The following changes have been made on the Manuscript “**Factors Influencing Tandem Learning in Mathematics**” in accordance with reviewers’ comments

|  |  |  |
| --- | --- | --- |
| Reviewer’s comments | Changes made | Section |
| Reviewer 1: Major weaknesses |  |  |
| The decision to exclude conference papers (as written in Section 2.1) substantially undermines the comprehensiveness of the review. Prestigious conferences such as LAK, EDM, and AIED are central to this research community and feature rigorous, peer-reviewed work. Their omission results in a skewed and incomplete representation of the field, lacking many state-of-the-art publications. | After careful consideration and thorough screening, we agree with the reviewer’s observation regarding the importance of including conference papers. Consequently, we have expanded our review to incorporate relevant publications. | Methods |
| The analysis lacks depth and is too simple for a research paper. For instance, Figure 2 presents a word cloud featuring terms such as "machine learning" and "data", which is not surprising: of course these terms would be highly prominent. Furthermore, the study relies solely on basic descriptive statistics without performing more sophisticated analyses, such as correlations or cross-tabulations of various aspects, which could have enriched the findings. | Added authorship analysis, and some deeper descriptive analysis. | Results |
| Section 2.2 mentions that the authors developed their own software despite the existence of a suitable Python package. However, the rationale for this decision is not provided. The paper should specify the limitations of the existing tool and how the new implementation offers improvements or additional functionality. | Since we added collaboration network and country productivity plot, using pyBibX, we changed the paragraph to:  *The 285 WOS records were processed with custom Python scripts (Author 1, 2025) and the pyBibX library, which leverage WOS’s standardized fields—keywords, affiliations, citations—to construct co-authorship, keyword-co-occurrence, and citation networks. These scripts ensure reproducible extraction of publication years, citation counts, and geographic metadata* | Methods |
| Citation counts are presented without accounting for the year of publication, which introduces bias favouring older papers. Normalizing citation data or contextualizing it with publication dates is essential for meaningful comparisons. | *Added the following paragraph:*  *However, raw citation counts tend to favor older publications simply due to having more time to accumulate references. To address this, we normalized citation data by calculating the average number of citations per year since publication, enabling a more equitable comparison across years. The resulting trend, shown in Figure 7, reveals that more recent publications have a higher citation rate per year, suggesting not only an increase in publication volume but also a growing immediacy and relevance of research in this area.*    *Figure 7: Normalized average citations per year.* | Results |
| The end of Section 5 claims that the dataset is available, but no repository link is provided. For the sake of transparency and reproducibility, anonymized access to the included 184 studies should be offered at the peer review stage. | Doda prof. Istenič. V mail sem Vam dodal še dodatne datoteke. | Study registration and data availability |
| Reviewer 1: Minor weaknesses |  |  |
| Section 1 would benefit from the use of subheadings to guide the reader through its various components. | Added subsections “1.1. Predictive Models and Machine Learning in Education”, “1.2. The Value of Bibliometric Analysis”, “1.3. Research Aims”, and “1.4. Research Questions”. | Introduction |
| Typesetting is sub-standard and requires improvement, e.g., Table 1 spans a page break awkwardly. | Authors decided, that would be best to do after the final review process. | General observations |
| Bar charts would be clearer if absolute counts were included above each bar. | Added absolute counts. | Results |
| A few typos are present, such as “alst search” in Section 2.1. | Another proof-reading was conducted by independent researcher. | General observations |
| Reviewer 2 |  |  |
| The authors stated that they analyzed 33 review articles from WOS, SCOPUS, and Taylor & Francis databases. However, in the latter part of the section, they mentioned that 184 articles were analyzed to identify key research themes, publication trends, and methodological advancements. Why are there inconsistencies in the number of articles analyzed? | See major change below. | Abstract |
| The authors stated that they analyzed 184 articles from WOS, Scopus, and Taylor & Francis databases; however, from the PRISMA diagram, it was observed that the 184 articles were from only the WOS database. Thus, the above statement made in the abstract section is misleading. | See major change below. | Abstract |
| If the authors identified gaps such as underutilization of clustering and reinforcement learning techniques, why do they still focus on ML methods, which, per the review, have been widely used? | Clarified that those techniques have been identified as underutilized in the reviewed article, not in our review. | Introduction |
| Although the authors stated that the results from the university’s register are 785, it was missing in Table 1. This information should also be captured in the table, including the search string applied, date of search, and number of results. | Added the information in the table. | Methods |
| From the box labeled Records removed before screening in the PRISMA diagram, it was observed that Records not in WOS (n = 9051), which include records from SCOPUS, TAYLOR & FRANCIS and University Registers. Could the authors please clarify the rationale behind analyzing other databases if the focus was solely on records from WOS? What informed them to limit it to only WOS during the Identification phase? | See major change below. | Methods |
| Since the authors adopted bibliometric analysis, they should consider co-authorship analysis. | Added authorship analysis. | Results (+ Methods) |
| The authors should discuss the results in the context of the two research questions. They should provide a subsection for each research question and discuss it. | See major change below. | Discussion and conclusion |
| The authors stated that “the discrepancies between databases highlight the importance of selecting appropriate sources to ensure a balanced and accurate representation of the research landscape”. How does this relate to their work? Can they explain this further? | Elaborated further:  *The discrepancies between databases highlight the importance of selecting appropriate sources to ensure a balanced and accurate representation of the research landscape. In our case, although the initial search included SCOPUS, Taylor & Francis, and university registers, we ultimately relied on WOS for the bibliometric analysis due to its standardized metadata, compatibility with bibliometric tools, and lower risk of duplication, which allowed for a more consistent and reproducible analysis*. | Discussion and conclusion |
| The authors should separate the conclusion from the discussion and provide key findings from their work. | Separated the section, but did not add anything. | Discussion and conclusion |
| The authors claimed that the reliance on bibliometric data from only the selected databases may introduce selection bias. However, there are inconsistencies regarding whether the authors used databases from WOS, Scopus, Taylor & Francis, and University Registers, or only the WOS database. The authors must address these inconsistencies. It makes reading difficult. | See major change below. | Limitations and future directions |
| The authors should recheck their references especially in regard to the date of publication. In particular, the authors listed this research below as published in 2024. However, it was published in 2022 instead.  Lynam, S., Cachia, M., & Stock, R. (2024). An evaluation of the factors that influence academic success as defined by engaged students. Educational Review, 76(3), 586–604. https://doi.org/10.1080/00131911.2022.2052808 | That is a mislead by the reviewer. The paper was published online in 2022, but the citation was taken from “Educational Review Volume 76, 2024 - Issue 3”. | General observations |

