

# 學習分析與教育資料探勘之研究現況與趨勢 - 應用書目計量分析法之探究

Current Status and Trends in Learning Analytics and Educational Data Mining Research - An Inquiry Using Bibliometric Analysis

王舒民<sup>1\*</sup>, 洪暉鈞<sup>2</sup>

<sup>1</sup> 中國文化大學 資訊管理學系

<sup>2</sup> 中央大學 網路學習科技研究所

\* scottie.wang@gmail.com

**【摘要】** 資訊科技的發展使用蒐集與分析學習者的行為歷程變得便利快速，也使得學習分析之應用更為多元。本研究應用書目計量分析法，針對 2015-2024 年間 Web of Science 資料庫中 1,138 篇學習分析相關文獻進行分析。研究透過引用分析與共現詞分析等方法，探討學習分析領域之研究現況與趨勢，其中 Computers & Education 具最高引用影響力分析顯示，此外，學習分析著重於學習行為研究，而教育資料探勘則側重於預測分析演算法的應用。本研究為學習分析領域研究者提供重要的文獻計量視角，可做為未來研究方法之參考。

**【關鍵詞】** 學習分析；教育資料探勘；書目計量分析；引用分析；共現詞分析

**Abstract:** Advances in information technology have made collecting and analyzing learner behavioral data more efficient and accessible, leading to diverse applications of learning analytics. This study employs bibliometric analysis to examine 1,138 learning analytics-related articles indexed in Web of Science database from 2015 to 2024. Through citation analysis and co-word analysis, we investigate research trends in learning analytics, with Computers & Education showing the highest citation impact. The analysis reveals that learning analytics research focuses on learning behavior, while educational data mining emphasizes predictive analytics algorithms. This study provides researchers with valuable bibliometric perspectives for future research methodologies in learning analytics.

**Keywords:** learning analytics, educational data mining, bibliometric analysis, citation analysis, co-word analysis

## 1. 前言

隨著資訊科技的快速發展，應用數位工具以輔助學習已成為一個常態。有別於實體世界，在數位環境中，運用適當的系統工具，可以蒐集、識別與紀錄學習者的行為與歷程，從而可據以設計不同的教學策略，提供適切的學習支持以提昇學習成效 (Palanci et al., 2024)。

近年來，在科技的輔助下，學習者於學習環境中之行為、成果之蒐集以較過往便利許多。如何能應用學習分析，無論是透過過往的資料來預測未來的表現，以做為教學資源的投入、課程設計，乃至於政策規劃之參考。或者是在數位學習環境中，設置即時的行為分析工具，透過學習代理人提供引導、鷹架支持等個人化的適性學習支持，都逐漸吸引到各方的關注 (Al-Shabandar et al., 2018; Casquero et al., 2016; Siemens, 2013)。

儘管過去已有數篇針對學習分析或教育資料探勘之系統化文獻回顧，惟其多使用質性之系統化文獻回顧法，限於需要深度閱讀與分析各文獻，往往分析的文章數較少，且多以特別主題進行文獻回顧 (Banihashem et al., 2018; Matcha et al., 2020; Palanci et al., 2024; Zacharoula & Anastasios, 2014)。有鑑於學習分析此一研究主題的重要性與價值，理解相關文獻之整體趨勢與概觀，有助於研究人員思考未來研究方向，促進此一主題領域之發展。為進一步瞭解學習分析相關主題研究領域之現況與趨勢，本研究之主要目的即為透過文獻回顧分析，定位學習分析領域中之重要期刊、高影響力文獻、重要研究貢獻者與主題發展趨勢。由上述，本研究應用書目計量分析法，以過去十年 (2015-2024) 期間所發表之文獻進行量化文獻回顧。期能

