Vol.7 No.1 (2024)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Thematic exploration of educational research after the COVID pandemic through topic modelling

Indu Bala[A]

Lewis Mitchell[B]

[A] School of Computer and Mathematical Sciences, The University of Adelaide, Adelaide, Australia, 5005

[B] School of Computer and Mathematical Sciences, The University of Adelaide, Adelaide, Australia, 5005

Keywords

Academic journal trends;
academic teaching;
BERTopic model;
educational research;
higher education;
learning analytics.

Correspondence

indu.bala@adelaide.edu.au [A]

Article Info

Received 20 February 2024
Received in revised form 21 March 2024
Accepted 22 April 2024
Available online 23 April 2024

DOI: https://doi.org/10.37074/jalt.2024.7.1.40

Abstract

The Journal of Applied Learning and Teaching is known for its focus on innovative practices in learning and teaching in higher education. In this study, we utilised BERTopic modelling to investigate trends and research within the journal. Our objective was to analyse thematic structures and identify emerging trends in a vast academic research corpus. BERTopic modelling enabled us to categorise academic texts into distinct topics, revealing underlying patterns and themes. Our analysis unveiled various topics, showcasing the journal's interdisciplinary nature. Particularly, articles from January 2021 to December 2023 shed light on global trends in learning and teaching amidst significant changes in the post-COVID era. We identified 17 frequent topics, categorised into four major thematic groups: Technology and Digital Learning in Education, Healthcare and Clinical Training, Educational Strategies and Outcomes, and Pandemic-Driven Social and Compassion Aspects in Education. We examined these themes and presented the findings, highlighting challenges and opportunities in higher education. This comprehensive analysis serves as a roadmap for future research, guiding scholars and practitioners in advancing applied learning and teaching.

Introduction

The education landscape is continually evolving, driven by technological advancements and global challenges. Innovative teaching methods and learning paradigms are necessary to maintain a positive learning environment (Escueta et al., 2017). Global challenges, such as COVID-19 and wars, can sometimes disrupt or alter the learning environment. Therefore, it is crucial to periodically evaluate the existing methods in light of current situations and prepare for the future. In this context, understanding the thematic progression in higher educational research is vital (Yunita, 2018).

In recent years, there has been a surge in research aiming to understand the thematic progression of educational landscapes, resulting in a significant increase in the application of artificial intelligence in academic research (Ai, 2017). This has introduced new methods for analysing scholarly literature. Additionally, there has been a growing demand for interdisciplinary approaches among scientific disciplines, leading to the development of numerous interdisciplinary programmes to address academic challenges. The outcomes of these programmes are annually published in various journals (Huston et al., 2018). However, it is essential to periodically analyse these research findings to comprehend current challenges in academia and gain insights for the future. Analysing the vast amount of published research is daunting due to the rapid advancement of science and technology. This underscores the importance of monitoring research trends to identify potential innovations. Research trends can be identified using various sources, including scientific literature, books, articles, and publications extensively reviewed by researchers worldwide (Ranjbar-Sahraei & Negenborn, 2017; Jiang et al., 2018). The analysis of these publications has proven helpful in identifying emerging topics and tracking their evolution over time. Topic modelling techniques offer one way to analyse these text-based published research articles, helping to uncover patterns and address more specific research questions (Amado et al., 2018).

Topic modelling has been effectively used to discern patterns and topics from scholarly publications. Researchers have previously endeavoured to determine trends through publications to analyse current and future directions. For example, Chen et al. (2020) utilised structural topic modelling to analyse articles published in the Computers & Education Journal, identifying research hotspots. They also analysed annual topic proportion trends and topic correlations, offering insights into potential future research directions within the journal's scope. Similarly, Pandur et al. (2020) employed a combined approach of Structural Topic Modelling and Latent Dirichlet Allocation (LDA) to extract topics and identify current trends from scientific papers in the field of social science from the Web of Science. Later, Lemay et al. (2021) conducted a structural topic modelling of articles in educational data mining and learning analytics, revealing thematic features of these two fields. They identified five significant topics within educational data mining and learning analytics and analysed the differences in research focus between the two disciplines. Furthermore, Nylander et al. (2022) applied topic modelling to publications in the International Journal of Lifelong Education, identifying predominant themes and examining the evolution of the journal's content over time. Hussain et al. (2022) crafted a multi-layered topic modelling approach that integrates situation awareness with an advanced hybrid machine learning technique to analyse students' textual feedback in academic environments. Maphosa and Maphosa (2023) employed the LDA for a bibliometric analysis on a subset of the Scopus database, explicitly examining the progression of Artificial Intelligence (AI) research in higher education (HE). Choi and Lee (2023) constructed a topic map covering areas such as biocompatible materials, structural materials, electrochemistry, and photonics, using it to discern national research priorities in materials science and to explore the competitive stances and strategic approaches of leading countries. Additionally, in recognition of COVID-19's extensive impact across research fields, Cao et al. (2023) utilised topic modelling on published abstracts to evaluate the pandemic's effect on research directions. By applying the LDA method, they delineated the research topics, trends, and topic correlations in COVID-19 studies, finding that the thematic similarity between topics increased with the scope of documents analysed.

