

## ***Learning Analytics* Desconectada: Um Estudo de Caso em Análise de Produções Textuais**

Elyda Freitas<sup>1,2</sup>, Hyan H. N. Batista<sup>3</sup>, Gabriel A. Barbosa<sup>3</sup>, Moesio Wenceslau<sup>3</sup>, Carlos Portela<sup>3</sup>, Seiji Isotani<sup>7</sup>, Thiago Cordeiro<sup>6</sup>, Ig Ibert Bittencourt<sup>6,7</sup>, Koiti Yasojima<sup>5</sup>, Álvaro Sobrinho<sup>3</sup>, Rodrigo Lisboa<sup>4</sup>, Rafael Ferreira Mello<sup>2,3</sup>

<sup>1</sup> Universidade de Pernambuco

<sup>2</sup> Centro de Estudos e Sistemas Avançados do Recife (CESAR)

<sup>3</sup> Universidade Federal Rural de Pernambuco

<sup>4</sup> Universidade Federal do Pará

<sup>5</sup> Universidade Federal Rural da Amazônia

<sup>6</sup> Universidade Federal de Alagoas

<sup>7</sup> Harvard Graduate School of Education

elyda.freitas@upe.br, rafael.mello@ufrpe.br

**Resumo.** *A utilização de Learning Analytics (LA) traz consigo diferentes benefícios às instituições de ensino. Porém, exige recursos computacionais e de internet inacessíveis às populações de baixa renda, tornando esta uma tecnologia que pode gerar desigualdade. Nesse contexto, este artigo tem dois objetivos: (i) apresentar o conceito de LA Desconectada, que permite a aplicação dessa tecnologia em ambientes com recursos limitados; e (ii) apresentar uma aplicação real para correção de produção textual de alunos de escolas públicas brasileiras, compatível com este conceito. O aplicativo permite a correção offline de redações escritas no papel e apresenta um dashboard impresso e com informações sumarizadas aos professores.*

### **1. Introdução**

Ao longo dos últimos anos, a tecnologia tornou-se uma ferramenta essencial para apoiar alunos e professores em diferentes atividades educacionais. Por exemplo, a utilização de ambientes virtuais de aprendizagem se tornou frequente, tanto em cursos presenciais, quanto a distância. Esse fenômeno fez crescer consideravelmente a quantidade de dados educacionais disponíveis, que podem ser usados para criação de aplicações de Inteligência Artificial (IA) [Gaftandzhieva et al. 2020]. O uso de IA na educação pode ajudar a melhorar o desempenho de alunos [Varanasi et al. 2018] e também apoiar práticas dos professores [Jivet et al. 2018].

Nesse contexto, surgiu o conceito de *Learning Analytics* (LA). Segundo a *Society for Learning Analytics Research* (SoLAR<sup>1</sup>), LA pode ser definida como o processo de: “*medição, coleta, análise e descrição de dados sobre estudantes e seus contextos, com o propósito de entender e otimizar o aprendizado e os ambientes em que ocorrem*”. LA

<sup>1</sup><https://www.solaresearch.org/>

é um campo de pesquisa recente, que se baseia na utilização de dados de interação de estudantes com plataformas digitais [Freitas et al. 2020].

Contudo, existe um problema crítico em vários países do Sul global: nem todas as instituições possuem acesso a infraestrutura (isto é, conexão com a internet e dispositivos) adequada para coletar dados das interações educacionais, para que possam, conseqüentemente, se beneficiar das aplicações de LA [Gašević 2018]. Além disso, a maioria das aplicações da área é focada e experimentada em países que possuem acesso mais adequado à infraestrutura necessária, o que pode gerar um aumento da desigualdade no acesso à educação [Puigjaner 2016]. Todo esse ciclo é oposto aos esforços mundiais para garantir educação de qualidade para todos [Ossai 2022].

