comunicação, que estão “[...] propiciando um aumento exponencial no volume e tipos de dados disponíveis, criando possibilidades sem precedentes para informar e transformar a sociedade [...]. Esta é a revolução dos dados” (THE UNITED NATIONS, 2014. p. 2).

Dessa forma, destaca-se a relevância em fomentar pesquisas realizadas na área de Educação e Ensino, com foco no (re)uso de dados, considerando os preceitos da descoberta e da verificação, por meio de uma avaliação mais aprofundada das informações identificadas no corpus. Assim, explicar o contexto de um fenômeno implicaria tomar a amostra na íntegra, em detrimento ao registro e à análise parcial dos dados ou inserção de resultados que não estão contemplados na amostra (viés do pesquisador) e que podem causar distorções e comprometer a validade interna.

PROCEDIMENTOS METODOLÓGICOS

9

A abordagem adotada para este estudo é de caráter teórico-reflexivo. Destaca a relevância do (re)uso dos dados, estratégias e ferramentas tecnológicas para o processo de investigação científica, tendo a informatização como elemento motivador e a pesquisa como eixo transversal no processo formativo.

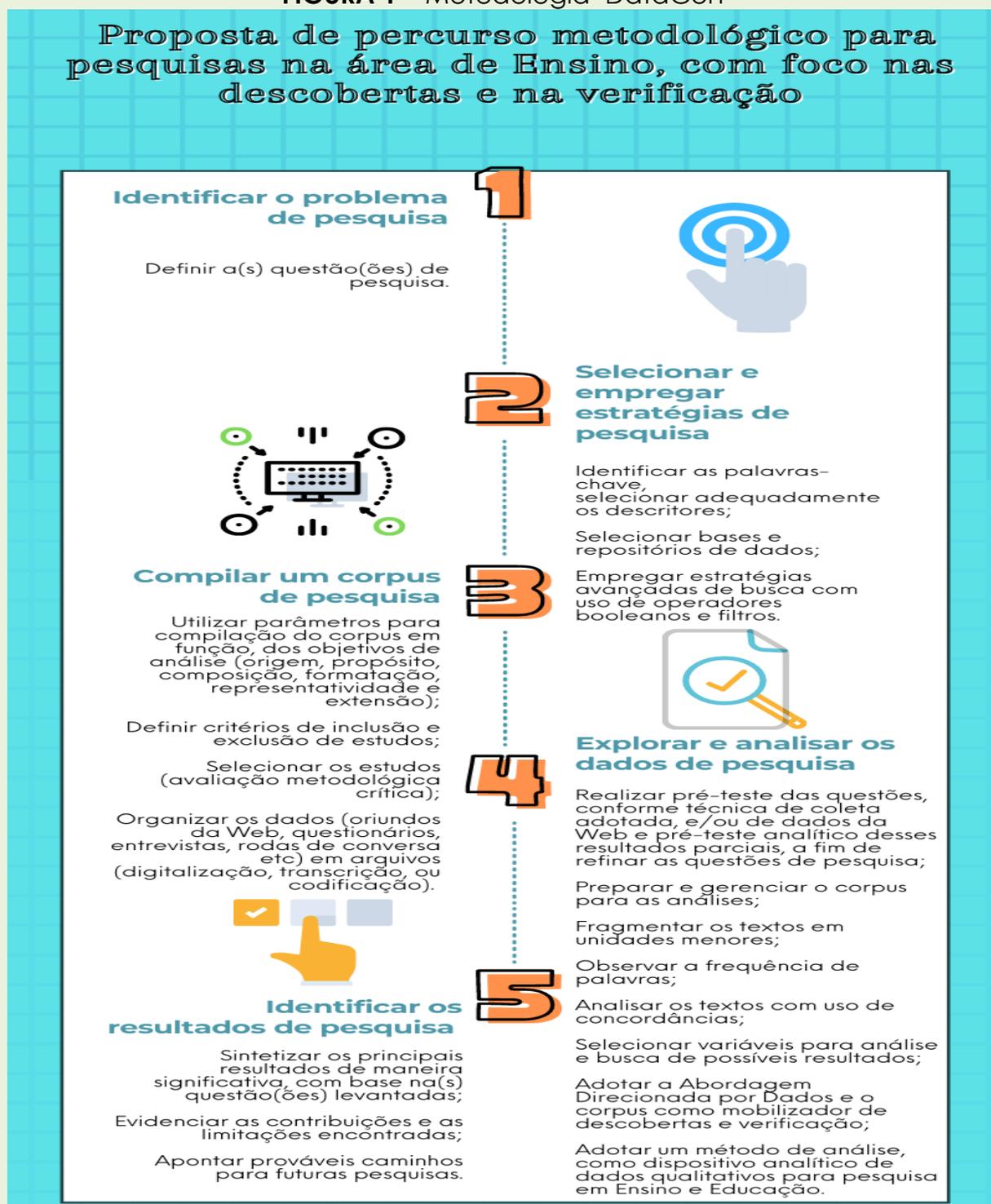
Ao utilizar ferramentas e princípios da Linguística de Corpus e da Abordagem Orientada por Dados no desenho do DataCon, evidenciam-se as oportunidades que se abrem com esse arcabouço teórico-metodológico e que podem contribuir para as pesquisas realizadas nas áreas de Educação e Ensino. Por meio dessa interface busca-se possibilitar a verificação de uma grande quantidade de dados, mediante a realização de uma análise automatizada, em um tempo mais reduzido e com devido rigor metodológico.

