BIS HUMANITIES AND SOCIAL SCIENCE
Volume 2 2025 | DOI: 10.31603/bishss.378
6th Borobudur International Symposium on Humanities and Social Science (BIS-HSS) 2024
UNIMMA PRESS
Universitas Muhammadiyah Magelang

Promoting sustainable teacher education: Evaluating internship outcomes through clustering analysis

Nur Baiti Nasution[1*], Dwi Hartanto[2], Dewi Mardhiyana[1], Amalia Fitri[1], Dewi Azizah[1], Muhamad Najibufahmi[1], Rini Utami[1], Sayyidatul Karimah[1], Nurina Hidayah[1]

[1] Pekalongan University, Pekalongan, Indonesia
[2] Indonesian Artificial Intelligence Society, Jakarta, Indonesia
[*]Corresponding author's email: nurbaitinasution86@gmail.com

Abstract

This study explores a clustering-based approach to analyze the performance of preservice teachers during their three-month internship. The analysis is driven by data from multiple assessment sources, including mentor teachers, field assistant lecturers, and program committee evaluations. Using K-Means clustering, this study groups preservice teachers based on similarities in their performance components, aiming to provide meaningful insights for differentiated feedback and targeted support. The elbow method determined the optimal number of clusters to be four, capturing diverse patterns of performance profiles. Results show that each cluster presents distinct characteristics in terms of teaching practice, portfolio preparation, media development, and final evaluation scores. These findings highlight the potential of clustering to guide reflective practices and decision-making in teacher training programs.

Keywords

Preservice teacher internship, Performance clustering, K-means algorithm, Teacher education assessment

Published:
May 30, 2025

Selection and Peer-review under the responsibility of the 6th BIS-HSS 2024 Committee

Introduction

An internship is generally defined as a structured, supervised practical experience that allows students to apply theoretical knowledge in real-world settings. In teacher education, an internship is typically longer and more immersive than a brief practice teaching session, serving as a bridge between being a student and assuming full teaching responsibilities. It often involves a variety of field experiences under the guidance of experienced supervisors, enabling interns to engage deeply with the teaching profession and school culture [1], [2], [3].

Internships are a critical component of teacher education programs. They provide future teachers with hands-on opportunities to develop and refine essential teaching skills, such as lesson planning, classroom management, and instructional delivery. These experiences help student teachers build confidence, adapt to diverse classroom environments, and understand the complexities of school systems. Internships also foster professional growth by exposing interns to real-life challenges and encouraging reflective practice, which is vital for effective teaching [2], [3], [4], [5], [6].

The internship program is instrumental in shaping the professional competencies of future teachers. It allows student teachers to put theoretical knowledge into practice, receive constructive feedback, and identify areas for improvement. Through direct interaction with students and collaboration with mentor teachers, interns develop socio-affective skills, pedagogical competence, content knowledge, and personal qualities necessary for effective teaching. The program also helps interns understand their strengths and weaknesses, promoting continuous professional development [7].

Studies and reports indicate that internship programs in teacher education are typically implemented during the final semester of a four-year degree or towards the end of the program. The duration can vary but often spans several weeks to a full semester, providing ample time for immersive learning experiences. The process usually includes several stages: deployment to schools, observation of experienced teachers, coaching and mentoring, and independent teaching practice. Evaluation methods commonly involve feedback from mentor teachers, university supervisors, and self-reflection by interns. Key points highlighted in the evaluation process include the development of teaching skills, adaptability, professionalism, and the ability to manage classroom dynamics. However, some studies note challenges such as infrequent supervision, limited access to resources, and the need for more objective and differentiated assessment methods [8].

Considering its importance, the effectiveness of these programs is critical, as they shape the professional competencies and readiness of future educators. To ensure quality and continuous improvement, it is essential for faculties to regularly evaluate internship programs, not only to assess student outcomes but also to inform the design of preparatory activities and to determine the suitability of partner schools as internship hosts [9], [10].

Traditional evaluation methods often rely on subjective assessments, such as mentor feedback, self-assessment, and peer evaluation, each of which brings unique perspectives but also inherent limitations [11], [12]. The integration of data-driven approaches, such as educational data mining and clustering algorithms, offers new opportunities to enhance the objectivity and depth of internship program evaluation [13]. In particular, unsupervised learning techniques like K-means clustering can reveal hidden patterns in student performance data, supporting differentiated feedback and targeted program improvements [14], [15].

By leveraging a combination of traditional and advanced analytic methods, faculties can develop more robust frameworks for evaluating and refining teaching internship programs, ultimately supporting the development of highly effective educators.