提供一個替代視角，供予學習分析與學習評估之相關研究者們做為參考。具體之研究方法與分析內容請見下節。

## 2. 研究方法

### 2.1. 書目計量分析法

本研究之主要目的為分析學習分析與學習評估相關研究主題之研究趨勢、主題發展，並定位核心作者、重要期刊與高影響力文獻等。由此，本研究使用書目計量分析法 (bibliometric analysis) 做為主要研究工具。有別於質性系統化文獻回顧分析專注於文本的深入分析與綜整比較研究成果之趨勢與發現，書目計量分析使用量化的觀點，萃取文稿的資訊，例如：文章名稱、發表期刊、作者、主題關鍵字等，並透過分析文稿間的引用關係，建立各文獻之網路關係，從而能呈現特定領域研究成果之趨勢、文獻關聯性、知識結構、主題發展，以及作者、機構等發表狀況。其主要的分析方式包括：引用分析 (citation analysis)、共被引分析 (co-citation analysis)、共現詞分析 (co-word analysis) 等，並透過主題關鍵字之叢集分析 (cluster analysis) 與網路分析，使用視覺化的技術呈現分析結果，其有助於研究者理解研究領域之發展趨勢、主要研究貢獻者等，做為深入探索與發展未來研究方向之參考。

### 2.2. 文獻蒐集與分析工具

本研究使用 Web of Science (以下簡稱 WoS) 做為主要資料蒐集來源。選擇該資料庫之主要考量為該資料庫所收錄文獻多為 SSCI 或 SCI 索引之期刊，為現今學術界較為普遍認可之索引指標，具相當程度之嚴謹性與代表性。本研究參考本子會議之主題，設定關鍵字於 WoS 中之主題欄位 (Topic) 進行文獻搜尋。WoS 之主題欄位為一個複合欄位，所設定之關鍵字將於文章名稱、摘要、作者提供關鍵字，以及 WoS 依據文章內容所自行加注之關鍵字 (Keywords plus) 等欄位中搜尋。本研究所設定之關鍵字如下所示：

(TS=("learning analytic\*" OR "educational data mining" OR "academic analytics" OR "education\* mining" OR "education\* data science"))

AND

TS=("learning assessment\*" OR "learning evaluation\*" OR "learning feedback" OR "student diagnosis" OR "learning outcome\*" OR "performance predict\*" OR "learning progress" OR "learning process\*" OR "academic performance" OR "student performance"))

文獻搜尋期間為 2015 – 2024 年合計十年，另為使搜集文獻更加聚焦與一致，本研究以期刊論文為主要分析標的，排除文獻回顧 (Review Article)、研討會文章 (Proceeding paper)、書籍與專章 (Book and Book chapter) 等文獻類型，本研究以英文文獻為主要分析標的。

綜合上述條件，於 WoS 資料庫中搜尋後，共得 1,138 篇文章做為本研究之分析資料集。

本研究使用 R 語言為分析工具，使用套件為 bibliometrix 4.3.0 版 (Aria & Cuccurullo, 2017)。並使用 Python 進行部份分析結果之資料整理與視覺化。

## 3. 資料分析與結果

### 3.1. 資料總覽與文獻發表趨勢

經前節所述進行文獻搜尋後，共取得 1,138 篇文獻，其整體概況描述如下。1,138 篇文章共來自於 341 份不同期刊，發表作者達 3,482 名，平均每篇作者數為 3.86 名，單一作者文獻為 68 篇，合計共引用 38,017 篇文獻。本主題於 2015-2024 年間發表數呈現穩定成長趨勢，年成長率達 29.09%。各年度發表篇數如下圖 1 所示，圖中呈現於 2019 年及 2024 年呈現兩次較大幅度的發表數量成長，均較前年成長 50 篇左右，顯示此一主題在近年來得到的關注與未來持續發展的潛力。