Diante desse contexto, este artigo apresenta um projeto que vem sendo desenvolvido em escolas públicas brasileiras, com o objetivo de sistematizar o uso de LA com recursos limitados. Mais especificamente, o projeto é uma parceria com o Ministério da Educação, com foco na recuperação de aprendizagem relacionada a produções textuais.

## **2. Learning Analytics Desconectada**

Vários benefícios da utilização de LA podem ser identificados na literatura, por exemplo, apoio institucional para evitar evasão, auxílio aos alunos na realização de atividades (de leitura, matemática, produção textual, entre outros) e suporte ao envio de *feedback* de qualidade sobre as atividades dos estudantes [Freitas et al. 2020, Sousa et al. 2021]. Esses benefícios ilustram as várias potencialidades dessa área de pesquisa.

Entretanto, a elevada desigualdade no acesso à tecnologia necessária para utilização de LA é uma realidade preocupante. Por exemplo, em um relatório recente da UNICEF, destaca-se que aproximadamente dois terços das crianças do mundo não possuem acesso à internet em sua residência [Thompson 2020]. Além disso, com a pandemia da COVID-19, foi possível evidenciar algumas realidades mundiais já existentes, que, muitas vezes, são negligenciadas. Estudos que analisam o cenário de acessibilidade à educação em diferentes países da América do Sul, África e Ásia mostraram que muitas escolas não possuem as condições necessárias para oferecer suporte a uma educação de qualidade [Lim et al. 2020, Rodríguez and Cobo 2022, Matias et al. 2022]. Essa realidade não é diferente no Brasil, onde muitas escolas não têm acesso à internet, nem equipamentos necessários para atividades educacionais básicas, muito menos para utilização de LA [Perine and Rowsell 2019, Stevanim et al. 2020].

Para sobrepor essas dificuldades, algumas iniciativas começaram a ser propostas com o objetivo de democratizar o acesso à educação em diferentes países. Por exemplo, a utilização de MOOC foi sugerida como uma possibilidade para expandir o acesso à educação de qualidade [Rambe and Moeti 2017]. Além disso, vários estudos apresentam possibilidades de direcionamento futuro para lidar com esse problema [Tlili et al. 2021, Renz and Hilbig 2020].

Como exemplos concretos de aplicações que buscam ampliar o acesso à utilização de LA no contexto de instituições com poucos recursos, podem ser citadas aquelas que apoiam-se na utilização de dispositivos móveis de baixo custo para capturar imagens de atividades desenvolvidas em papel, sendo o processamento realizado em um servidor externo. Nesta linha, existem aplicações relacionadas a diferentes áreas, como matemática [Zanibbi and Blostein 2012] e produção textual [Sharma and Jayagopi 2018].

Diante deste contexto, os autores deste artigo propõem o conceito de *Learning Analytics* Desconectada (do inglês *Unplugged Learning Analytics* - ULA). ULA se caracteriza por aplicações que utilizam dispositivos de baixo custo, em contextos com acesso à internet de baixa qualidade, para capturar dados dos alunos em atividades *offline* e gerar visualizações que possam ser interpretáveis e informativas quando impressas (e não apenas em *dashboards* interativos, que é o mais comum na área). Por exemplo, nesta definição, a captura de uma produção textual que os alunos desenvolveram no papel é considerada uma atividade *offline*.

Além dos exemplos mencionados, em uma pesquisa recente, os autores de [Patel et al. 2022] propuseram um sistema tutor inteligente para problemas de matemática, que pode ser classificado como ULA, pois são coletadas informações *offline*, com a sugestão de envio de um relatório impresso para o aluno. Contudo, os autores focam apenas na técnica de reconhecimento de caracteres matemáticos, sem considerar as especificidades de todo o processo, as quais tornam a aplicação democrática.

Além de contribuir com a definição do termo *Learning Analytics* Desconectada, este artigo tem por objetivo apresentar uma aplicação prática para utilização em escolas públicas, a qual pode ser utilizada como modelo para aplicações ULA. Na próxima seção, detalha-se a aplicação compatível com ULA, tendo por base um projeto de análise de produções textuais em escolas públicas brasileiras. O fluxo de atividades apresentado detalha cada etapa do desenvolvimento e, nas considerações finais, são levantadas possibilidades de generalização da proposta.