Focusing on trends in higher education post-COVID-19 pandemic, this study aims to analyse a collection of journal abstracts from the Journal of Applied Learning and Teaching (JALT). JALT serves as a crucial hub, offering a global platform for new ideas and insights in the realm of higher education. Employing the sophisticated BERTopic modelling technique, this research identifies the most prominent themes and patterns that have emerged in recent years. This illuminates the evolving nature of academic research, especially in education and teaching, after the COVID-19 pandemic. Our study reveals detailed trends and shifts in higher education methodologies. The objectives of this study are:

- To conduct a BERTopic modelling-based analysis of the Journal of Applied Learning and Teaching, pinpointing the major research topics that have arisen following the COVID-19 pandemic.
- To provide a broader range of thematic research areas and analyse their annual distribution, extracting significant insights.

Methodology

Text mining techniques are designed to unearth valuable knowledge that may be hidden or not immediately apparent within a vast amount of textual data. Standard text mining methods include unsupervised and supervised techniques such as text categorisation, text clustering, document summarisation, and keyword extraction (Gurcan & Cagiltay, 2023; Bala, 2023). Topic modelling, a subset of text mining, is an unsupervised machine-learning technique that identifies topics within a collection of documents. For our analysis, we utilised the BERTopic model, an advanced method that typically does not require extensive data preprocessing. However, given the complexity of our dataset and our goal to extract meaningful topics accurately, we performed a thorough data preprocessing, which we detail in the subsequent sections.

Data source and preprocessing

We extracted data from published research articles on the Journal of Applied Teaching and Learning (JALT) website from January 2021 to December 2023. The dataset comprises the title, research type, publication date, DOI link, abstract, and keywords for 144 articles. To pre-process the text from the abstract collection, we implemented the following steps:

- Lowercasing: All text data is converted to lowercase to standardise the text, ensuring that words in different cases (e.g., "Abstract", "abstract", "ABSTRACT") are treated as the same (Alasadi & Bhaya, 2017).

- Removing punctuation and numbers: Using regular expressions, the text is stripped of punctuation and numbers, identifying and removing non-word characters and numerical digits. This refinement focuses the analysis on the lexical content, removing extraneous elements that could skew the textual analysis.

- Tokenising the text: The cleaned text is tokenised and split into individual words or tokens using the word tokenise function from the Natural Language Toolkit (NLTK). This step breaks the text into words, provides a list of tokens, and sets the stage for further processing, such as stemming or lemmatisation (Hardeniya et al., 2016).

- Removing stopwords: Common words that typically carry little meaning (stopwords), such as 'the', 'is', 'in', etc., are removed using the NLTK-provided list of English stopwords. Filtering out these words eliminates those not likely to be significant for the analysis.

- Re-constructing the text: The remaining words (tokens) are then joined to form the final pre-processed text, which contains only the relevant and meaningful words ready for analysis.

Topic modelling

BERTopic is a state-of-the-art topic modelling method in Natural Language Processing (NLP), employing transformer embeddings and clustering algorithms to extract meaningful topics from text (Grootendorst, 2022). The BERTopic process involves several key steps:

- Embedding of documents: Documents are converted into numerical formats using the sentence-transformer model "Paraphrase-multilingual-MiniLM-L12-v2". This pre-trained model effectively generates sentence embeddings and facilitates semantic search (Reimers & Gurevych, 2019).

- Reducing of dimensionality: BERTopic reduces the dimensionality of these embeddings using the Uniform Manifold Approximation and Projection (UMAP) technique, maintaining the integrity of local and global structures in reduced dimensions (McInnes et al., 2018).

- Clustering process: The core of topic extraction involves clustering the reduced embeddings into related groups using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which excels at identifying clusters of varying densities and is resilient to noise (McInnes et al., 2017).

- Topic tokenisation: CountVectoriser, which converts text data into a numerical format, is used for tokenising topics, facilitating efficient analysis and interpretation of the results (Hu & Zhang, 2022).

- Weighting and Topic Differentiation: A class-based TF-IDF (c-TF-IDF) approach, which focuses on clusters rather than individual documents, is used to differentiate clusters. This method identifies unique aspects of documents within each cluster, treating each cluster as a single document and analysing word frequencies. This approach precisely defines clusters and suggests topic names (Grootendorst, 2022; Bala et al., 2023). Figure 1 visualises the working process of the BERTopic model.

BERTopic provides a flexible modelling approach that does not require the pre-definition of the number of topics. However, setting a potential topic range can help guide the model to capture broader themes (Wang et al., 2023; Bala et al., 2024). After training the model, it is crucial to evaluate the topics to ensure they are coherent, interpretable, and accurately reflect the main themes of the text data. This evaluation often involves assessing the coherence and distinctiveness of the topics. To further refine the results, topics can be merged based on similarity. This process entails setting a similarity threshold score; topics with a score above this threshold are grouped. This reduces the granularity of the topics and allows for broader thematic categories, making them easier to analyse and interpret.

In this study, we enhanced our analysis by setting a high similarity threshold of 0.83. Applying this to the 17 topics obtained from the BERTopic model, we successfully grouped them into four distinct thematic categories. This consolidation provided a clearer, more thematic organisation of the topics, aiding in a more straightforward understanding and interpretation of the subject matter in the corpus.

