

Data-Driven Insights into Non-Purchasing Behaviours through Latent Dirichlet Allocation: Analysing Study Material Acquisition Among University Students

Moeketsi Mosia^a


a. Centre for Teaching, Learning and Programme Development, Sol Plaatje University, Kimberley, South Africa.
Email: moeketsi.mosia@spu.ac.za

Article Info

Received: November 19, 2023

Accepted: January 8, 2024

Published: March 15, 2024

 10.46303/jcve.2024.5

How to cite

Mosia, M. (2024). Data-driven insights into non-purchasing behaviours through latent dirichlet allocation: Analysing study material acquisition among university students. *Journal of Culture and Values in Education*, 7(1), 72-82.
<https://doi.org/10.46303/jcve.2024.5>

Copyright license

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

ABSTRACT

The advent of technology has shifted higher education students' educational resource acquisition from traditional printed textbooks to digital formats. This shift is underscored by recent scholarship. The current study extended the discourse on student educational resource acquisition, beyond digital preference. It offers a nuanced view of strategies encompassing digital resources and traditional methods such as library and peer borrowing. The aim of this study was to investigate non-purchasing behaviours of university students in acquiring study material. An online survey link was sent through email to all registered students in the university, out of which a total of 1500 participated – this sample included all students from the four schools. The survey sought to answer the following questions: How do you access books? Besides buying, what other strategies have you employed to access books? Data were analysed using Dirichlet allocation for topic modelling to extract the topics from participants' qualitative responses. The study's findings reveal a complex interplay of socio-economic factors that influence student behaviour. Digital downloads emerged as a prominent strategy, indicating a reliance on online resources. However, borrowing from libraries and peers was significant, which highlights the role of social learning communities. This study contributes to the understanding of how the cost of textbooks impacts academic choices and success, emphasising the social justice implications of textbook affordability. The rise of alternative media suggests a need to reimagine educational resources. It potentially reflects a generational shift in media consumption. This study's insights advocate for educational institutions and policymakers to reassess resource provisioning to align with the evolving landscape of higher education students' academic resource acquisition. This will ensure equitable access to educational material.

KEYWORDS

Data-driven; non-purchasing behaviours; latent dirichlet allocation; study material.

INTRODUCTION

The advent of technology has catalysed a paradigmatic shift across multiple sectors, and significantly within higher education. The paradigmatic shift includes the move from traditional printed texts to digital resources, which is a profound transformation in educational resource delivery. This shift has equally captured scholarly attention, thus, prompting investigations into its consequences for student success. Scholars offer an incisive discourse on the adoption of technology in academia, and highlight the salient benefits of openly licensed textbooks. This argument is buttressed by Harjono et al. (2021), who underscore the affordability, portability, and accessibility of electronic textbooks as pivotal determinants of student preferences. Challenges to access study material are not only faced by undergraduate students. In fact, doctoral students also face difficulties in accessing affordable study material – which prompts libraries to support them through Open Educational Resources (OER) and other avenues (Harjono et al., 2020).

Furthermore, Appedu et al. (2021) present a nuanced perspective that correlates the challenges of accessing educational resources with economic impediments. Appedu et al.'s (2021) study delineates the adversities that are confronted by first-generation students, who disproportionately struggle to access educational resources. Appedu et al. (2021) further observe that first-generation students frequently devise coping strategies to contend with the financial burden of textbook procurement. Although these strategies are well-intentioned, they may inadvertently compromise students' educational success. The barriers to students' access to educational resources underscore the social justice dimension of textbook affordability, as argued by Jenkins et al. (2020), who view OER as a potential equaliser.

Previous studies have shown that early and easier access to educational resources may heighten students' academic engagement (Barkhuizen, 2013; Kim et al., 2019). Kim et al., (2019)'s study found that students who had access to textbooks during the first week of their academic programmes had significantly better academic outcomes. Sansom et al. (2021) support the forgoing findings: they observed no negative effect on student performance in general chemistry, when OER was adopted, even after accounting for various academic factors. Furthermore, Magro and Tabaei (2019) indicate that OER can improve learning outcomes in the psychology domain by enhancing textbook access. These summative findings emphasise the success metrics linked to the use of free textbooks for students who may have been marginalised by the prohibitive cost of study material. The findings are echoed by Bali et al. (2020) who report on the development of comprehensive strategies to make course content more affordable through cross-unit collaborations.