### 3. Projeto para Produção Textual

O projeto para Produção Textual tem por objetivo a análise de redações escritas em língua portuguesa a fim de corrigir os textos e apresentar aos professores um conjunto de análises realizadas nesses textos, considerando critérios específicos. Os textos são produções narrativas feitas por alunos do quinto ao nono ano do ensino fundamental. O fluxo de atividades do projeto, aderente ao conceito de ULA, é apresentado como modelo a seguir.

A análise de produções textuais permite, por exemplo, reduzir o esforço no trabalho dos professores, trazer *feedbacks* instantâneos para os estudantes e diminuir os custos de correção. Neste contexto, as produções textuais precisam ser coletadas, armazenadas e processadas. Além disso, os professores precisam receber a informação analisada de forma sistemática. O processamento das redações, por sua vez, é uma atividade especialmente desafiadora, pois demanda um conhecimento bastante denso de áreas como Visão Computacional e Processamento de Linguagem Natural (PLN). Na Figura 1 é apresentado o fluxo das etapas desenvolvidas neste projeto, incluindo a captura, processamento utilizando IA e devolutiva para o professor. Vale destacar que é possível separar a fase de IA em duas grandes etapas: extração e análise textual. A primeira fase consiste de uma *pipeline* de algoritmos de Visão Computacional e a segunda, de PLN.

#### 3.1. Captura dos dados

Antes de se adentrar às etapas de processamento de imagens e linguagens naturais, tem-se a etapa de coleta das redações. O primeiro passo para esta etapa foi o desenvolvimento de uma folha de resposta especialmente formatada para aumentar a precisão dos algoritmos de visão computacional, de forma que facilitasse a extração textual. Essa folha



Figura 1. Arquitetura ULA (*Unplugged Learning Analytics*) do projeto.

é, em seguida, recolhida pelo professor que, por sua vez, usa um aplicativo móvel para fotografá-la e enviá-la para avaliação. Este aplicativo foi desenvolvido em Andriod e possui como principal característica o fato de utilizar celulares de baixo custo para capturar as imagens. Junto com os dados de imagens, também foram coletadas informações a respeito dos ciclos, escolas, estudantes e professores. O objetivo é oferecer *feedback* da maneira mais personalizada possível.

### 3.2. Aplicação de Inteligência Artificial

A partir das imagens coletadas, deve-se seguir o fluxo de extração do texto, utilizando visão computacional e processamento dos critérios propostos, utilizando PLN.

#### 3.2.1. Visão Computacional

Com as imagens e demais dados armazenados, o professor pode solicitar a correção da prova pelo próprio aplicativo, que realiza uma requisição a uma REST API que é responsável por iniciar a avaliação. Antes que a análise textual aconteça, é necessário extrair-se o conteúdo textual do documento em processamento, denominado Reconhecimento Inteligente de Caracteres (ICR, do inglês *Intelligent Character Recognition*). Para que essa extração ocorra são necessários vários algoritmos de processamento de imagens, conforme descrito a seguir.

Inicialmente, é realizada a binarização da imagem. Esse processo simplifica a

paleta de cores da imagem da folha resposta para apenas preto e branco com o objetivo de aumentar a eficácia de modelos de *machine learning* vindouros. Nessa etapa, entretanto, a binarização é especialmente utilizada para extrair as regiões de interesse (ROI, do inglês *Region of Interest*) da folha: contornos, para diferenciá-la do plano de fundo, a área de texto e a área do *QR-code*, que guarda informações a respeito da escola e da prova.