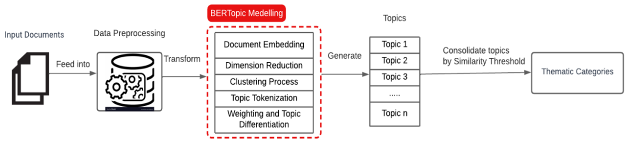

Figure 1. Graphical representation of BERTopic model.

Results

Data analysis

For the experimental evaluation, the data were analysed, and a notable observation was that a majority of the articles incorporated references to COVID, varying from minor mentions to significant discussions. Figure 2(a) displays the year-wise number of publications in the visual representation, indicating a consistent annual increase in published articles. Complementing this, Figure 2(b) offers a histogram depicting the word count distribution across the abstracts, highlighted by a distribution curve. The most frequent abstract length hovers around the 200-word mark, a crucial detail for text analysis considerations.

We pre-processed the dataset using Python packages before conducting the topic modelling analysis on abstracts. This preprocessing resulted in a total word count of 31,887 in the abstracts, which was the foundation for our topic modelling exercise. This preparatory step ensured that the subsequent analysis was conducted on a refined dataset, poised to yield more accurate and meaningful insights into the prevalent topics of academic discourse in higher education.

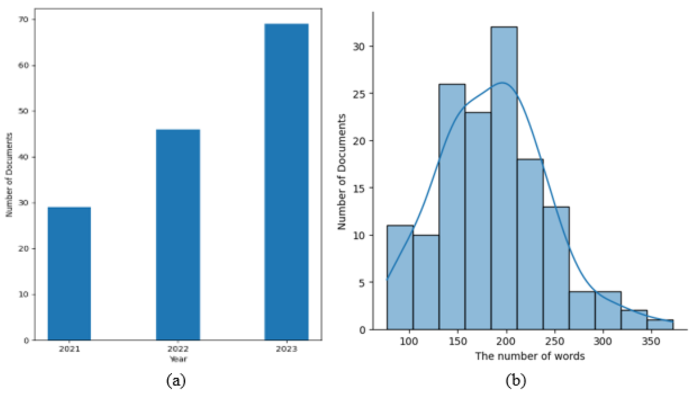

Figure 2. (a) Number of publications per year, (b) Word distributions of documents.

Topic modelling results analysis

In this study, we applied the BERTopic model to a collection of abstracts, identifying 17 distinct topics. These were visually presented in Figure 3 through a bar graph that illustrates the distribution of topics across the dataset. Each bar's height indicates the prevalence or score of a topic, accompanied by the top 15 high-scoring words associated with these topics. This facilitated the assignment of descriptive labels to each topic, derived from the most significant and recurring terms that reflect their central themes. The highest-scoring terms within each topic were pivotal in the labelling process, as they most represent the topic's content.

Delving into the specifics, Topic 0, 'AI in Education,' features terms like 'AI,' 'ChatGPT,' and 'education,' suggesting a focus on AI tools such as chatbots in educational settings. Meanwhile, Topic 1, 'COVID-19 and Education,' includes words like 'COVID,' 'pandemic,' and 'online,' indicating discussions about the pandemic's impact on educational practices, particularly the shift to online learning. Topic 2, 'Online Learning Environments,' characterised by terms such as 'online,' 'learning,' and 'social,' points to the evolving nature of digital learning spaces and tools.

Further topics range from 'Educational Feedback and Outcomes,' examining the efficacy of feedback in learning, to 'Mathematics Education in Sub-Saharan Africa,' highlighting unique regional challenges and teaching methodologies. 'Digital Learning Platforms and Practices' underscores the growing trend in digital education, while 'Integrated Programming and Engineering Equation' and 'Teacher Professional Development and Learning' highlight the expanding scope of interdisciplinary educational approaches and continuous educator development.

In healthcare education, 'Clinical Skills and Healthcare Training' emphasises the importance of practical training. At the same time, 'Academic Connectedness and Resilience' highlights the significance of mental resilience and social bonds in academic settings. The theme of collaboration is further explored in 'Collaborative Online Teaching and Learning,' showcasing the value of community and interaction in online education. 'Graphic Design and Modern Pedagogical Approaches in Higher Education' contrasts this by delving into the role of design in teaching methods. At the same time, 'UTAUT Model and Technology Acceptance' examines the theoretical aspects of technology acceptance in education. The final themes, 'Knowledge Management and Learning Programs' and 'Reading Processes and Political Contexts in Education,' explore the strategic aspects of knowledge management in educational environments and the interplay between political narratives and educational content, respectively.