Essa constatação faz referência à importância dos *softwares* que facilitam a análise e a interpretação de dados qualitativos, visando ao aumento do alcance e da profundidade dessas análises. Nesse sentido, de acordo com Creswell (2007, p. 197), o material a ser analisado “[...] pode ser

melhorado com o uso de programas de computador com *software* qualitativo”, sendo, nas palavras de Bardin (2011, p. 37), “[...] algo fastidioso quando feito manualmente”.

A proposta foi construída a partir de três dimensões, com base em Sirvent (1995, 1999, 2007), a saber: epistemológica, estratégia geral e técnica. A primeira envolve aspectos como a intencionalidade, o problema, as hipóteses e a função da teoria. A segunda refere-se ao tipo de desenho (quantitativo, qualitativo, misto etc) e a seleção de casos, por meio de critérios amostrais, apoiando-se, sempre que possível, em noções básicas da representatividade estatística. E a terceira inclui a seleção e a aplicação de técnicas de obtenção, análise e validação das informações empíricas. Para atender às dimensões, a proposta se desenvolve mediante cinco etapas, conforme Figura 1.

FIGURA 1 – Metodologia DataCon



Fonte: A autora.

As etapas evidenciam os elementos constitutivos da proposta: compilação de corpus, uso de ferramental específico (software, concordanciador, listas de frequências), escolha de método de análise pertinente ao tema e à orientação epistemológica, uso de estratégias para

realização da coleta sistematizada, fragmentação de material textual, conversão de dados em variáveis e/ou categorias de análise, identificação de unidades temáticas, foco nas descobertas sobre características e propriedades do fenômeno investigado, entre outros aspectos.

Contextualizando o DataCon: possibilidades para a pesquisa e o ensino

O DataCon² integra o uso de um *software* de pesquisa qualitativa e uma proposta de percurso metodológico orientado por dados. Tem como finalidade auxiliar no desenho, gerenciamento, análise e compreensão do fenômeno, de maneira significativa, congruente, consistente e válida. As técnicas e estratégias empregadas estimulam a verificação e as descobertas, como meios potenciais e inovadores na construção de projetos de pesquisa.

O elemento relacionado à verificação está atrelado à melhora da visualização e maior processamento permitido pelo uso de *softwares* e técnicas que possibilitam a análise de uma grande quantidade de dados, com o auxílio de recursos gráficos e representações quantitativas e/ou qualitativas. Soma-se a isso, a potencial identificação de evidências e descobertas de pesquisa que, possivelmente, não seriam observadas em uma análise manual.

Com o uso do DataCon, prevê-se diminuir limitações tecnológicas que estão associados à coleta e ao armazenamento de dados, facilitar o acesso às informações compiladas e à análise de um conjunto maior de textos, a fim de fomentar a geração de percepções objetivas e resultados pautados em dados válidos e relevantes, potencialmente úteis que podem gerar novos conhecimentos, causar impacto acadêmico e social.

Busca-se, dessa maneira, fomentar a compreensão do fenômeno de pesquisa, com auxílio de descrições empíricas que emergem das análises realizadas, a partir de categorias pré-estabelecidas ou por meio da

² Cabe ressaltar que o projeto DataCon: Desenho metodológico, gerenciamento e análise de dados conta com um *software*, que está em processo de desenvolvimento. O projeto foi interrompido devido à suspensão das atividades acadêmicas, em 2020, como medida preventiva contra a pandemia do coronavírus.

exploração das diferentes relações evidenciadas pelos próprios dados. Esse processo se configura sob a perspectiva das descobertas, graças aos movimentos e relações que se estabelecem entre o pesquisador e a dimensão epistemológica, capazes de gerar *insights* valiosos (descritivos, metodológicos, teóricos) de relevância para uma análise orientada por dados.