Method

This study was conducted within the context of the teaching internship program at FKIP Universitas Pekalongan. A total of 109 students from three study programs—Pendidikan Matematika, Pendidikan Bahasa Inggris, and Pendidikan Bahasa dan Sastra Indonesia—participated in three-month internship program, during which they were required to: 1) Prepare lesson plans and instructional media, 2) Teach at least six times under mentor supervision, and 3) Compile a teaching portfolio. Each of these activities was assessed separately: mentor teachers evaluated teaching practice and learning media, while field assistant lecturers assessed the lesson plan portfolios. The final evaluation was an aggregate score from all components, including assessments by the internship committee.

Data were collected from multiple sources to comprehensively assess each participant's performance, including: Scores from Supervising Teacher (MA), Scores for Teaching Media (Med), Scores for Lesson Plans (Port), Pre-Internship Training Pretest (Pred), Peer Evaluation Scores (PA), Self-Assessment Scores (SA). K-Means clustering was used to group students based on their collected scores. Prior to clustering, the Elbow Method was applied to determine the optimal number of clusters. The quality of the clustering was evaluated using two metrics: the Silhouette Score and the Davies-Bouldin Index, which measure cluster compactness and separation.

To support interpretation of the clustering results, Principal Component Analysis (PCA) was conducted as a post-analysis step to reduce dimensionality and visualize the clusters in a two-dimensional space. All data processing and analysis procedures were performed using Python programming language in the Google Colab environment. After determining the optimal number of clusters, the characteristics of each cluster were analyzed in detail to interpret the performance profiles and identify strengths and weaknesses across key internship components.

Results and Discussion

Results

The elbow plot used in the clustering process displayed a visible bend at k=4 (Figure 1), suggesting that four clusters offered a balanced trade-off between data fit and model complexity. The silhouette scores for different numbers of clusters, ranging from 2 to 7. The highest silhouette score is observed when the data is grouped into 4 clusters, indicating that this number of clusters likely provides the best balance between cluster compactness and separation among clusters. The silhouette score achieved for 4 clusters was 0.46, and the corresponding Davies-Bouldin Index was 1.92. These values suggest that the clustering solution is reasonably well-structured and sufficiently separated.

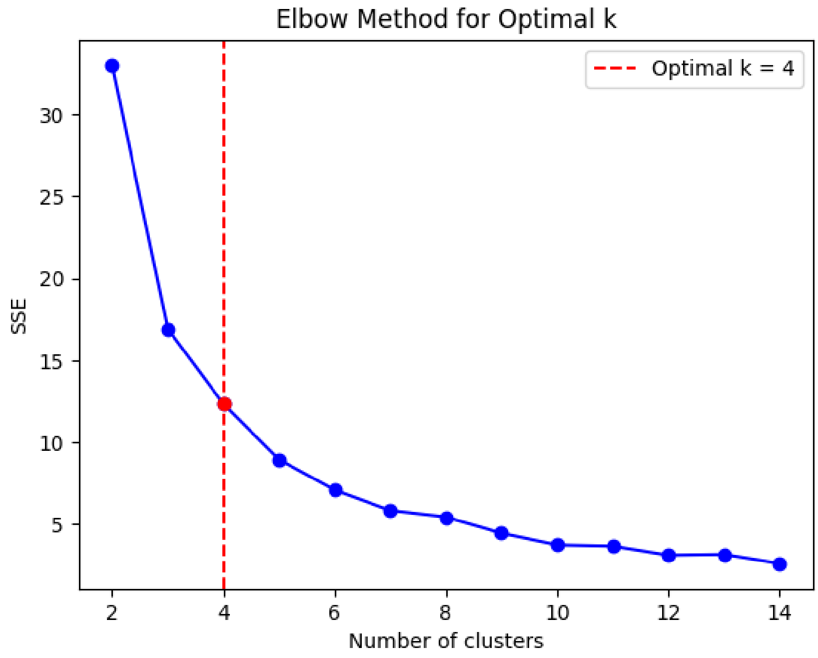

Figure 1. Result of the elbow method

Figure 2 shows the number of members in each cluster. Cluster 3 has the highest number of members (38), followed closely by Cluster 1 (35), and Cluster 0 (27), while Cluster 2 has the smallest membership with only 9 students. This indicates that Cluster 2 represents a relatively unique subgroup with specific characteristics, while the larger clusters might reflect more common performance profiles among preservice teachers.