These topics are labelled according to the most frequent terms mentioned in Table 1, with the labels derived from a combination of the most weighted terms within each topic. These terms serve as indicators of the central themes being addressed in the collection of texts, thus guiding readers to grasp the core ideas quickly. Upon careful analysis, we observed that some topics bear similarities to others. We employed an intertopic distance map to delve deeper, as shown in Figure 4(a). This map, generated using dimensionality reduction techniques such as UMAP, visualises the distances between different topics identified in the dataset. It represents high-dimensional data in two or three dimensions, aiming to preserve relative distances as accurately as possible. This visualisation aids in distinguishing between topics and demonstrates the distinctness of each topic from the others. Well-separated clusters in the map suggest distinct topics, while areas where clusters overlap indicate similarity. Additionally, the map reveals relationships between topics, allowing us to infer potential connections or common themes. Topics in proximity may share specific keywords or relate to similar subject matter. This is a diagnostic tool to evaluate the quality of topic modelling, where a good topic model is characterised by well-defined, non-overlapping clusters corresponding to coherent and distinct topics.

Figure 4(a) indicates some topics' similarities to others. We employed additional visualisations to validate these similarities further and understand the generated topics' pattern. We visualised this using a hierarchical clustering map in Figure 4(b) and a similarity matrix heatmap in Figure 5. The hierarchical clustering map, or dendrogram, visually groups similar topics together in a tree-like structure. Each branch in this structure represents a possible grouping, with branch lengths indicating the level of similarity between topics. Similar topics cluster under the same branch and are positioned closer, aiding in determining a cutoff level for merging related topics and thereby simplifying the topic model. In contrast, the similarity matrix heatmap visualises pairwise topic similarities in a matrix format, where warmer colours represent higher similarity and cooler colours have lower similarity. Each matrix cell shows the similarity score between two topics, with the diagonal indicating maximum similarity, as a topic is always 100% similar to itself. This heatmap is instrumental in quickly identifying topics that are similar or distinct. It also helps verify the coherence of the topics generated by the model, where coherent topics should exhibit higher similarity scores with themselves and lower scores with unrelated topics.

Therefore, based on these visualisations, we opted for a similarity threshold value of 0.83 to consolidate the topics. This thresholding led to four thematic categories, each characterised by a thematic similarity of 0.83 or higher. The details of these thematic groups are discussed in the following section, highlighting their significance and interrelations.

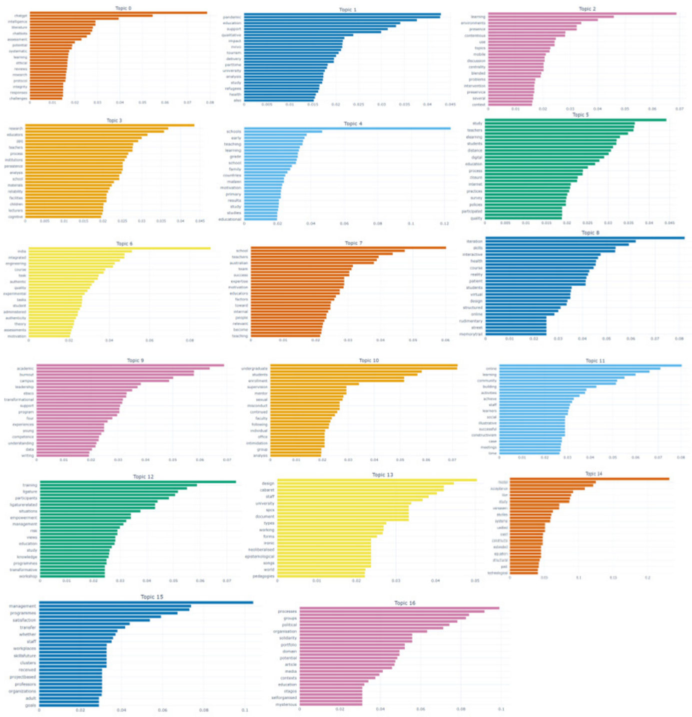

Figure 3. Topic word scores.

Table 1: Topics and their labels based on the most frequent terms.