Students' perceptions play a significant role in the impact of Zero Textbook Cost (ZTC) adoptions, they influence purchase and access behaviours (Pfannenstiel et al., 2020). However, Pfannenstiel et al. (2020) caution that despite the recognised importance of online accessibility, a large proportion of students delay acquiring textbooks until the semester is underway. In the

context of the current study, similar observations were made. Initially, the funding for the students' study material was centralised and administered through the university, which forced students to use the funding only for study material such as textbooks. However, when the funder changed the process and paid the funds directly into students' bank accounts, it was observed that most students did not use the funds for purchasing textbooks. This context prompted this study – to explore what strategies are used by students to access the study material. To achieve this aim, a Latent Dirichlet allocation (LDA) was used to collate the topics that emerged from students' shared strategies.

LITERATURE REVIEW

In this section, the paper conceptualises the study by discussing existing related literature. This locates the study within the broad body of knowledge on the student's non-purchase of textbooks. Secondly, the section reviews literature on LDA as a lens to approach a problem through making sense of large text data.

Related Studies

In the realm of higher education, a growing body of research has been focusing on the complexities surrounding students' non-purchasing behaviours in relation to textbooks and learning materials. This literature reveals a multifaceted landscape where economic factors, technological advancements, and social justice issues intersect. Central to this discourse is the impact of textbook costs on students' academic choices and success. Studies by Martin et al. (2017) and Wittkower & Lo (2019) have shown a direct correlation between the high cost of textbooks and students' academic performance, with many students opting not to purchase required materials due to financial constraints. Furthermore, other studies shed light on the phenomenon of students' non-purchasing behaviours, revealing an unequal financial burden which is placed on students (Appedu et al., 2021; Becker et al., 2022).

In an attempt to resolve the challenge faced by students in accessing study material, there have been ongoing considerations and discussions on the exploration of Open Educational Resources (OER). OER is primarily viewed as a potential solution to alleviate the financial burden on students, with the expectation that it will reduce costs and lead to improved student success (Hilton et al., 2016). Offering contrasting viewpoints, Jenkins et al. (2020) view OER through a social justice lens, and argue that textbook affordability is an essential aspect of achieving equitable education.

Furthermore, from students' perspectives other studies, have contributed to the complexities involved in of the adoption of OER (Falc, 2013; & Su, 2021). Falc (2013) and Su (2021) reveal a spectrum of student perceptions towards e-textbooks, ranging from neutral to mixed. This diversity in opinions suggests that the adoption of OER might not be a completely sufficient solution to address the financial burdens that are faced by students. This suggests that there is a misalignment between students' expectations for access and efficiency of the designed learning material. (Philip & Moon, 2013). Adding to this discussion, also from a student

perspective, Pfannenstiel et al. (2020) argue that students' prior experiences impact their textbook purchasing behaviour – especially as it relates to zero textbook cost material. This insight is crucial for understanding the dynamics involved in the textbook selection and the adoption of cost-effective alternatives. Moreover, the link between textbook usage and student learning outcomes has been increasingly emphasised. The study by Pavešić and Cankar (2022) underscores the need for systematic collection of data on textbook use to inform the development of future textbooks and learning material, thereby improving educational quality.

Topic Modelling with Latent Dirichlet Allocation

Topic modelling is an analytical method that employs unsupervised machine learning algorithms to uncover latent themes within textual data. This technique has been found to be, amongst others, particularly beneficial in consumer research for understanding consumer behaviours and preferences, thus aiding in marketing and product development strategies. For instance, Jung et al. (2023) utilised structural topic modelling (STM) to examine social media data, and revealed the approach-avoidance behaviours of Indonesian and Malaysian Muslim consumers towards Korean beauty products. Jung et al.'s (2023) study illustrates STM's efficacy in cross-cultural consumer behaviour analysis.