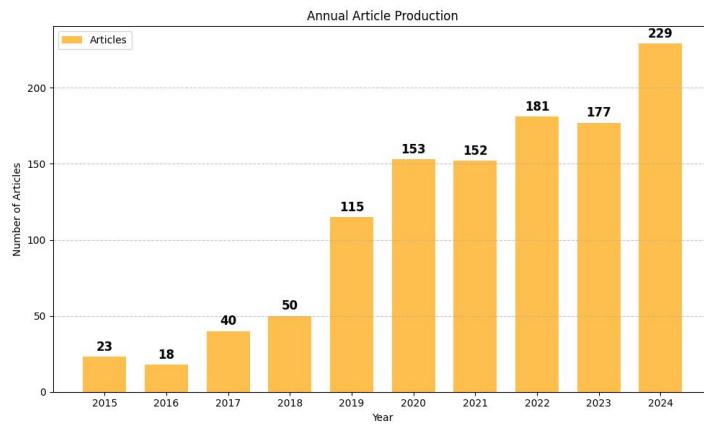


圖 1 2015 – 2024 年間每年發表篇數

### 3.2. 重要作者與主要發表期刊

本研究進一步分析本主題領域之重要作者與發表數量，以及主要發表期刊。下表 1 顯示了主要的作者發表篇數與排名。為更準確地呈現作者之貢獻，表中亦列出作者數調整後之發表篇數。其計算作者於每篇文獻之貢獻度，例如：一篇文章若有兩名作者時，則各計 0.5 發表篇數，依此類推後，加總個別作者之發表篇數後呈現作為參考。由表 1 可看出，在計算加權平均後，作者發表數量排名呈現些許異動，例如：Temperlaar 採計加權發表數量後，排名將提昇至第二名，僅次於 Saqr。透過兩個數字的比較，可以更完整地呈現研究者的生產力。

表 1 作者發表篇數與排名

作者	發表文章篇數	依作者數調整發表篇數*	作者	發表文章篇數	依作者數調整發表篇數*
Gasevic, D.	25	4.28	Kloos, CD.	12	2.88
Saqr, M.	14	5.67	Muñoz-Merino, PJ.	10	2.52
Ogata, H.	13	3.22	Tempelaar, D.	10	4.42
Pardom, A.	13	2.69	Xing, WL.	10	2.30
Rienties, B.	13	3.78	Doneva, R.	8	2.13

圖 2 顯示了學習分析領域累積發表數量前十名之期刊。其中 Education and Information Technologies 十年來累積共出版 61 篇相關文獻。其次為 IEEE Access，再者為 Applied Sciences Basel。由錯誤！找不到參照來源。亦可發現於 IEEE Transactions on Learning Technologies 後，相關文章數量大致趨緩。此一曲線亦符合書目計量學中的布萊德福定律（Bradford's law），概念與長尾效應相近，亦即特定主題的文獻，大多集中於少數的核心期刊，而其它文獻則分布於大量的非核心期刊。故相關主題之研究者，可以參考本研究之分析結果，閱看核心期刊之相關文獻，亦可做為規劃研究投稿標的時之參考。

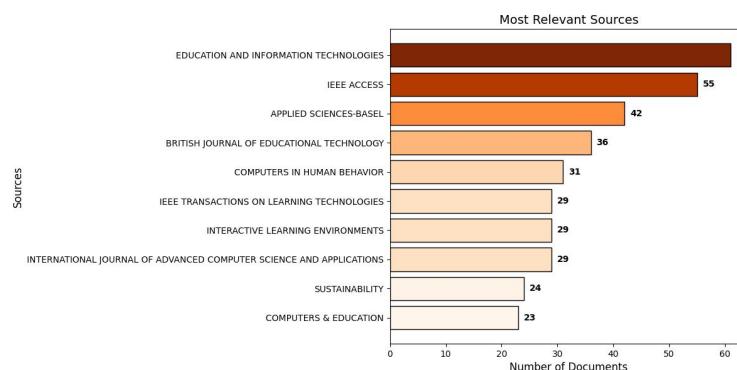


圖 2 期刊發表相關文章數前十名

### 3.3. 引用分析

透過引用分析，可以提供另一個視角，定位出主題領域的核心文獻與發表期刊。延續前節之各期刊發表數量，如圖 3 所示，以被引用次數<sup>6</sup>進行排序，則可發現 Computer & Education 之引用次數遠高於其它期刊。原發表數量為第一之 Education and Information Technologies 則掉落至第七名。進一步以文章為單位分析引用次數，分析結果如下圖 4 所示，前十名高被引用之文章，即有四名發表於 Computers & Education。限於篇幅以下僅就前三名引用次數較高之文獻進行探述。Asif et al. (2017)分析 210 名大學生四年的學習資料，並用以預測學生的畢業表現，研究結果發現使用入學成績與前兩年的課程表現，即能合理預測學生的畢業表現。其應用教育資料探勘 (educational data mining) 之分析方式，包括：預測分析、分群分析，並進行視覺化分析結果。Tempelaar et al. (2015) 則分析 922 位進行混成學習 (blended learning) 之學習數據，包括：系統使用紀錄與各項作業與測驗成績以建立預測模式。分析後發現形成性評量最能有效預測學生表現，。Tomasevic et al. (2020) 的研究則著重於應用多個不同的機器學習演算法進行比較分析，其研究資料集主要來之英國開放大學學習分析資料集 (Open University Learning Analytics dataset, OULAD)，其包含 22 個課程模組，共 32,593 名學生之資料。研究結果發現，類神經網路 (ANN) 在預測學生考試表現有最準確的結果。此三篇研究文獻，其主題均為預測學生之學業表現，也顯示此一主題的研究集中度。