Generated Topics	Created Labels	Most frequent terms
Topic 0	AI in Education	AI, ChatGPT, education, intelligence, artificial, literature, higher, chatbot, content, assessment, paper, potential, academic, systematic, generative, learning, human, ethical, benefits, reviews, using, research, tools, protocol, teachers, integrity, assessments, responses, impact, challenges
Topic 1	COVID-19 and Education	COVID, pandemic, online, education, students, support, higher, qualitative, research, impact, provided, NVivo, student, tourism, hybrid, delivery, wellbeing, part-time, model, university, future, analysis, impacted, study, graduate, refugees, data, health, identified, compassion
Topic 2	Online Learning Environments	online, learning, social, environments, simulation, presence, students, contentious, classes, use, interconnectivity, topics, using, mobile, distance, discussion, challenged, centrality, network, blended, levels, problems, various, intervention, books, preservice, include, several, multiple, context
Topic 3	Educational Feedback and Outcomes	feedback, research, instructional, educators, outcomes, students, teachers, academics, process, books, institutions, psychomotor, persistence, gender, analysis, factor, school, affective, materials, study, reliability, publishing, facilities, questionnaire, children, picture, lecturers, researchers, cognitive
Topic 4	Mathematics Education in Sub-Saharan Africa	mathematics, schools, Subsaharan, early, learners, teaching, ppd, Africa, learning, children, grade, teachers, school, family, quality, countries, Tanzania, Malawi, interventions, motivation, challenge, primary, curriculum, results, low, study, performance, studies, approaches, educational
Topic 5	Digital Learning Platforms and Practices	teaching, study, countries, teachers, school, e-learning, platforms, students, learners, distance, pd, digital, refugee, education, refugees, process, closures, closure, impact, internet, time, practices, used, survey, six, policies, access, participated, gap, quality
Topic 6	Integrated Programming and Engineering Equation	writing, India, students, integrated, programming, engineering, approach, course, thinking, task, skills, authentic, accounting, quality, engagement, experimental, developing, tasks, education, student, study, administered, findings, authenticity, problem-solving, theory, traditional, assessments, classroom, motivation
Topic 7	Teacher Professional Development and Learning	learning, school, routines, teachers, us, Australian, students, team, professional, success, motivating, expertise, teacher-led, motivation, experiences, educators, peers, factors, academic, toward, achievement, internal, teacher, people, development, relevant, study, become, knowledge, teaching
Topic 8	Clinical Shills and Healthcare Training	clinical, iteration, assessment, skills, game, interactive, nursing, health, application, course, user, reality, mixed, patient, experience, students, OSCEs, virtual, Singapore, design, serious, structured, resulted, online, realism, rudimentary, self-confidence, street, persons, memory trail
Topic 9	Academic Connectedness and Resilience	connectedness, academic, level, burnout, resilience, campus, PhD, leadership, students, EBSCO, findings, transformational, study, support, student, programme, graduate, empathy, bass, experiences, self-efficacy, young, individuals, competence, provide, understanding, kindness, counsel, within, humaneness
Topic 10	Research in Undergraduate Education	research, undergraduate, hdr, students, concurrent, enrolment, campus, supervision, matriculation, mentor, researchers, sexual, stem, misconduct, respondents, continued, university, faculty, opportunity, following, training, individual, member, office, eraus, intimidation, Embry-Riddle, group, factors, analysis
Topic 11	Collaborative Online Teaching and Learning	teaching, online, collaborative, learning, collaboration, community, Jamboard, building, applied, activities, model, achieve, strong, staff, presence, learners, providing, social, deep, illustrative, scaffolded, successful, collaborativism, constructivism, due, case, support, meetings, contemporary, time
Topic 12	Healthcare Simulation and Training	healthcare, training, simulation-based, ligature, within, participants, health, ligature related, professionals, situations, ipe, empowerment, theme, management, care, risk, demonstrated, views, effectively, education, nursing, study, experiences, knowledge, themes, programmes, firsthand, transformative, transformed, workshop
Topic 13	Graphic Design and Modern Pedagogical Approaches in Higher Education	graphic, design, sbp, cabaret, modern, staff, pedagogic, university, kabarett, specs, tlrs, document, ite, types, pedagogy, working, article, forms, critique, ironic, tlr, neoliberalised, sociality, epistemological, regimes, songs, charged, world, performative, pedagogies
Topic 14	UTAUT Model and Technology Acceptance	utaut, model, use, acceptance, technology, like, educational, study, new, osmasem, theory, studies, empirical, systems, predicting, unified, intentional, used, initial, constructs, robust, extended, proven, equation, done, structural, framework, past, modelling, technological
Topic 15	Knowledge Management and Learning Programs	knowledge, management, motivation, programmes, project, satisfaction, design, transfer, learn, whether, learning, staff, academic, workplaces, Paulo, skillsfuture, elucidate, clusters, são, received, real life, project-based, intent, professors, distinct, organisations, three, adult, orientation, goals
Topic 16	Reading Processes and Political Contexts in Education	reading, processes, social, groups, neoliberalist, political, organisation, movements, solidarity, Aotearoa New Zealand, portfolio, critiques, domain, reforms, potential, particular, article, tertiary, media, forms, contexts, examine, education, tap, otagos, heart, self organised, suggestion, mysterious

Figure 4. (a) Intertopic distance map (b) Hierarchical Clustering.

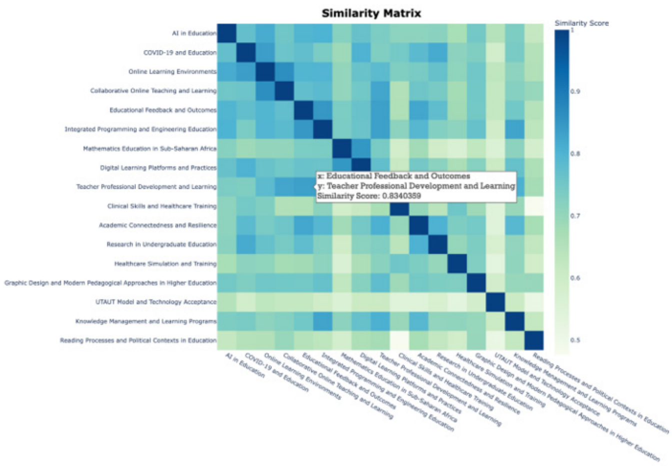

Figure 5. The heatmap presented a similarity matrix between the topics.