Com base na identificação de regiões de interesse, segue-se então a segmentação das linhas. O objetivo dessa etapa é identificar em um texto manuscrito não somente o contorno das imagens, mas também, e especialmente, a ordem em que elas aparecem na linha. Por fim, a cada uma das segmentações é aplicado um algoritmo de reconhecimento de texto manuscrito (do inglês, *Handwritten Text Recognition*). Esse tipo de algoritmo mapeia uma imagem a uma sequência de caracteres. Anexo a esse algoritmo realiza-se ainda uma etapa de pós-processamento de correção ortográfica, para aumentar a qualidade da informação extraída. A estrutura de dados, obtida como resultado dessa última etapa, é essencial para o processo como um todo, pois é com base nela que os algoritmos de PLN vão executar a avaliação da redação.

### 3.2.2. Processamento de Linguagem Natural

Após a extração do texto a partir de uma imagem de manuscrito, inicia-se o processo de Processamento de Linguagem Natural para predição de valores de diferentes aspectos avaliativos relacionados às produções textuais. Os aspectos considerados nesta etapa de processamento são: **Registro Formal**; **Coerência Temática**; **Tipologia Textual**; e **Coesão**. Detalhes sobre cada um desses aspectos podem ser encontrados na sec-*resultados*.

Inicialmente, realiza-se uma análise de viabilidade de correção do texto, classificando-o em *viável para correção* ou *inviável para correção*. Produções textuais são consideradas inviáveis quando, por exemplo, não satisfazem um nível ortográfico mínimo, não possuem uma quantidade mínima de linhas, apresentam cópia do texto motivador, quando fogem ao tema ou tipologia, entre outros. Os textos considerados viáveis são analisados de forma independente para cada um dos aspectos avaliativos considerados e, ao final desse processo, a etapa de IA (Visão Computacional e PLN) é concluída e os resultados são passados para a etapa de *devolutiva*.

### 3.3. Devolutiva

Essa é a última etapa da ULA, onde os resultados produzidos pelo módulo de Inteligência Artificial são sumarizados e apresentados através de *feedbacks* imprimíveis para os professores e gestores. Para essa etapa foi executado um ciclo de *design thinking* que resultou numa interface amigável e adequada para ser utilizada como *dashboard* interativo ou extração de pdf imprimível.

## 4. Progresso do Desenvolvimento

Nesta seção, são apresentadas as etapas já implementadas da aplicação proposta, focando nas abordagens de IA utilizadas em cada uma das etapas de Visão Computacional e Processamento de Linguagem Natural, que é o foco deste *workshop*.

#### 4.1. Algoritmos de Visão Computacional

Essa subseção discute os algoritmos responsáveis pela conversão das imagens, obtidas por meio da aplicação, para textos digitais. Na tabela:CV-*alg* estão resumidos os objetivos em cada uma dessas atividades.

**Tabela 1. Etapas do processo de Visão Computacional.**

| Etapa                                | Descrição  |
|--------------------------------------|--|
| Binarização                          | Conversão de imagens multi-tonais, capturadas pelo <i>smartphone</i> , para bi-tonais, facilitando as demais etapas de visão computacional.  |
| Determinação da Área de Interesse    | Identificação de regiões de interesse na folha de produção textual. Algumas dessas regiões são as que contêm a produção propriamente dita e o QR-code.                               |
| Extração da Linha                    | Responsável pelo reconhecimento de linhas e palavras na imagem da folha de redação. O resultado dessa etapa é crucial para conversão dos textos manuscritos para um formato digital. |
| Reconhecimento de Textos Manuscritos | Conversão de textos manuscritos para textos digitais (ASCII, Unicode). Etapa crucial para dar início às etapas de PLN.   |

A primeira etapa é a binarização de imagens. Ela pode ser entendida como a conversão de uma imagem multi-tonal para uma bi-tonal. No contexto deste trabalho, é necessário transformar a imagem digitalizada da folha resposta com vários tons de cores em uma imagem com apenas dois tons de cores: preto e branco. Esse tipo de pré-processamento é essencial para aumentar a precisão de algoritmos de visão computacional. Nesse sentido, foram analisados mais de 60 algoritmos de binarização para escolher o mais apropriado para cada uma das etapas. Com os experimentos realizados, foi possível identificar que o melhor algoritmo para binarização final foi o proposto em [Bataineh et al. 2011].