Sendo assim, acredita-se que a proposta pode ser utilizada tanto no âmbito acadêmico como no pedagógico. No primeiro cenário, por meio desse percurso, o estudante pode evidenciar e aprofundar-se em diferentes aspectos relacionados à problemática de sua pesquisa, verificar a forma como o tema já foi tratado em estudos prévios, identificar os progressos alcançados por outros pesquisadores, observar as tendências de pesquisa sob o prisma cronológico, reconhecer lacunas, com as quais é possível desenvolver novas pesquisas, identificar contribuições e limitações existentes, com o intuito de verificar o que já foi superado e atualizar os conhecimentos sobre o objeto de estudo.

A fim de alcançar essas metas, propõe-se o uso de ferramentas da Linguística de Corpus, aplicadas à pesquisa e de estratégias metodológicas com foco no uso de pistas lexicais como unidade central de análise.

Nesse sentido, a proposta oferece um caminho, um percurso metodológico, que orienta o que fazer da investigação científica e, a partir dela, amplia-se o horizonte metodológico. Sob uma perspectiva multidimensional, é possível construir novas trajetórias e fomentar questionamentos, tais como: Que tipo de variáveis são identificadas? Quais variáveis estão relacionadas entre si? Que tipo de evidências textuais são encontradas para responder a problemática de pesquisa? Quais as possíveis relações entre os fatores analisados? Quais categorias ou unidades de sentido podem ser depreendidas do contexto textual? entre outras questões.

Por outro lado, no âmbito pedagógico, acredita-se que o professor também pode beneficiar-se dessa ferramenta para compilar um *corpus* de aprendizes, organizar dados coletados de diferentes atividades realizadas em

sala de aula, otimizar o tempo do gerenciamento das informações coletadas, potencializar a curadoria dos dados oriundos dos contextos de aprendizagem, ter em mãos possíveis evidências que demonstram a evolução das aprendizagens e dificuldades dos estudantes, assim como salvaguardar os dados educacionais, para que possam ser analisados sob uma perspectiva cronológica e/ou utilizados como recursos informacionais imprescindíveis para futuras pesquisas.

Isso implica considerar a relação entre a prática profissional e a investigação educativa, compreendendo que mediante a compilação, gerenciamento e análise de dados oriundos do contexto de ensino é possível realizar uma (auto)avaliação dos desempenhos alcançados, compreender os resultados de aprendizagem, permitir maior nível de detalhamento para a tomada de decisões, impulsionar novas oportunidades, assim como estimar e promover inovação para as lacunas existentes no âmbito educacional.

Dessa forma, “expostos aos dados, professores e alunos discutem e, em conjunto, problematizam possíveis caminhos para a produção do conhecimento” (ROCHA, 2017, p. 142), mediante uma postura “dialógica, aberta, curiosa, indagadora e não apassivada [...]”. O que importa é que professor e aluno se assumam epistemologicamente curiosos” (FREIRE, 2002, p. 96).

O professor também pode problematizar as especificidades de sua própria prática, transpondo os limites do ambiente acadêmico, com o intuito de entender a relação dialética entre teoria e realidade, de forma contínua e dinâmica, e fortalecer o processo de ensino-aprendizagem, a partir da transversalidade da pesquisa e da disposição investigativa, considerando que uma não deveria existir sem a outra.

Em ambos os cenários, com o uso da proposta, busca-se investir em um percurso autônomo e emancipatório, por meio do uso de recursos tecnológicos e estratégias com as quais, tanto estudantes, como professores (atores que planejam, executam e responsabilizam-se diretamente pela atividade de pesquisar e publicar os resultados de pesquisa) podem gerenciar

informações, promover (trans)formações e (re)configurar a cultura investigativa.

Vale ressaltar que a utilização da proposta envolve diferentes processos cognitivos e metacognitivos, por meio de um percurso imersivo que permite a construção de novas trajetórias investigativas. Para tanto, faz-se necessária a utilização de estratégias que permitam movimentos de idas e vindas, aproximação e distanciamento dos dados, relações subjetivas com o fenômeno, promoção de atividades sistemáticas e criativas, desenvolvimento de habilidades de busca, seleção, coleta de dados, fomentando o *input* potencial de pesquisa, com devido rigor metodológico para o desenho, recorte, gerenciamento, realização de análises e inferências.