Bai et al. (2023) expanded upon Latent Dirichlet Allocation (LDA) using STM to incorporate document information into customer concern analysis within the sharing economy context. This enabled a nuanced understanding of evolving customer concerns, offering businesses actionable insights to enhance their services. Zhuo (2022) introduced word vectors to traditional LDA to refine consumer demand behaviour analysis. The comparison between the enhanced and traditional models aimed to more precisely and efficiently understand consumer demands, underscoring the improved accuracy word vectors can bring to topic modelling. Lastly, Wang and Yang (2021) discussed the application of topic modelling to extract numerical measures from consumers' textual product evaluations. This method provides a potent means to decode the reasons for consumers' word choices in reviews.

More broadly, topic modelling has emerged as a transformative technique in qualitative research. It presents a scalable and reproducible method for analysing textual data. It utilises algorithms such as LDA to detect underlying themes within a corpus of documents, thus, facilitating theory development and the discovery of emerging patterns. Hannigan et al. (2019) emphasised the significance of topic modelling in management research, proposing that it unlocks the potential to distil new theories from extensive qualitative data sets. This method is influential because of its ability to systematically unearth insights that traditional analysis might overlook.

Dieckmann et al. (2022) portrayed topic modelling as an automated means of theme identification within a body of text. This systematised analysis streamlines the interpretation of qualitative data, which enables the revelation of hidden patterns and relationships. In

healthcare, Shaw et al. (2022) demonstrate the application of topic modelling for decoding health-related information from social media, offering crucial insights for healthcare stakeholders. Similarly, Rutkowski et al. (2022) showed how STM can refine the analysis of qualitative data by revealing latent topics, thereby enhancing the analytical rigour.

METHODOLOGY

Data collection for this study was conducted via an open-ended questionnaire that was designed to probe the following inquiries:

1. What alternative methods or strategies, other than purchasing from an on-campus bookstore, do students employ to obtain textbooks?
2. What are the presumed practices of peers who do not purchase textbooks to access study materials?

The above questions facilitated discussions that elicited insights into the myriad strategies that students adopt to acquire textbooks without resorting to purchasing from the on-campus bookstore. The qualitative data amassed from the survey consisted of a diverse range of descriptive tactics that were utilised by students. This survey was administered to a cohort of 1500 students from four distinct academic schools. No specific group of students was sampled. Rather, all registered students were targeted. Data collection was strategically administered during the fourth term of the academic year, a juncture at which it is assumed that all students intending to purchase textbooks would have already done so.

The preprocessing of the dataset involved tokenisation, a process in which sentences were segmented into discrete elements, namely words or terms. Subsequently, frequent terms in natural language, often termed 'stop words', including, 'the', 'and', and 'is', were systematically excised from the dataset. Furthermore, the data was lowercased to ensure uniformity and stemming was applied to reduce words to their root form.

Topic Modelling – Latent Dirichlet Allocation

The core of the analysis employed Latent Dirichlet Allocation (LDA), a generative statistical model of the corpus (Blei et al., 2003). Mathematically, as Blei et al., (2003) put it, for a corpus of D documents and a chosen number of topics T , the LDA model aims to infer the topic distribution θ_d for each document and the word distribution β_t for each topic, under the assumption that K topics are present across the corpus.

The generative process for each document $d \in D$ is as follows:

1. Choose $N_d \sim \text{Poisson}(\xi)$ to denote the number of words in document d
2. Choose $\theta \sim \text{Dir}(\alpha)$ to denote the distribution of the topic in document d
3. For each of the word positions $i \in \{1, \dots, N_d\}$, in document d :
 - a. Choose a topic $z_{d,i} \sim \text{Multinomial}(\theta_d)$
 - b. Choose a word $w_{d,i}$ from $p(w_{d,i} | z_{d,i}, \beta_{z_{d,i}})$

The observed data $w_{d,i}$ are the words in the document which mean are the word from student k , and the latent variables θ_d and $z_{d,i}$ are the topic structures that we aim to infer. The

distributions β_t are the parameters of the multinomial distribution over words for topic t , which are also inferred from the data. The prior distributions for θ_d and β_t are Dirichlet distributions with hyperparameters α and η respectively. To perform the inference of the latent variables we use variational Bayesian methods utilising a variational expectation-maximisation algorithm to approximate the posterior distributions of interest.