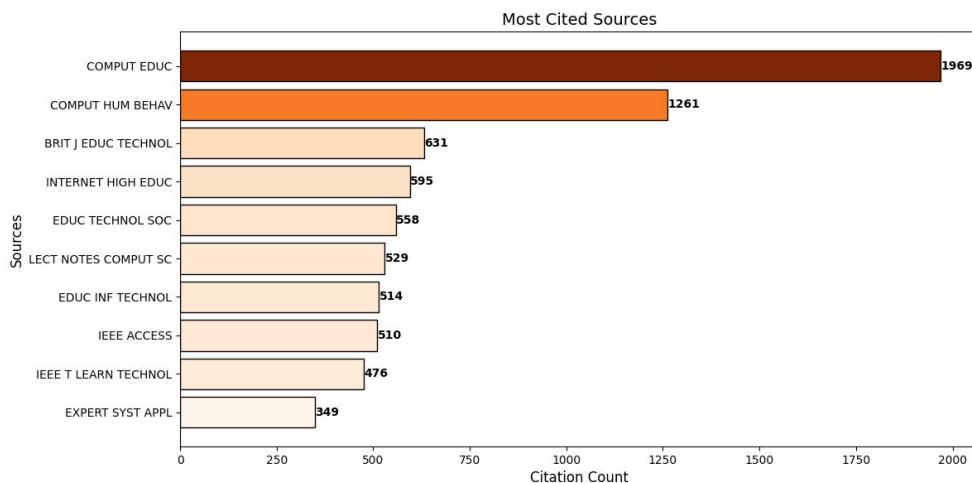


圖 3 期刊累積引用次數前十名

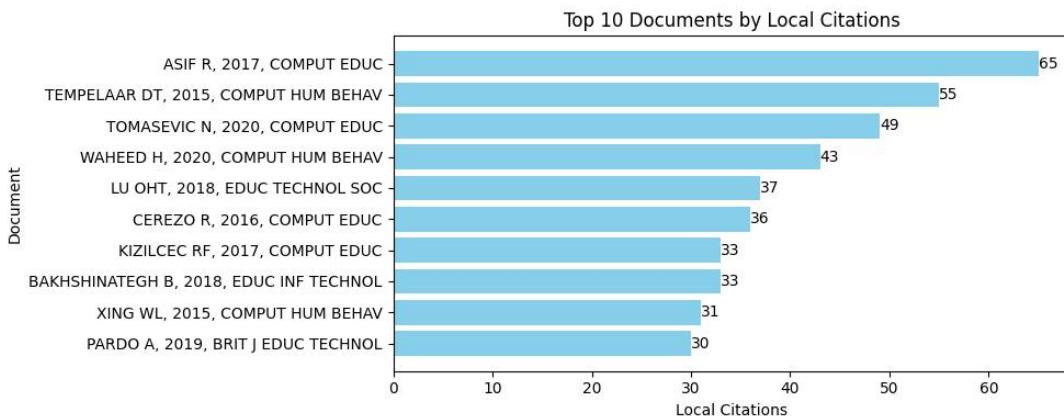


圖 4 文章累積引用次數前十名

### 3.4. 共現詞分析

本研究進一步應用書目計量分析中之共現詞分析 (co-occurrence)，分析關鍵字所形成的

<sup>6</sup> 本節所指引用次數，均為本資料集中之引用次數 (local citation)，用以比較引用關係之相對重要性。

共現網路。分析結果如下圖 5 所示。

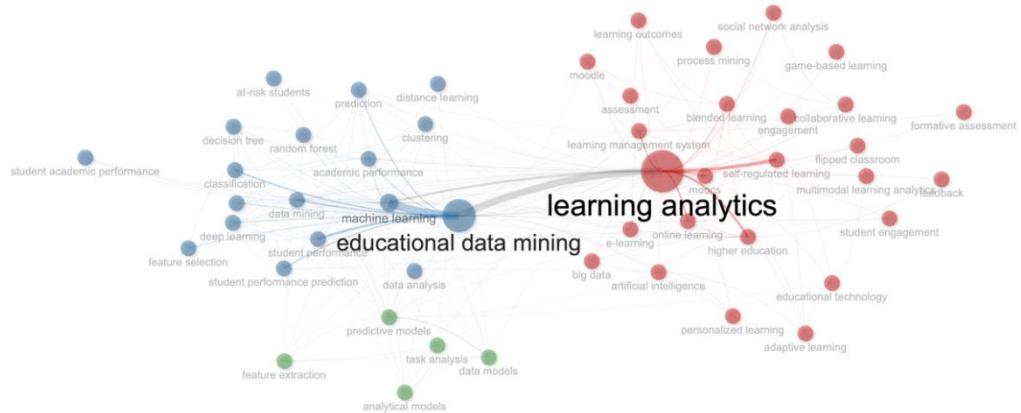


圖 5 共現詞網路圖

圖中呈現了二個主要叢集，一為以 learning analytics 為核心之共現詞網絡，其多為學習形式（如：blending learning, online learning, collaborative learning）與學生行為等主題關鍵字。而另一個主要叢集為以 educational data mining 為核心之共現詞網絡，其中主現了較多的分析方法，如 clustering, feature selection, machine learning 等，並以學生之學業表現為預測標的。比較兩個叢集，可以發現兩種類型的研究有些微不同的著重點。Learning analytics 較為側重不同學習情境中的學生學習行為。而 educational data mining 則有較多關於預測分析的演算法討論。