Tais movimentos epistêmicos visam a contribuir para o desenvolvimento de um percurso investigativo autônomo, reflexivo e colaborativo, redução das incertezas sobre os conhecimentos adquiridos a respeito do objeto de estudo, e a diminuição das interferências das percepções subjetivas que podem influenciar na coleta dos dados e na interpretação dos dados, sujeita, muitas vezes, a diferentes vieses dos pesquisadores.

Nesse sentido, no momento das análises, evidencia-se a relevância da escolha de um método, de acordo com o posicionamento epistemológico e com os objetivos da pesquisa, e a utilização de *softwares* qualitativos, “os quais não têm o intuito de quantificar ou criar um viés positivista” (MOZZATO; GRZYBOVSKI, 2011, p. 743).

Dessa forma, a inferência do pesquisador, a qual se faz necessária, não é totalmente neutra; no entanto ele deve procurar interferir minimamente de maneira pessoal. Isso não quer dizer ser acrítico ou não fazer inferências, aspectos que uma análise de conteúdo em pesquisas qualitativas exige (MOZZATO; GRZYBOVSKI, 2011, p.740).

A análise de conteúdo é citada como uma das possibilidades. Trata-se de uma técnica em crescente utilização e legitimação nos estudos qualitativos no âmbito educacional, cujo propósito é “ultrapassar o senso comum do subjetivismo e alcançar o rigor científico necessário” (MOZZATO; GRZYBOVSKI, 2011, p. 736). Apesar da influência positivista que sofre, devido ao fato de ter em sua origem a quantificação,

logo se compreendeu que esta técnica pode ser aplicada também na análise qualitativa, pois sua característica é a inferência (variáveis inferidas a partir de variáveis de inferência no nível da mensagem), quer estas estejam baseadas ou não em indicadores quantitativos (MOZZATO; GRZYBOVSKI, 2011, p. 736-7).

A fim de extrair sentido dos dados e buscar maior profundidade nas análises, é necessário, em alguns casos, (des)construir experiências investigativas, com o objetivo de consolidar novas práticas que tenham como foco a análise e a explicação interpretativa, constituídas de maneira criativa e flexível, em que se integram teoria e empiria, objetividades e subjetividades, com devida observação do contexto global, local, social e histórico dos participantes.

Isso implica a adoção de uma visão multidimensional que envolve tanto a hermenêutica (processo reflexivo e analítico), como a métrica, em que procura-se estabelecer relação do pesquisador com os objetos/sujeitos da pesquisa e com a literatura científica, identificar evidências passíveis de responder as questões norteadoras, possibilitar a descrição, a classificação e o agrupamento de categorias, de acordo com a área, finalidade, objetivos e métodos empregados.

Essa demanda investigativa, portanto, é impulsionada pelos dados, devidamente organizados (*corpus*), por recursos tecnológicos, especialmente, os *softwares* que oportunizam as condições para facilitar as análises qualitativas (BARDIN, 2011; MOZZATO; GRZYBOVSKI, 2011) e por desenhos sistemáticos explícitos, passíveis de reprodução e reuso em futuras investigações, como é o caso do DataCon. Apresenta potencial para a extração, análise e síntese dos resultados de forma exponencial, compreensível e útil para a geração de novas ideias e a construção de conhecimentos.

Para melhor esclarecimento e considerando o exposto até aqui, vale ressaltar que não se trata de uma visão simplista e prescritiva do uso de técnicas e estratégias na realização das pesquisas, devido, principalmente, às “limitações intrínsecas de qualquer método de investigação científica derivada dos contextos específicos que surgem” (LEMUS-DELGADO; PEREZ

NAVARRO, 2020, p. 52) e ao fato de que a epistemologia empirista não é tomada como uma forma superior de conhecimento (LEMUS-DELGADO; PEREZ NAVARRO, 2020).

Os resultados, portanto, estão interrelacionados a diferentes elementos como a amostra, as ferramentas utilizadas, as teorias que amparam e certos pontos de vista e perspectivas do pesquisador, já que os fenômenos não se abstraem dos contextos de maneira neutra e objetiva, como verdades absolutas. Isso implica compreender que “os dados, por mais vastos que sejam, não significam nada até que damos a eles uma interpretação, que é formulada e reformulada a partir de contextos específicos do saber” (LEMUS-DELGADO; PEREZ NAVARRO, 2020, p. 58).