Model Implementation and Analysis

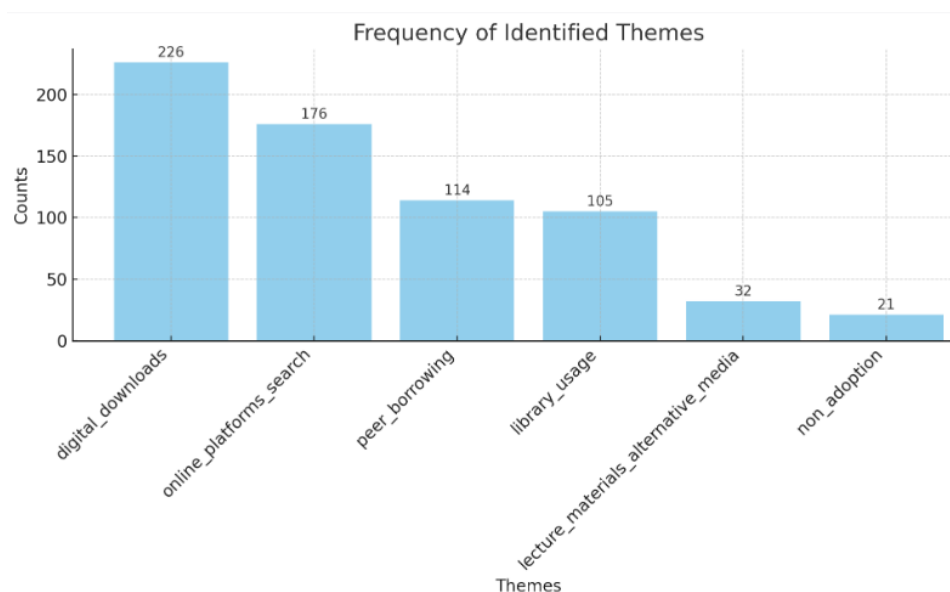
To perform the analysis Python library scikit-learn CountVectorizer was employed to convert the text documents into a matrix of token counts. We set the upper threshold for the proportion of documents in which a term must appear to 95% and a lower threshold of 2 documents. This matrix served as the input for the LDA model, which was instantiated with $T = 5$ to extract five topics from the data. The model was fit to the data, and the resulting topic-word distributions were examined to interpret the themes from the students' responses.

RESULTS AND DISCUSSIONS

To answer the question: "What alternative strategies, other than purchasing, that students employ to obtain textbooks?" a descriptive analysis was performed, and the results are reported in the bar chart below (figure 1). Figure 1 shows the frequency of identified themes from the students' responses. Each bar represents a theme. The height of each bar indicates the number of mentions for each theme.

Figure 1.

Systematic Breakdown of the Most Frequent Themes



The theme 'Digital Downloads' emerged as the most prevalent, indicating a substantial number of students reporting having acquired books from the internet. This trend suggests that students significantly relied on digital resources. The prominence of 'Library' as the fourth most cited strategy underscores the role that institutional support plays through providing resources

via the library. The fact that 'Library' does not rank as the foremost strategy presents an opportunity for libraries to amplify their digital support services to better meet students' resource access needs. Furthermore, it was interesting to observe that one of the most common strategies among students was to borrow textbooks from their peers, a practice that was, in fact, more common than using the library. More so, there was a move for more alternative media such as YouTube videos and other forms of media to replace traditional textbooks. A few students mentioned not exploring other options, which could indicate either satisfaction with their current methods or a lack of awareness about the alternatives.

Figure 2.

Top 16 words used across the topics.