A segunda etapa é a determinação da área de interesse. Para que se fossem determinadas as áreas de interesse da folha de resposta, foi utilizada uma série de técnicas e métodos de processamento de imagens que usaram como base marcações impressas na própria folha. O resultado gerado pela binarização é passado para o detector de marcações, que as identifica na imagem, empregando métodos de detecção de formas. Para identificar quais das formas encontradas são marcações, o sistema usa a forma e tamanho relativo do *QR-code*, que é, por sua vez, detectado usando o OpenCV <sup>2</sup>.

Ao longo do processo de experimentação foram testados vários algoritmos a fim de se realizar a extração de linhas de forma bem-sucedida. Entre elas, abordagens mais comuns como projeção ortogonal do histograma. Todas elas, entretanto, geravam perdas significativas do texto. A solução para esses problemas, foi, ao invés de focar na linha, focar na palavra. O *framework* introduzido por [He et al. 2016] possibilita que este problema seja resolvido. Mais especificamente, ele obtém a posição da palavra na folha e, em seguida, usando-a como entrada, realiza a inferência da linha. Esse método conse-

<sup>2</sup><https://github.com/opencv/opencv>

que recuperar de 92% a 95% do conteúdo textual da página, o que representa um avanço significativo quando comparado com outros métodos testados.

Por fim, a etapa de extração de linhas fornece as entradas necessárias para se executar, finalmente, a principal etapa do processo: o reconhecimento de textos manuscritos. O seu propósito é ser uma função que converte uma imagem em texto. Foram testados os algoritmos SimpleHTR, Decoupled e HTR-Flor tanto em conjuntos de dados em inglês quanto em português e seus desempenhos foram mensurados usando CER (do inglês, *Character Error Recognition*) e o WER (do inglês, *Word Error Recognition*). É relevante destacar que o conjunto em português foi anotado manualmente por uma equipe de especialistas do time do qual os autores deste trabalho fazem parte. As melhores pontuações foram apresentadas pelo HTR-Flor, que apresentou um aprimoramento em 2% dos resultados com relação aos outros métodos experimentados.

#### 4.2. Algoritmos de Processamento de Linguagem Natural

Nesta subseção são discutidas as abordagens para análise de viabilidade da correção e predição dos níveis dos aspectos avaliativos. Em especial, esta seção discute as soluções desenvolvidas até o momento para lidar com cada uma das etapas de PLN. Na tabela: NLP-*alg*, essas etapas são resumidas. A validação e análise experimental das soluções descritas neste artigo vêm sendo realizadas em conjuntos de dados presentes na literatura, apresentando resultados promissores.

**Tabela 2. Etapas do processo de Processamento de Linguagem Natural.**

| Etapa                  | Descrição  |
|------------------------|--|
| Análise de Viabilidade | Classificação do texto em <i>viável para correção</i> ou <i>inviável para correção</i> de acordo com algumas características. Caso viável, as avaliações de cada um dos aspectos são realizadas de forma independente. |
| Registro Formal        | Classificação em 5 níveis de acordo com: adequação vocabular; concordância verbal e nominal; regência verbal e nominal; pontuação; outras características. Níveis maiores indicam maior domínio do registro formal.    |
| Coerência Temática     | Classificação em 5 níveis de acordo com a adequação do texto aos textos motivadores e tema sugerido. Níveis maiores indicam maior coerência.   |
| Tipologia Textual      | Classificação em 5 níveis de acordo com a adequação do texto à tipologia esperada. Níveis maiores indicam maior adequação.   |
| Coesão                 | Classificação em 5 níveis de acordo com: encadeamento lógico e semântico de períodos; emprego correto de articuladores; emprego correto de conectivos; outras características. Níveis maiores indicam maior coesão.    |

Para a análise de viabilidade de correções das produções textuais, deve-se considerar algumas características: (i) **Nível Ortográfico**: para análise de outras etapas, existem requisitos mínimos de nível ortográfico; (ii) **Fuga ao Tema e Cópia**: outro critério de viabilidade se refere à adequação com o tema e cópia dos textos motivadores. A solução utilizada para classificar com base nesses critérios consiste em analisar a similaridade do texto com o texto motivador e classificá-lo em: cópia (alta similaridade), fuga ao tema (baixa similaridade), ou nenhum dos dois. Em particular, a produção textual é comparada em relação ao texto motivador por meio do algoritmo BERT (*Bidirectional Encoder Representations from Transformers*) [Devlin et al. 2018].