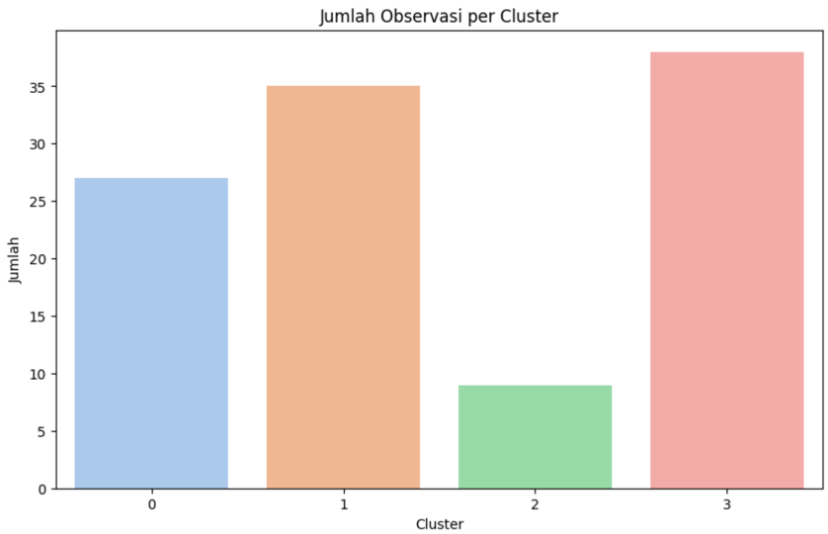

Figure 2. Numbers of observations in each cluster

Figure 3 is the scatter plot represents the clustering results projected onto two principal components after dimensionality reduction using Principal Component Analysis (PCA). Each point corresponds to a student, and the colors indicate membership to one of four clusters (clusters 0, 1, 2, and 3). PCA was applied to reduce data dimensionality and facilitate visual interpretation by plotting the data in only two dimensions.

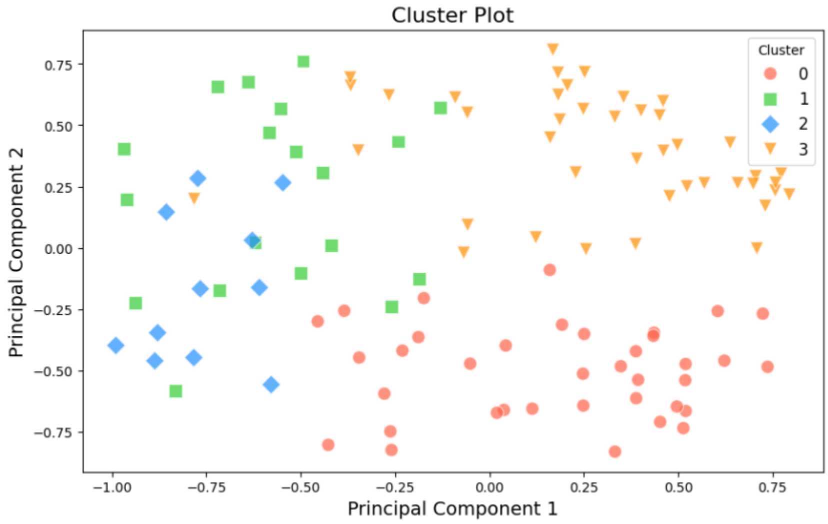

Figure 3. Clustering result

Each cluster reflected a distinct pattern of internship performance. Comparison of the characteristics for each cluster was shown in Figure 4. Below is the characteristics for each cluster:

1. **Cluster 1:** Low in all performance components. The bar chart below illustrates that this cluster consistently scores below average across all evaluated aspects. Among these, the peer assessment (PA) component is notably the lowest, indicating that students in this group not only struggle in formal evaluations but also receive the least favorable feedback from their peers.
2. **Cluster 2:** Displays a mixed performance profile. While this group scores positively on predictive assessment (Pred), self-assessment (SA), and peer assessment (PA), they show relatively lower scores in portfolio quality (Port), media development (Med), and particularly in mentor assessment (MA), which appears as the lowest component in this cluster. This suggests that although the students may demonstrate confidence and peer support, their performance in practical teaching components evaluated by mentors may require targeted improvement.
3. **Cluster 3:** Characterized by high performance in media development (Med), as reflected by the significantly above-average score in this component. Additionally, this cluster performs moderately well in mentor assessment (MA) and peer assessment (PA). However, their portfolio (Port) and predictive assessment (Pred) scores remain relatively low. This pattern suggests that students in this group may excel in designing instructional media but require support in documentation and planning aspects of their teaching practice.
4. **Cluster 4:** Demonstrates strong performance in portfolio preparation (Port) and mentor assessment (MA), with portfolio scores being the highest among all evaluated components. However, students in this cluster score relatively lower in

media development (Med), suggesting that while they are capable in planning and executing teaching activities, they may need further support in creating or utilizing instructional media. Other components such as predictive assessment (Pred), self-assessment (SA), and peer assessment (PA) fall around the average range.