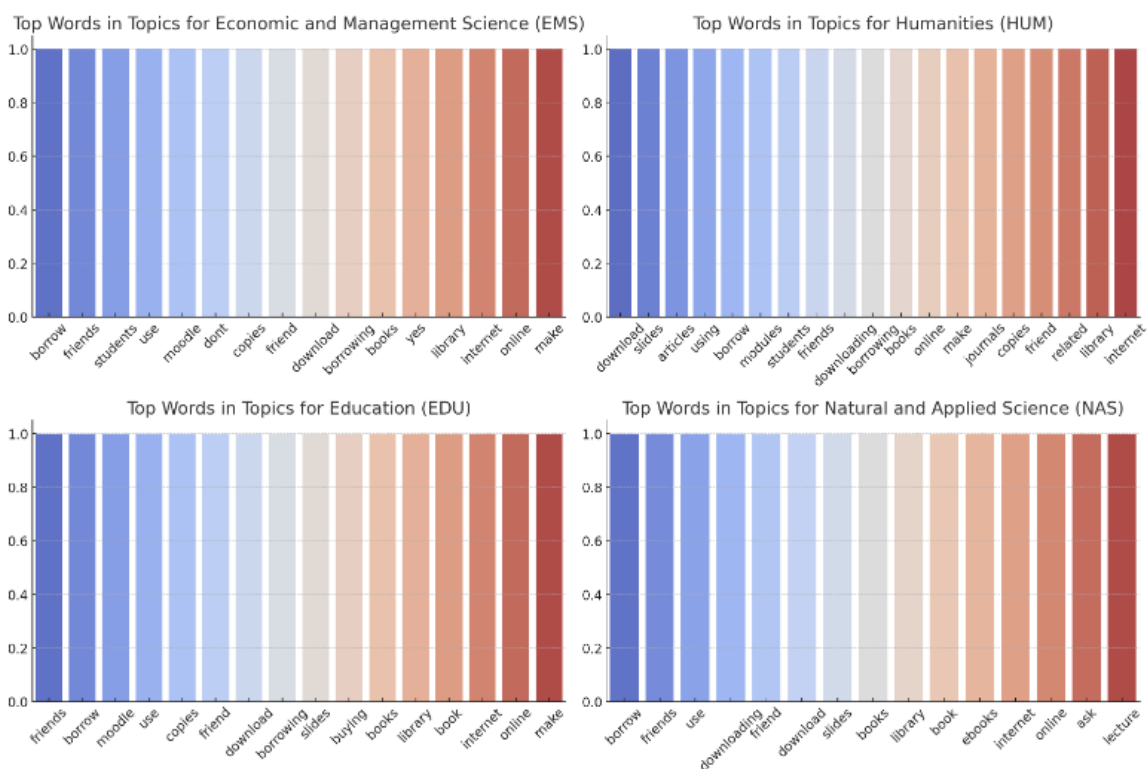


Figure 2 demonstrates that the distribution of the three most frequently used words within the topic is fairly consistent across all four schools. This descriptive analysis was insufficient to make conclusive claims. Therefore, a more robust method of inferential analysis was needed to extract topics from students' reflections on their strategies – thus, the next section presents the findings of the LDA.

Latent Dirichlet Allocation

In the exploration of strategies that students employed to access books without direct purchase LDA topic modelling was applied to their textual responses. The analysis revealed five distinct topics, each representing a unique strategy. The first topic underscored the use of lecture material and alternative media, with significant weight attributed to terms such as 'lecture slides' and 'PDFs,' which pointed to a reliance on institutional provisions and self-curated content. The second topic highlighted the traditional approach of borrowing from libraries and

friends, signalling the enduring relevance of social and educational networks in resource sharing. Peer-to-peer exchange was central to the third topic, reflecting students' collaborative ethos to share prescribed material and notes. The fourth topic was dominated by online access and downloading behaviours, illustrating a pronounced shift towards digital resources, where terms such as 'online' and 'internet' were prevalent. The fifth topic revealed resourceful behaviours in digital and second-hand book access, with search engines and second-hand purchases indicating a blend of cost-effective and accessible strategies.

This multifaceted thematic landscape, as depicted in figure 1, illustrates a dynamic range of book acquisition strategies with a clear predominance of digital and online resources, supplemented by traditional borrowing. The transition towards electronic resources suggests an adaptive response to the financial and logistical challenges posed by textbook procurement. The visualisation provided in figure 1, a bar chart of the top words within each topic, offers a clear and immediate representation of the thematic weightings and underscores the key findings of this analysis. The participants' responses were categorised using term document metrics analysis and the most frequent phrases were visualised in Figure 2.

DISCUSSION OF FINDINGS

The literature review conducted for this study indicates that transformative shifts are occurring in how students acquire learning material. This evolution is underscored by the work of Jenkins et al. (2020) and Harjono et al. (2020), which emphasise the affordability and accessibility of electronic textbooks as a barrier. Consistent with the current study's findings, it was revealed that students are increasingly resorting to digital downloading of learning material. The preference for digital resources can be attributed to students' ease of access and portability, in agreement with the observations made by Appedu et al. (2021).