Na etapa de classificação dos níveis de registro formal, múltiplas abordagens são utilizadas para análise de cada uma das características desse aspecto. Com o resultado individual das características, os resultados são agregados para predição do nível final de registro formal. Em especial, as características analisadas são: *pontuação*, por meio de rotulagem de sequência usando o BERT [Devlin et al. 2018]; *concordância*, por meio de um analisador morfológico usando o CoGrOO<sup>3</sup>; e *regência*, por meio de *Named Entity Recognition* (NER) associado a regras gramaticais.

De forma similar à realizada na análise de viabilidade de correção, a coerência temática é classificada em 5 níveis, de acordo com a similaridade da produção de texto aos textos motivadores e o tema sugerido. A similaridade é calculada por meio do algoritmo baseado no BERT [Devlin et al. 2018], e múltiplos *thresholds* são definidos para classificação dos níveis.

Para classificar a adequação do texto à tipologia proposta, faz-se necessário um conjunto de textos anotados manualmente e diferentes métodos capazes de avaliar cada um dos critérios para esse aspecto. Vide a ausência de tais anotações na literatura atual, a solução encontrada é classificar os textos de forma probabilística (probabilidade do texto possuir a tipologia esperada) e converter essa classificação para os 5 níveis de adequação. Em particular, diferentes classificadores (*e.g.*, *Support Vector Machines*, *Random Forest*) vêm sendo aplicados para classificação de textos em tipos textuais através de diferentes características extraídas pelo CohMetrix PT-BR [Camelo et al. 2020].

Por fim, para estimar o nível de coesão de uma produção textual, a solução desenvolvida consiste em extrair múltiplas características relacionadas a coesão (*e.g.*, uso de conectivos, diversidade léxica, legibilidade, sobreposição de frases) e utilizar regressores (*e.g.*, *Extremely Randomized Trees*, *Support Vector Machines*, *Perceptron Multicamadas*) para predição do nível de adequação.

## 5. Considerações Finais

Este artigo possui dois objetivos principais: apresentação do conceito de *Learning Analytics* Desconectada e a aplicação prática deste conceito em um projeto que está sendo desenvolvido no contexto das escolas públicas brasileiras. Apesar do projeto ainda estar em fase inicial do desenvolvimento, já foram coletadas em torno de 900 mil imagens de produções textuais de alunos, que estão sendo usadas para o desenvolvimento e validação dos modelos de inteligência artificial que serão utilizados.

Como implicações práticas da proposta, acreditamos que o conceito de ULA está alinhado com o objetivo de desenvolvimento sustentável da ONU relacionado à promoção de educação de qualidade para todos<sup>4</sup>, pois ele possibilita a utilização de inteligência artificial em instituições que possuem poucos recursos disponíveis. É relevante destacar que, além do projeto de produção textual, em setembro será iniciado outro projeto relacionado à análise de equações matemáticas, que utilizará a mesma arquitetura proposta neste artigo.