Thematic groups

We employed a clustering method focused on thematic resemblance to form coherent clusters from the identified topics. This involved an in-depth examination of each topic, merging those with intersecting or related themes, following the methodology suggested by Glazkova (2021). We used a benchmark similarity score of 0.83 for grouping, ensuring high relevance and cohesion within each cluster. Subsequently, descriptive labels were carefully selected for each cluster, aiming to capture the essence of the combined topics succinctly.

The first thematic group, “Technology and Digital Learning in Education,” unites topics like “AI in Education,” “Online Learning Environments,” “Digital Learning Platforms and Practices,” and “UTAUT Model and Technology Acceptance.” This cluster reflects the increasing integration of technology in educational settings. The second group, titled “Healthcare and Clinical Training,” encompasses “Clinical Skills and Healthcare Training” along with “Healthcare Simulation and Training.” This cluster highlights the importance of hands-on training and simulation in healthcare education. The third group, “Educational Strategies and Outcomes,” includes a diverse array of topics such as “Educational Feedback and Outcomes,” “Teacher Professional Development and Learning,” “Mathematics Education in Sub-Saharan Africa,” “Collaborative Online Teaching and Learning,” and “Graphic Design and Modern Pedagogical Approaches in Higher Education,” reflecting various educational methods and outcomes. Lastly, the “Pandemic-Driven Social and Compassionate Aspects in Education” group brings together themes like “COVID-19 and Education,” “Academic Connectedness and Resilience,” “Research in Undergraduate Education,” “Integrated Programming and Engineering Education,” “Knowledge Management and Learning Programs,” and “Reading Processes and Political Contexts in Education,” focusing on the educational impact of the pandemic.

These clusters categorise a broad range of topics into more concise and focused groups, each distinguished by its unique thematic focus. This clustering simplifies understanding of diverse topics and highlights distinct areas within the broader educational context.

Discussion

Using topic modelling, we closely examined 17 topics that detail the trends of publications in the Journal of Applied Learning and Teaching. Subsequently, we extracted four thematic groups to explore the broader interests of researchers, particularly addressing various challenges in learning and teaching.

The first thematic group focuses on Technology and Digital Learning in Education, highlighting the growing significance of AI and technological innovations in education. As depicted in Figure 6, terms like ‘AI’, ‘ChatGPT’, and ‘artificial’ significantly emphasise the role of AI, especially the use of AI-driven tools such as chatbots in educational contexts. The frequent appearance of words like ‘online’, ‘e-learning’, and ‘blended’ indicates a shift towards digital and blended learning environments. This group also explores the ethical aspects of digital learning, including concerns about academic honesty and maintaining integrity in a digital academic landscape. Key terms such as ‘assessment’, ‘feedback’, and ‘engagement’ reflect a strong interest in evaluating the impact of these technologies, particularly regarding educational outcomes and student involvement. Figure 7 further illustrates a consistent increase in the importance of technological and digital tools in educational settings, aligning with the study (Rudolph et al., 2023). The group envisions a future where digital platforms are central to content delivery and student engagement. However, it also points to the unethical use of tools like ChatGPT in assessments, which breaches ethical standards. Solutions include constructive feedback, rigorous supervision, interactive methods, and teamwork activities to ensure adherence to ethical norms in the evolving educational landscape.

The second thematic group, ‘Healthcare and Clinical Training,’ emphasises the crucial role of practical, experiential learning in healthcare education. As shown in Figure 6, there is a distinct focus on healthcare education, particularly in nursing and clinical disciplines. Terms like ‘clinical’, ‘nursing’, ‘skills’, and ‘assessment’ are frequently mentioned, underscoring the importance of hands-on skill development and evaluation in this sector. Notably, using terms such as ‘virtual’, ‘interactive’, ‘reality’, and ‘game’ indicates the adoption of advanced, technology-driven training methods such as virtual reality and interactive simulations. This innovative approach aims to enhance learning and increase healthcare practitioners' confidence. Furthermore, the presence of terms like 'gender', 'equality', and 'female' in this thematic group suggests attention to gender perspectives in healthcare training, potentially addressing issues of diversity and inclusiveness. Essentially, this group reflects a trend towards more engaging and immersive learning experiences through simulations and virtual reality, which is particularly vital in the dynamic and constantly evolving healthcare sector. The need for continuous training and skill enhancement became especially prominent during the COVID-19 pandemic. As depicted in Figure 7, the peak interest in this area occurred around 2021, coinciding with the heightened impact of the COVID-19 pandemic, indicating an intensified focus on healthcare training during this period.

The third thematic group, 'Educational Strategies and Outcomes,' delves into various educational tactics and their impacts. The terminology depicted in Figure 6 paints a broad picture of education and learning. Terms such as 'feedback', 'instructional', 'outcomes', and 'pedagogy' highlight a focus on teaching methods, student evaluation, and education results. Regular references to 'students', 'teachers', 'educators', and 'learners' emphasise the central figures in the educational process. Furthermore, words like 'research', 'study', 'analysis', and 'publishing' suggest a scholarly approach to education, potentially encompassing academic research and the field of educational publishing. The mention of topics like 'mathematics', 'cognitive', 'affective', and 'psychomotor' indicates an exploration of various learning domains and subject areas, showing a wide array of educational interests and approaches. This theme's distribution, as seen in Figure 7, indicates steady progress and possibly modest enhancements in educational strategies and outcomes. Overall, this group underscores the critical role of teacher development in elevating educational quality. The specific focus on mathematics education in Sub-Saharan Africa signals attention to region-specific educational challenges and tailored solutions, including innovative pedagogical methods and graphic design, pointing to a shift towards more dynamic and engaging teaching techniques.