Antes de concluir essa seção, cabe ainda destacar a relevância do Datacon, considerando alguns trabalhos correlatos. Com base em uma investigação prévia, foi possível perceber que é recorrente o interesse pela temática relacionada ao ensino de Metodologia Científica, com foco para o uso de gamificação (SANTOS; SILVA; MATTAR, 2017) e de jogos digitais “[...] que se propõem a desenvolver nos alunos diversas habilidades de pesquisa e de pensamento científico” (MATTAR; SOUZA; BEDUSCHI, 2017, p. 3). De maneira geral, esses recursos têm como intuito apoiar o ensino dessa disciplina na graduação.

Outros trabalhos ainda versam sobre “softwares que podem facilitar os procedimentos de análise, otimizando tempo e gerando análises mais ricas” (SCHLOSSER; FRASSON; CANTORANI, 2019, p. 539), sugeridos como ferramentas de apoio ao pesquisador.

Apesar da existência desses estudos que tratam sobre o ensino de metodologia e uso de *softwares*, ressalta-se o caráter diferenciado do DataCon, caracterizando-se como uma ferramenta que contempla diferentes dimensões: tecnológica, epistemológica e metodológica, e objetiva a (re)construção dos conhecimentos por meio da verificação e das descobertas de pesquisa.

Além do exposto, outro aspecto inovador deve-se à filosofia principal da proposta, que visa a estimular o planejamento da gestão dos dados de pesquisa na área educacional. Esses dados que são coletados, utilizados, processados, analisados e gerados são elementos essenciais e integrais do percurso metodológico. Para tanto, faz-se necessário planejar a coleta, realizar a limpeza de dados, compilar amostras robustas e confiáveis e armazenar, de forma adequada, para que esses recursos informacionais possam ser acessíveis e reutilizáveis.

Contextualizando o DataCon: possibilidades e desafios no contexto da Educação Aberta e da Ciência Aberta

Neste estudo, também ganham espaço algumas reflexões sobre o uso da proposta metodológica DataCon, pensada a partir dos movimentos da Educação Aberta e da Ciência Aberta, com especial destaque para a criação de *softwares* livres e (re)uso de dados abertos educacionais, gerados em ambientes de pesquisa.

Segundo Guy (2016, p. 166), os dados educacionais abertos são uma área de interesse relativamente nova, com escassa literatura, sendo necessário, portanto, ampliar os estudos, a fim de responder questões tais como: “[...] o que são exatamente dados educacionais abertos, por que os dados abertos são relevantes para a educação e como podemos usar conjuntos de dados abertos para atender às necessidades educacionais?”.

Para a autora, o termo “dados educacionais abertos” ainda permanece vagamente definido, podendo ser usado para fazer referência a todos os dados disponíveis abertamente que possam ser usados para fins educacionais (caracterizados como um subconjunto de recursos educacionais abertos que são disponibilizados para uso no ensino-aprendizagem) e a dados abertos que são divulgados por instituições acadêmicas, relacionados ao desempenho institucional e aos estudantes (dados do curso, currículo, objetivos de aprendizagem, avaliações de desempenho etc).

Ferreira, Rodrigues e Souza (2021, p. 1194) argumentam que a área de pesquisa relacionada aos dados educacionais abertos vem crescendo nos últimos anos. Contudo, de acordo com os autores, “[...] esta área de pesquisa ainda precisa avançar, trazendo soluções para que a comunidade educacional possa visualizar os dados abertos e tomar decisões para melhoria dos seus processos educacionais”.

A urgência desses avanços deve-se ao fato de que “o processo de pesquisa não consiste apenas em aprender e descobrir, mas também em compartilhar essas descobertas com outras pessoas, para que a sociedade como um todo possa se beneficiar dos esforços do indivíduo” (CALLAGHAN, 2019, p. 2).

Sob essa premissa e considerando que “dados de pesquisa podem e devem ser vistos sob diversas perspectivas” (SALES; SAYÃO, 2019, p. 46), neste estudo, destaca-se como um dos possíveis caminhos para avançar na abertura, compartilhamento e reutilização dos dados, a compilação de *corpus* em formato de produto educacional. Configura-se como um produto interlocutivo, podendo ser utilizado como uma fonte de referência em investigações científicas e desenvolvimento de propostas de ensino-aprendizagem voltadas para contextos educacionais formais e não-formais.