---

<sup>3</sup><http://cogroo.org/>

<sup>4</sup><https://sdgs.un.org/goals/goal4>



## Referências

- Bataineh, B., Abdullah, S. N. H. S., and Omar, K. (2011). An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. *Pattern Recognition Letters*, 32(14):1805–1813.
- Camelo, R., Justino, S., and de Mello, R. F. L. (2020). Coh-metrix pt-br: Uma api web de análise textual para a educação. In *Anais dos Workshops do IX Congresso Brasileiro de Informática na Educação*, pages 179–186. SBC.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.
- Freitas, E., Falcão, T. P., and Mello, R. F. (2020). Desmistificando a adoção de learning analytics: um guia conciso sobre ferramentas e instrumentos. *Sociedade Brasileira de Computação*.
- Gaftandzhieva, S., Docheva, M., and Doneva, R. (2020). A comprehensive approach to learning analytics in Bulgarian school education. *Education and Information Technologies*.
- Gašević, D. (2018). Include us all! directions for adoption of learning analytics in the global south. *Learning analytics for the global south*, pages 1–22.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Jivet, I., Scheffel, M., Specht, M., and Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 31–40.
- Lim, C. P., Ra, S., Chin, B., and Wang, T. (2020). Information and communication technologies (ict) for access to quality education in the global south: A case study of sri lanka. *Education and Information Technologies*, 25(4):2447–2462.
- Matias, T. P., Imperador, A. M., Botezelli, L., Riondet-Costa, D. R. T., de Oliveira Costa, V. A., and de Alcântara Moreira, L. (2022). Vulnerable populations in south america and facing the pandemic: what can our experiences reveal to the world? *Research, Society and Development*, 11(4):e59511427924–e59511427924.
- Ossai, A. (2022). Sustainable development goal four (sdg4): Challenges and the way forward. *International Journal of Advance Research in Education & Literature*, 8(2).
- Patel, N., Thakkar, M., Rabadiya, B., Patel, D., Malvi, S., Sharma, A., and Lomas, D. (2022). Equitable access to intelligent tutoring systems through paper-digital integration. In Crossley, S. and Popescu, E., editors, *Intelligent Tutoring Systems*, pages 255–263, Cham. Springer International Publishing.
- Perine, C. M. and Rowsell, J. (2019). Making it work in the global south: Stories of digital divides in a brazilian context. In *Stories from inequity to justice in literacy education*, pages 185–202. Routledge.
- Puigjaner, R. (2016). Progressing toward digital equity. In *IFIP World Information Technology Forum*, pages 109–120. Springer.

- Rambe, P. and Moeti, M. (2017). Disrupting and democratising higher education provision or entrenching academic elitism: towards a model of moocs adoption at african universities. *Educational Technology Research and Development*, 65(3):631–651.
- Renz, A. and Hilbig, R. (2020). Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies. *International Journal of Educational Technology in Higher Education*, 17(1):1–21.
- Rodríguez, M. R. B. and Cobo, C. (2022). Covid-19 and education in the global south: Emergency remote learning solutions with long-term implications.
- Sharma, A. and Jayagopi, D. B. (2018). Automated grading of handwritten essays. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 279–284. IEEE.
- Sousa, E. B. d., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., and Gašević, D. (2021). Applications of learning analytics in high schools: a systematic literature review. *Frontiers in Artificial Intelligence*, 4:737891.
- Stevanim, L. F. et al. (2020). Exclusão nada remota: desigualdades sociais e digitais dificultam a garantia do direito à educação na pandemia.
- Thompson, G. (2020). Two thirds of the world’s school-age children have no internet access at home, new unicef-itu report says. <https://www.unicef.org/press-releases/two-thirds-worlds-school-age-children-have-no-internet-access-home-new-unicef-itu>.
- Tlili, A., Burgos, D., Altınay, F., Altınay, Z., Huang, R., and Jemni, M. (2021). Remote special education during covid-19: A combined bibliometric, content and thematic analysis. In *2021 international conference on advanced learning technologies (ICALT)*, pages 325–329. IEEE.
- Varanasi, M. R., Fischetti, J. C., and Smith, M. W. (2018). Analytics framework for k-12 school systems. In *Data Leadership for K-12 Schools in a Time of Accountability*, pages 206–233. IGI Global.
- Zanibbi, R. and Blostein, D. (2012). Recognition and retrieval of mathematical expressions. *International Journal on Document Analysis and Recognition (IJ DAR)*, 15(4):331–357.