However, the present study extends beyond merely identifying a preference for digital content. It offers a nuanced comprehension of the strategies employed by students, encompassing digital resources and conventional methods such as borrowing books from libraries as well as peers. Such borrowing practices further emphasise the significance of social learning communities. Furthermore, this indicates a more complex behaviour than previously understood, where economic factors drive the transition to digital media as well as the reinforcement of traditional social and educational networks (Wittkower & Lo, 2019). This analysis provides insight into how textbook costs influence academic decisions and success, as highlighted by Martin et al. (2017). Additionally, the continued reliance on borrowing accentuates the critical demand for alternative strategies, directly reflecting the socioeconomic status of students and thereby underscoring the social justice aspects of textbook affordability. This connects with Jenkins et al. (2020), who view OER through a social justice lens, positing textbook affordability as essential for equitable education.

This study shows the proliferation of alternative media, including instructional videos, suggesting a need for textbook alternatives and a rethinking of educational material. Despite past research, such as Johnston and Salaz (2019), showing a preference for print over digital media, our analysis highlights a need for more adaptive study material that support a variety of learning and teaching modalities. This may suggest a generational shift in media use, which might affect instructional resource design. The Social Sciences Animation Video-Based Teaching Material (S2AV) research by Arif et al. (2023) supports these findings.

This article promotes new and interesting teaching resources to assist students. Arif et al. (2023) even said that animated video-based resources can improve students' learning experience, corroborating the current study's finding that textbooks are not essential. Apit et al. (2023) suggest that STEM-based mobile learning apps can help teachers and students learn. Non-print textbooks imply that the technology-driven approach may improve learning results compared to textbook-based techniques (Apit et al., 2023).

CONCLUSION

The current study aimed to enhance our comprehension of student study material purchasing behaviour. The study's findings disclosed that students employ strategies that are economical, social, and technological in nature. The findings highlighted that NSFAS-funded students may use their funds for alternative needs. The paper advocates that institutions of higher learning may need a reassessment of how educational resources are conceptualised, accessed, and distributed to students. Furthermore, it is essential for policymakers to acknowledge the changing landscape of student resource acquisition, which demands innovative models to guarantee equal access to educational material. It is important to note that this study was delineated to one South African university and thus for future it may be necessary to include more universities including those in other countries.

REFERENCES

- Appedu, S., Elmquist, M., Wertzberger, J., & Birch, S. K. (2021). Inequitable impacts of textbook costs at a small, private college: results from a textbook survey at Gettysburg College. *Open Praxis*, 13(1), 69. <https://doi.org/10.5944/openpraxis.13.1.1147>
- Arif, V. R., Afnan, M., Usmiyatun, U., & Lestari, C. Y. (2023). Development of Social Studies Animation Video (S2AV) Teaching Materials on the Material "Plurality of Indonesian Society" for Junior High School Students. *Assyfa Learning Journal*, 1(1), 01-11.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Bali, M., Cronin, C., & Jhangiani, R. (2020). Framing open educational practices from a social justice perspective. *Journal of Interactive Media in Education*, 2020(1). <https://doi.org/10.5334/jime.565>