Major change:

To resolve prior inconsistencies in the reported number of analyzed articles and clarify our database usage, we have restructured the methodology into three distinct and transparent analytical scopes, each answering a specific research question. First, we conducted a systematic review of 33 review articles retrieved from WOS, SCOPUS, and Taylor & Francis to identify overarching methodological trends (RQ1). Second, a qualitative synthesis of the 10 most cited WOS articles was performed to explore key algorithmic and interpretability practices (RQ2). Third, we carried out a bibliometric analysis based exclusively on 285 WOS articles, which were retained after database filtering and PRISMA screening (RQ3). By explicitly separating these three reviews—each with its own dataset, aim, and method—we now guide the reader clearly through our research questions, eliminate confusion over article counts, and transparently justify our final reliance on WOS data for bibliometric analysis due to its standardization and completeness.

Re-written abstract:

Numerous review studies have explored machine learning (ML)’s role in predicting academic success. In this paper, we perform three complementary analyses: (1) a systematic review of 33 review articles drawn from Web of Science (WOS), SCOPUS, and Taylor & Francis to identify prevailing methodologies and gaps; (2) a qualitative synthesis of the 10 most cited WOS articles to highlight influential algorithms and interpretability techniques; and (3) a bibliometric analysis of 285 WOS articles to chart publication trends, country-level productivity, and international collaboration networks. Our structured findings reveal an increasing focus on ensemble methods and explainable artificial intelligence (AI; e.g., SHAP, LIME), sustained growth in publication volume (2019–2024), and a leading output from China alongside strong collaboration hubs in Australia, India, and Europe. Persistent challenges include inconsistent terminology, data heterogeneity, and limited real-world deployment of ML models.