Finally, the fourth thematic group, 'Pandemic-Driven Social and Compassion Aspects in Education,' sheds light on the societal and emotional dimensions influencing education during the pandemic (Figure 6). This group, marked by terms like 'COVID', 'pandemic', 'online', 'education', 'wellbeing', and 'hybrid', captures the diverse ways the COVID-19 pandemic reshaped educational practices. It includes the shift to online and hybrid learning models, emphasising the necessary adjustments in educational approaches during this period. The focus on student welfare, mental health, and resilience reflects students' psychological challenges. Additionally, this group highlights the pandemic's significant impact on higher education, with terms like 'university', 'graduate', and 'PhD' indicating changes in teaching, learning, and research dynamics. Including research-oriented terms implies investigating the pandemic's effects through qualitative studies and data analysis. Moreover, this group underscores the importance of maintaining community bonds and academic connectedness despite physical distancing. Keywords like 'skillsfuture' and 'mentoring' emphasise skill enhancement and career preparation, adapting to changing job markets. The use of terms like 'connectedness', 'compassion', 'empathy', 'community', 'motivation', and 'kindness' highlights the importance of social and emotional support during these challenging times. As shown in Figure 7, interest in this thematic area showed a sharp increase up to 2022, followed by a decline, suggesting a peak in concern for the pandemic's social and compassionate aspects in education, which appears to have decreased by 2023. Overall, this thematic category offers a holistic view of the educational sector's adaptation to the pandemic, emphasising digital transformation, emotional support mechanisms, and preparation for future challenges.

Each thematic group provides a lens through which to understand current trends, challenges, and future directions in education and training. They collectively highlight the dynamic nature of education, the growing influence of technology, and the importance of contextual and culturally sensitive approaches.

Figure 6. Word clouds of thematic groups.

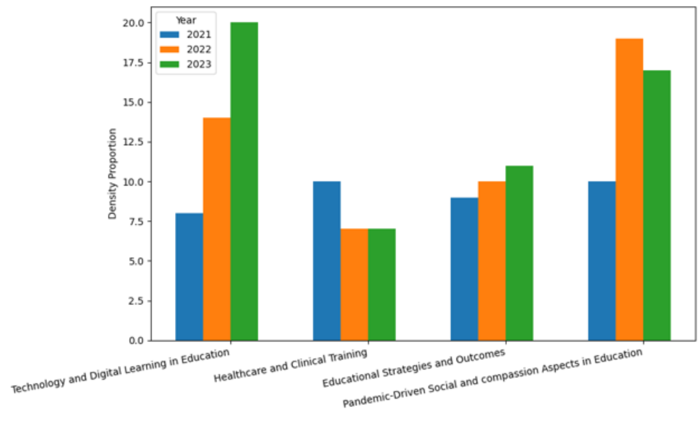

Figure 7. Year-wise densities distribution of thematic groups.

Conclusion

In conclusion, applying BERTopic modelling to the Journal of Applied Learning and Teaching has provided critical insights into the evolving landscape of higher education in the post-COVID era. Through a detailed analysis of academic texts from January 2021 to December 2023, we identified key research themes and trends within higher education. We meticulously organised 17 predominant topics into four broad thematic groups: Technology and Digital Learning in Education, Healthcare and Clinical Training, Educational

Strategies and Outcomes, and Pandemic-Driven Social and Compassionate Aspects in Education. These groups highlight the journal's interdisciplinary approach and offer a comprehensive understanding of the shifts in educational practices and research focus areas. Our annual examination of these themes, illustrated through various data visualisations, underscores the dynamic challenges and opportunities identified by researchers. The study emphasises the growing importance of technology and digital learning, the intricacies of healthcare training, the efficacy of educational strategies, and the social and emotional impacts of the pandemic, all of which are pivotal in shaping the future trajectory of higher education. This rigorous analysis directs attention to crucial areas for further research and development, which hold the potential to refine applied learning and teaching methodologies significantly. As the educational sector continues to evolve in the aftermath and beyond the pandemic, this study is a strategic foundation for educators and policymakers to enhance resilience and adaptability. It provides invaluable guidance for addressing the ongoing global transformations in education, ensuring that the sector remains responsive and forward-thinking in these changing times.