A fim de ilustrar tal possibilidade, cita-se como exemplo o Banco de Dados Linguísticos Reci, compilado durante os anos de 2016 e 2017 (SILVA; COELHO, 2018). Esse *corpus* é composto pela transcrição de entrevistas de 36 falantes, do núcleo da Reserva Extrativista Cazumbá-Iracema, localizado no Acre. Sua organização, em formato de produto educacional, levou em conta a necessidade de ampliar a visibilidade, a replicabilidade, o impacto e o (re)uso de dados linguísticos, tanto no contexto pedagógico (criação de propostas de ensino-aprendizagem de Língua Portuguesa, baseadas na utilização de dados reais), como no contexto acadêmico (geração de pesquisas no âmbito da Linguística, com análises exploratórias e empiricamente fundamentadas).

Esse tipo de produto educacional é uma possibilidade de atender a demanda por dados, fomentar uma nova cultura investigativa, gerando amostras que podem ser abertas e reproduzíveis, por meio de repositórios institucionais específicos, potencializando o acesso, a aplicabilidade e os impactos a serem gerados na comunidade científica da área e na sociedade.

Isso implica ir além do uso restrito pelo pesquisador (em um período determinado da formação acadêmica - mestrado ou doutorado), permitindo que o *corpus* da pesquisa possa ser disponibilizado por tempo indeterminado, localizável, acessível, interoperável e reutilizável. Assim, “uma vez que os conjuntos de dados são depositados no arquivo permanente, eles se tornam produtos fixos e são citáveis e publicáveis” (CALLAGHAN, 2019, p. 6).

Nessa perspectiva, busca-se um movimento de ruptura com as práticas tradicionais de pesquisa na área educacional, com o objetivo de consolidar “maior detalhamento de metodologias e gerenciamento de dados, de forma que eles possam ser acessados por toda a sociedade”, por meio da adoção de licenças que permitam a reutilização e a redistribuição.

Para tanto, é imprescindível a observação dos “[...] aspectos éticos preconizados por documentos regulatórios no âmbito de cada país” (FREITAG et al., 2021, p. 5) e a organização sistemática e rigorosa com relação à seleção e compilação das amostras, levando em conta critérios relacionados à compreensibilidade e à usabilidade.

Esse processo envolve uma série de práticas, limitações e desafios relacionados a: mudanças de atitudes por parte do pesquisador, exigência de esforço e tempo, aperfeiçoamento da infraestrutura tecnológica, armazenamento em repositórios confiáveis, demanda de investimentos, assim como as particularidades dos diferentes tipos de amostras e suas condições, com relação à documentação de apoio, geração de metadados e atendimento às necessidades das comunidades de usuários.

CONSIDERAÇÕES FINAIS

Na tessitura das reflexões, apresentou-se uma proposta metodológica para a construção de projetos de pesquisa, com foco no desenho, busca, seleção, gerenciamento, exploração e desenvolvimento analítico de dados científicos. As possibilidades evidenciadas com o uso do DataCon demonstram um percurso investigativo autônomo, apoiado em evidências textuais.

Com o intuito de tornar o processo de investigação na área educacional mais eficiente e menos complexo, foram destacadas potencialidades da ferramenta para o âmbito acadêmico e pedagógico, sob a égide das descobertas, da verificação, da imersão, do consumo de dados, assim como do uso de estratégias metodológicas e tecnológicas.

Também ganharam especial atenção algumas reflexões a respeito da importância da Ciência aberta e da Educação Aberta, evidenciando a incorporação inclusiva, equitativa e colaborativa de tecnologias, práticas pedagógicas e formatos abertos apoiados na criação, (re)uso, combinação e (re)distribuição de recursos e dados educacionais.

Em consonância com esses paradigmas, fomenta-se um movimento de partilha dos dados gerados, de forma colaborativa, transparente, sustentável e confiável, a fim de permitir a abertura ao diálogo com outros pesquisadores em direção a interdisciplinaridade. Dada essa ênfase, destacou-se a elaboração de produtos educacionais (*corpus* de dados linguísticos) e produções técnicas, desenvolvidos a partir dos dados, como uma das possibilidades em prol da geração de pesquisas informadas, reprodutíveis e reutilizáveis.

Essa produção, além de promover novas aprendizagens e gerar impacto social, também pode fomentar possibilidades de participação em novas redes de colaboração com outros pesquisadores e instituições, potencializar a validação dos resultados e imprimir qualidade nas práticas educativas e nas pesquisas publicadas, levando em conta a transparência em busca do avanço do conhecimento nas áreas de Ensino e Educação.