### 3.5. 主題演進趨勢

本研究進一步探討 2015-2024 年研究各研究主題叢集之演進。應用 Bibiliometrix 提供主題演進 (Thematic evolution) 分析功能，將過去十年分為三個階段，並使用作者提供之關鍵字 (Author's keywords)，建立主題演化圖。其分析方式主要就關鍵字進行叢集分析，探討共現詞叢集在不同時間區間之變化，分析結果如下圖 6 所示。在 2015-2018 這個區間，關鍵字所形成的叢集較為零碎，分為兩大部份，一為機器學習、資料探勘相關演算法 (圖 6 上半部)，二為學習情境與學習分析相關議題。而在 2019-2021 這個期間，學習分析、教育資料探勘與資料探勘各自形成一個主要叢集，代表有許多以此為主題的文章在這個區間發表。最後，在 2022-2024 這個區間，可以發現發表的主題再次趨於零散，同時出現新的科技應用，如：eye-tracking 與新興的分析方法，如：epistemic network analysis；此外，關於學生的學習策略，如：自我調節策略 (self-regulation) 也開始形成主題叢集。透過分析研究主題的演化，有助於概略地掌握學習分析領域中，各研究叢集所探尋的研究主題現況與未來可以進一步探究的議題。

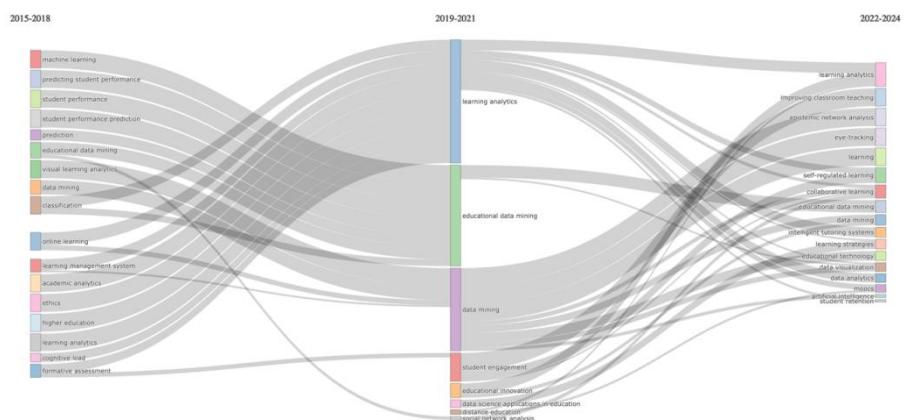


圖 6 研究主題演化圖 (Thematic Evolution)

#### 4. 討論與結論

本研究應用書目計量分析法，就學習分析與學習評估相關文獻進行分析。研究之文獻資料集為 2015-2024 間收錄於 Web of Science 資料中之相關文獻，合計共 1,138 篇。相關研究結果茲分述說明如下。