References

Ai, B. (2017). Constructing an academic identity in Australia: An autoethnographic narrative. *Higher Education Research & Development, 36*(6), 1095-1107. https://doi.org/10.1080/07294360.2017.1303459

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences, 12*(16), 4102-4107. http://dx.doi.org/10.3923/jeasci.2017.4102.4107

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on big data in marketing: A text mining and topic modeling-based literature analysis. *European Research on Management and Business Economics, 24*(1), 1-7. https://doi.org/10.1016/j.iedeen.2017.06.002

Bala, I. (2023). Natural language processing in medical science and healthcare. *Medicon Medical Sciences, 4*, 01-02. https://doi.org/10.55162/MCMS.04.088

Bala, I., Kelly, T.-L., Lim, R., Gillam, M. H., & Mitchell, L. (2023). An effective approach for multiclass classification of adverse events using machine learning. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE32021924

Bala, I., Kelly, T.-L., Stanford, T., Gillam, M. H., & Mitchell, L. (2024). Machine learning-based analysis of adverse events in mesh implant surgery reports. *Social Network Analysis and Mining, 14*(1), 63. https://doi.org/10.1007/s13278-024-01229-6

Cao, Q., Cheng, X., & Liao, S. (2023). A comparison study of topic modeling-based literature analysis by using full texts and abstracts of scientific articles: A case of COVID-19 research. *Library Hi Tech, 41*(2), 543-569. https://doi.org/10.1108/LHT-03-2022-0144

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of computers & education. *Computers & Education, 151*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Choi, J., & Lee, B. (2023). Quantitative topic analysis of materials science literature using natural language processing. *ACS Applied Materials & Interfaces, 16*(2), 1957-1968. https://doi.org/10.1021/acsami.3c12301

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). Education technology: An evidence-based review. *NBER Working Paper Series*. https://doi.org/10.3386/w23744

Glazkova, A. (2021). Identifying topics of scientific articles with BERT-based approaches and topic modeling. In M. Gupta & G. Ramakrishnan (Eds.), *Trends and applications in knowledge discovery and data mining*. Springer International Publishing. https://doi.org/10.1007/978-3-030-75015-2_10

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv preprint arXiv:2203.05794. https://doi.org/10.48550/arXiv.2203.05794

Gurcan, F., & Cagiltay, N. E. (2023). Research trends on distance learning: A text mining-based literature review from 2008 to 2018. *Interactive Learning Environments, 31*(2), 1007-1028. https://doi.org/10.1080/10494820.2020.1815795

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural language processing: Python and NLTK*. Packt Publishing Ltd.

Hu, X., & Zhang, R. (2022, June). Text classification based on machine learning. In *2022 IEEE International Conference on Artificial Intelligence and Computer Applications* (ICAICA) (pp. 911-916). IEEE. https://doi.org/10.1109/ICAICA54878.2022.9844556

Hussain, S., Ayoub, M., Jilani, G., Yu, Y., Khan, A., Wahid, J. A., ... & Weiyan, H. (2022). Aspect2Labels: A novelistic decision support system for higher educational institutions by using multi-layer topic modelling approach. *Expert Systems with Applications, 209*, 118119. https://doi.org/10.1016/j.eswa.2022.118119

Huston, C. L., Phillips, B., Jeffries, P., Todero, C., Rich, J., Knecht, P., ... & Lewis, M. P. (2018, January). The academic-practice gap: Strategies for an enduring problem. *Nursing Forum, 53*(1), 27-34. https://doi.org/10.1111/nuf.12216

Jiang, S., Prasad, A., Kan, M. Y., & Sugiyama, K. (2018, August). Identifying emergent research trends by key authors and phrases. In *Proceedings of the 27th international conference on computational linguistics* (pp. 259-269). https://aclanthology.org/C18-1022

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence, 2*, 100016. https://doi.org/10.1016/j.caeai.2021.100016

Maphosa V., & Maphosa M. (2023). Artificial intelligence in higher education: A bibliometric analysis and topic modeling approach. *Applied Artificial Intelligence, 37*(1), https://doi.org/10.1080/08839514.2023.2261730

McInnes, L., Healy, J., & Astels, S. (2017). HDBSCAN: Hierarchical density based clustering. *The Journal of Open Source Software, 2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). *Umap: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426. https://doi.org/10.48550/arXiv.1802.03426

Nylander, E., Fejes A., & Milana M. (2022). Exploring the themes of the territory: A topic modelling approach to 40 years of publications in International Journal of Lifelong Education (1982–2021). *International Journal of Lifelong Education, 41*(1), 27-44. https://doi.org/10.1080/02601370.2021.2015636

Pandur, M. B., Dobša, J., & Kronegger, L. (2020, January). *Topic modelling in social sciences: Case study of web of science* [Conference Paper]. Central European Conference on Information and Intelligent Systems, Varazdin. https://www.proquest.com/openview/f5b35b8664922bcf629d4a57a6344a45/1?pq-origsite=gscholar&cbl=1986354

Ranjbar-Sahraei, B., & Negenborn, R. R. (2017). *Research positioning & trend identification: A data-analytics toolbox.* TU Delft. http://resolver.tudelft.nl/uuid:e736746b-7cba-4239-ac87-49634ccb43e4

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using Siamese BERT-networks.* arXiv preprint arXiv:1908.10084. https://doi.org/10.48550/arXiv.1908.10084

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Wang, Z., Chen, J., Chen, J., & Chen, H. (2023). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics*, 1-26. https://doi.org/10.1007/s11192-023-04776-5

Yunita, S. (2018). Theme and thematic progression in students' recount texts. *Indonesian Journal of Applied Linguistics, 7*(3), 524-530. https://doi.org/10.17509/ijal.v7i3.9797