Entre as principais potencialidades evidenciadas para o uso do DataCon, citam-se: obter e processar um conjunto maior de dados, fomentar a construção de produtos educacionais acessíveis e reutilizáveis, transgredir e inovar a maneira como os estudantes têm realizado o percurso metodológico das pesquisas na área das Ciências Humanas e Sociais, potencializar a identificação e análise de tendências, limitações, evidências de aprendizagens e avanços no campo educacional, reconhecer as principais lacunas de investigação sobre o fenômeno de pesquisa, entre outras.

Para concluir a discussão, projeta-se o interesse em estudos futuros que possam tratar das contribuições, limitações e implicações do DataCon em contexto de aplicação. Outras reflexões podem somar-se, contemplando os desafios impostos pelas novas práticas relacionados à gestão de dados (educacionais) abertos que afetam os pesquisadores e as instâncias de publicação, a saber: as condições de armazenamento, disponibilização, a recuperação de (meta)dados, o sigilo e direito de uso dos dados, as licenças de uso, os aspectos éticos e de integridade, demonstrando um campo fértil para estudo, intercâmbio e inovação.

REFERÊNCIAS

AIELLO-SINDONI, M. Dificultades en el aprendizaje de la metodología de la investigación Magis. **Revista Internacional de Investigación en Educación**, v. 2, n. 3, 2009, p. 141-155. Disponível em:

<https://www.redalyc.org/pdf/2810/281021558008.pdf>. Acesso em: 15 fev. 2022.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BERBER SARDINHA, Tony. **Linguística de corpus**. Barueri, São Paulo: Editora: Manole, 2004.

BERBER SARDINHA, T.; SHEPHERD, T. M. G. O Corpus, novas tecnologias e mídias no ensino de inglês: para uma pedagogia do terceiro milênio. In: BERBER SARDINHA, T. et al. **Tecnologias e Mídias no Ensino de Inglês**: o corpus nas receitas. São Paulo: Macmillan, 2012, p. 6-14.

BRAVIN, C.; PIEVI, N. **Documento Metodológico Orientador para la Investigación Educativa**. Coordinación de Investigación del INFD, Argentina, 2008.

CALLAGHAN, S. Research Data Publication: Moving Beyond the Metaphor. **Data Science Journal**, v. 18, n.39, 2019, p. 1–7.

CHARAUDEAU, P. “Dize-me qual é teu corpus, eu te direi qual é a tua problemática”. **Revista Diadorim**, v.10, 2011, p. 1-23. Disponível em: <https://revistas.ufrj.br/index.php/diadorim/article/view/3932/15637>. Acesso em: 15 fev. 2022.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**, 2. ed. Porto Alegre: Artmed, 2007.

FERREIRA, L. A.; RODRIGUES, R. L.; SOUZA, R. N. P. M. Dados Abertos Educacionais Brasileiros: um Mapeamento Sistemático da Literatura. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 32, 2021, Porto Alegre. **Anais...** Porto Alegre: Sociedade Brasileira de Computação, 2021, p. 1186-1195. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18141/17975>

FREIRE, P. **Pedagogia da autonomia: saberes necessários à prática educativa**. 12 ed. São Paulo: Paz e Terra, 2002.

FREITAG, R. M. K. et al. Desafios da gestão de dados linguísticos e a Ciência Aberta. **Cadernos de Linguística**, v. 2, n. 1, 2021, p. 1-19. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/307>. Acesso em: 15 fev. 2022.

FREITAS, M. F. Q.; SOUZA, J. Pensar a formação e a pesquisa na pós-graduação stricto sensu. **Educar em Revista**, Curitiba, v. 34, n. 71, 2018, p. 9-18.

GUY, M. The Open Education Working Group: Bringing People, Projects and Data Together. In: MOUROMTSEV, D.; D'AQUIN, M. (Orgs.). **Open Data for Education: Linked, Shared, and Reusable Data for Teaching and Learning**. Suíça: Springer, 2016, p.166-187.

JOHNS, T. Should you be persuaded: two samples of data-driven learning materials. In: JOHNS, T.; KING, P. **Classroom concordancing**. Birmingham, England: Centre for English Language Studies, University of Birmingham, 1991, p. 1-16.

LAROCCA, P.; ROSSO, A. J.; SOUZA, A. P. A formulação dos objetivos de pesquisa na pós-graduação em Educação: uma discussão necessária.