首先，學習分析領域在過去十年以來便呈現穩定的成長趨勢，顯示此一主題研究領域之重要性。2024 年單年已有 229 篇相關文獻發表。其次，透過圖 2 呈現的分析結果，可以定位出相關文獻主要發表的核心期刊為何。有志於學習分析與學習評估之研究者可參看核心期刊所發表之文章以進一步理解相關研究現況。此外，亦可將核心期刊做為主要投稿標的，提昇研究主題與所投期刊之主題領域切合度。第三，透過引用分析，可定位出主要的核心研究文獻。研究結果發現 Computers & Education 中所發表之研究文獻，為本次文獻資料集中各文獻主要引用之標的。參看圖 4 之分析結果，本研究亦進一步摘述引用次數排名前三之文獻。其共同趨勢為應用不同的資料探勘或機器學習之演算法於預測學生之學業表現。未來可再進一步深入地分析高引用文獻之研究主題，以理解本主題研究領域之研究熱點為何。最後，透過共現詞分析，由視覺化的共現詞網路圖，可區分學習分析與教育資料探勘各有側重之研究重點。具體而言，學習分析著重於學習環境中學生學習行為、學習策略之探究。而教育資料探勘則多以演算法或如何選取資料特徵值以達到較好的預測結果為主要研究方向。易言之，其較偏向技術之探討，而非課室中之行為分析與學習支持。

#### 5. 研究限制與未來研究方向

本研究應用書目計量分析，其有別於質性的系統化文獻回顧。書目計量分析著重於分析研究文獻間之關係，而非深入探究個別文獻之內涵。也因為這樣的分析方式，才能同時探看較大數量之文獻所呈現之總體趨勢。然而，書目計量分析亦有其研究限制。如同系統化文獻回顧之常見限制，本研究之研究分析僅限於收錄於 Web of Science 資料庫中之文獻，無法就未發表或其它資料庫之文獻引用狀況進行分析；此外，未符合本研究所設定關鍵字篩選結果之文獻亦不在分析範圍內。未來研究可參考本研究之研究方法，使用其它關鍵字或設定不同的資料蒐集期間，以探討學習分析相關主題之研究現況與趨勢。

再者，本研究所使用之共現詞分析均使用英文關鍵字進行分析。以其它語系發表之潛在重要文獻將不在本研究之分析範圍中，未來研究亦可參考納入其它語系之文獻，並於進行相關語詞之資料清理與修正後進行分析，以獲得更為全面之研究概觀。

最後，限於篇幅，本研究僅就發表趨勢、核心文獻、引用關係與共現詞進行分析。未來研究可以作者為主題，探討共被引關係，或分析資料集外之引用關係，包括分析所蒐集之研究文獻於 Web of Science 資料庫中之引用狀況，或使用其它資料庫之索引資料進行分析，將可呈現更為多元之觀點，做為相關研究興趣者之參考。

#### 參考文獻

- Al-Shabandar, R., Hussain, A. J., Liatsis, P., & Keight, R. (2018). Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach. *IEEE Access*, 6, 73669-73685. <https://doi.org/10.1109/ACCESS.2018.2876755>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/https://doi.org/10.1016/j.joi.2017.08.007>
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194. <https://doi.org/10.1016/j.compedu.2017.05.007>

- Banihashem, S. K., Aliabadi, K., Pourroostaei Ardakani, S., Delaver, A., & Nili Ahmadabadi, M. (2018). Learning Analytics: A Systematic Literature Review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 9(2). <https://doi.org/10.5812/ijvlms.63024>
- Casquero, O., Ovelar, R., Romo, J., Benito, M., & Alberdi, M. (2016). Students' personal networks in virtual and personal learning environments: a case study in higher education using learning analytics approach. *Interactive Learning Environments*, 24(1), 49-67. <https://doi.org/10.1080/10494820.2013.817441>
- Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226-245. <https://doi.org/10.1109/TLT.2019.2916802>
- Palancı, A., Yılmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29(17), 22629-22650. <https://doi.org/10.1007/s10639-024-12737-5>
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157-167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 18, Article 103676. <https://doi.org/10.1016/j.compedu.2019.103676>
- Zacharoula, P., & Anastasios, A. E. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Journal of Educational Technology & Society*, 17(4), 49-64. <http://www.jstor.org/stable/jeductechsoci.17.4.49>