-
- Barkhuizen, N., Rothmann, S., & Vijver, F. J. R. v. d. (2013). Burnout and work engagement of academics in higher education institutions: effects of dispositional optimism. *Stress and Health*, 30(4), 322-332. <https://doi.org/10.1002/smi.2520>
- Bai, S., Bi, X., Han, C., Yang, M., & He, H. (2023). Exploring the over-time variation in customer concerns on sharing economy services. *Journal of Global Information Management*, 31(1), 1-21. <https://doi.org/10.4018/jgim.321435>
- Dieckmann, N., Stoyles, S., Aebischer, J., & Álvarez, H. (2022). Dissertation topics in nursing. *Nursing Research*, 71(5), 380-386. <https://doi.org/10.1097/nnr.0000000000000604>
- Falc, E. O. (2013). An assessment of college students' attitudes towards using an online e-textbook. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, 9, 001-012. <https://doi.org/10.28945/1767>
- Harjono, A., Gunawan, G., Adawiyah, R., & Herayanti, L. (2020). An interactive e-book for physics to improve students' conceptual mastery. *International Journal of Emerging Technologies in Learning (IJET)*, 15(05), 40. <https://doi.org/10.3991/ijet.v15i05.10967>
- Hannigan, T., Haans, R., Vakili, K., Tchalian, H., Glaser, V., Wang, M., ... & Jennings, P. (2019). Topic modeling in management research: rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632. <https://doi.org/10.5465/annals.2017.0099>
- Hilton, J. (2016). Open educational resources and college textbook choices: a review of research on efficacy and perceptions. *Educational Technology Research and Development*, 64(4), 573-590. <https://doi.org/10.1007/s11423-016-9434-9>
- Jung, M., Lee, Y., & Chung, J. (2023). Cross-national consumer research using structural topic modelling: consumers' approach-avoidance behaviours. *International Journal of Consumer Studies*, 47(5), 1692-1713. <https://doi.org/10.1111/ijcs.12923>
- Rutkowski, R., Lee, J., Coller, R., & Werner, N. (2022). How can text mining support qualitative data analysis?. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1), 2319-2323. <https://doi.org/10.1177/1071181322661535>
- Shaw, G., Zimmerman, M., Vasquez-Huot, L., & Karami, A. (2022). Deciphering latent health information in social media using a mixed-methods design. *Healthcare*, 10(11), 2320. <https://doi.org/10.3390/healthcare10112320>
- Jenkins, J. J., Sánchez, L. A., Schraedley, M., Hannans, J., Navick, N., & Young, J. (2020). Textbook broke: textbook affordability as a social justice issue. *Journal of Interactive Media in Education*, 2020(1), 3. <https://doi.org/10.5334/jime.549>
- Kim, H. J., Hong, A. J., & Song, H. (2019). The roles of academic engagement and digital readiness in students' achievements in university e-learning environments. *International Journal of Educational Technology in Higher Education*, 16(1). <https://doi.org/10.1186/s41239-019-0152-3>
-

- Magro, J. and Tabaei, S. (2019). Results from a psychology oer pilot program: faculty and student perceptions, cost savings, and academic outcomes. *Open Praxis*, 12(1), 83. <https://doi.org/10.5944/openpraxis.12.1.1007>
- Martin, M. T., Belikov, O. M., Hilton III, J., Wiley, D., & Fischer, L. (2017). Analysis of Student and Faculty Perceptions of Textbook Costs in Higher Education. *Open Praxis*, 9(1), 79-91. DOI: <https://doi.org/10.5944/openpraxis.9.1.432>
- Philip, G. and Moon, S. (2013). An investigation of student expectation, perceived performance and satisfaction of e-textbooks. *Journal of Information Technology Education: Innovations in Practice*, 12, 287-298. <https://doi.org/10.28945/1903>
- Pfannenstiel, A. N., Redcay, A., & Albert, D. R. (2020). Student perceptions of textbooks: prior behaviors and beliefs can influence zero textbook cost (ztc) adoption impact. *Open Praxis*, 12(4), 555. <https://doi.org/10.5944/openpraxis.12.4.1119>
- Pavešić, B. J. and Cankar, G. (2022). Textbooks and students' knowledge. *Center for Educational Policy Studies Journal*, 12(2), 29-65. <https://doi.org/10.26529/cepsj.1283>
- Sansom, R., Clinton, V., & Fischer, L. (2021). Let students choose: examining the impact of open educational resources on performance in general chemistry. *Journal of Chemical Education*, 98(3), 745-755. <https://doi.org/10.1021/acs.jchemed.0c00595>
- Wittkower, L. R. and Lo, L. S. (2019). Undergraduate student perspectives on textbook costs and implications for academic success. *Open Praxis*, 12(1), 115. <https://doi.org/10.5944/openpraxis.12.1.1036>
- Wang, Z. and Yang, F. (2021). Mining numerical measure of consumers' product evaluation expressed in words based on latent dirichlet allocation. *Journal of Modelling in Management*, 18(1), 147-170. <https://doi.org/10.1108/jm2-07-2021-0163>
- Zhuo, J. (2022). Consumer demand behavior mining and product recommendation based on online product review mining and fuzzy sets. *Mathematical Problems in Engineering*, 2022, 1-9. <https://doi.org/10.1155/2022/1216475>