Revista Brasileira de Pós-Graduação, v. 2, n. 3, 2005, p. 118-133. Disponível em: <https://rbpg.capes.gov.br/index.php/rbpg/article/view/62/59>. Acesso em: 1 fev. 2022.

LEMUS-DELGADO, D.; PEREZ NAVARRO, R. Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos. **Colombia Internacional**, n.102, 2020, p.41-62. Disponível em: <http://www.scielo.org.co/pdf/rci/n102/0121-5612-rci-102-41.pdf>. Acesso em: 15 fev. 2022.

MATTAR, J., SOUZA, A. L. M.; BEDUSCHI, J. O. Games para o ensino de metodologia científica: revisão de literatura e boas práticas. **Educação, formação e tecnologias**. v. 10, n.1 p. 3-19. Disponível em: <https://eft.educom.pt/index.php/eft/article/view/573/262>. Acesso em: 15 fev. 2022.

MÉNDEZ, E. Open science por defecto. la nueva normalidad para la investigación. **ARBOR Ciencia, Pensamiento y Cultura**, Vol. 197, n.799, 2021, p. 1-20. Disponível em: <https://arbor.revistas.csic.es/index.php/arbor/article/view/2402/3600>. Acesso em: 15 fev. 2022.

MOZZATO, A. R.; GRZYBOVSKI, D. Análise de conteúdo como técnica de análise de dados qualitativos no campo da administração: potencial e desafios. **Revista de Administração Contemporânea**, v.15, n.4, 2011, p. 731-747. Disponível em: <https://rac.anpad.org.br/index.php/rac/article/view/874>. Acesso em: 15 mar. 2022.

PINTO, P. S. B.; DE PAULA, M. M.; ALVES, F. J. S. A relação entre as experiências com a pesquisa científica nos cursos de graduação e as dificuldades para elaboração de artigos no mestrado acadêmico. **SINERGIA**, v.14, n.1, 2010, p. 21-33. Disponível em: <https://periodicos.furg.br/sinergia/article/view/1796>. Acesso em: 11 mar. 2022.

ROCHA, J. M. P. Abordagem orientada por dados: desafios e aplicações no ensino de LE. **Revista do GEL**, v. 14, n. 3, 2017, p. 135-153. Disponível em: <https://revistas.gel.org.br/rg/article/view/1897>. Acesso em: 11 mar. 2022.

SALES, L.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, 2019, p. 31-48. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337>. Acesso em: 3 fev. 2022.

SANTOS, B. R.; SILVA, C. C.; MATTAR, J. Gamificação no ensino dos princípios da metodologia científica. **Revista Sapere**, 2017, p. 107-125. Disponível em:



<https://www.revistasapere.inf.br/site2/artigos/fael/08%20Santos>. Acesso em: 3 fev. 2022.

SCHLOSSER, D. F.; FRASSON, A. C., CANTORANI, J. R. H. Softwares livres para análise de dados qualitativos. **Revista Brasileira de Ensino de Ciência e Tecnologia**, v. 12, n. 1, 2019, p. 539-550. Disponível em: <https://periodicos.utfpr.edu.br/rbect/article/view/9550>. Acesso em: 3 fev. 2022.

SILVA, R. G.; COELHO, I. M. W. S. C. **Banco de Dados Linguísticos Reci – Corpus escrito**. DVD (PDF) Digital, 2017. Disponível em: https://drive.google.com/file/d/1YekKfpwDUZ23VMZG-mOqE1PWZmUmOQ_w/view. Acesso em: 10 jan. 2022.

SIRVENT, M. T. **El proceso de investigación**. Las dimensiones de la metodología y la construcción del dato científico. Ficha I. Universidade Nacional de Buenos Aires, Buenos Aires, 1995.

SIRVENT, M. T. **Cuadro comparativo entre lógicas según dimensiones del diseño de investigación**. Ficha III. Universidade Nacional de Buenos Aires, Buenos Aires, 1999.

SIRVENT, M. T. **El proceso de investigación**. Investigación y Estadística Educacional I. Ficha de Cátedra, 2007.

THE UNITED NATIONS. **A World that Counts Mobilising the Data Revolution for Sustainable Development**, 2014. Disponível em: <https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>. Acesso em: 15 fev. 2022.

Recebido em: 05 de Novembro de 2021.

Aprovado em: 05 de Agosto de 2022.

Publicado em: 22 de Agosto de 2022.