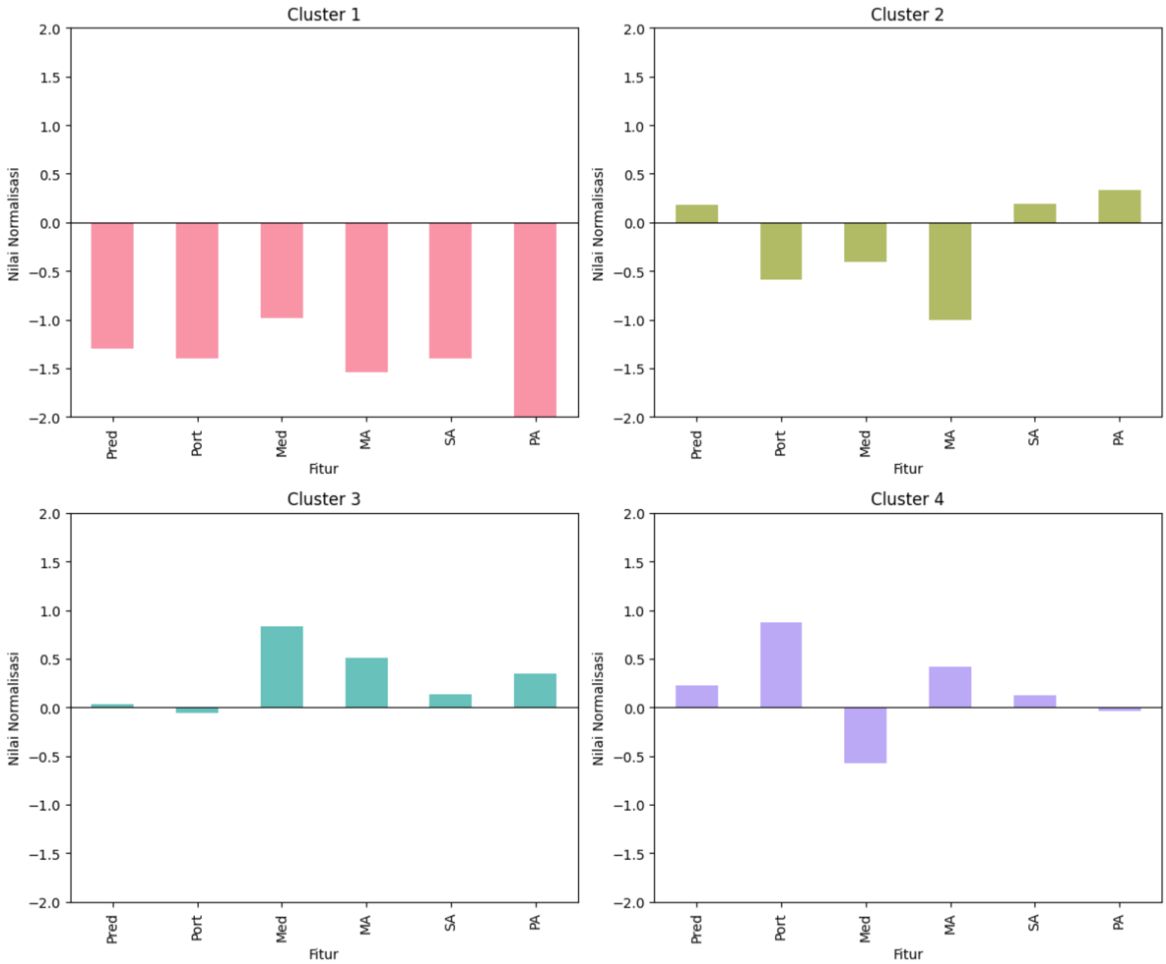

Figure 4. Characteristics of each clusters

Discussion

This study identified four distinct clusters of teaching internship performance, each characterized by unique strengths and areas for improvement across components such as predictive assessment, self-assessment, peer assessment, portfolio quality, media development, and mentor assessment. The findings highlight the diversity of pre-service teachers' competencies, with some excelling in reflective and peer-supported domains, while others demonstrate stronger practical or planning skills. These results align with previous research emphasizing the importance of differentiated support and targeted feedback in teacher education programs to address varied developmental needs among interns [16], [17], [18], [19].

The results of this study provide a deeper understanding of performance variations among preservice teachers during internships, which aligns with the call for data-driven approaches in personalized education [15]. The four identified clusters offer empirical evidence that student performance is not homogenous, reinforcing claims by Darling-

Hammond that effective teacher preparation must accommodate diverse developmental trajectories [9].

The finding that Cluster 1 scored consistently low across all components—including peer and self-assessments—supports earlier studies (e.g., [11]) suggesting that weak performers may lack both self-regulatory capacities and peer recognition. In contrast, Cluster 4's high portfolio and teaching evaluation scores align with research by Zeichner, which emphasizes the role of reflective planning in effective teaching [10].

Interestingly, the existence of Cluster 2, which scored highly in peer/self assessments but poorly in mentor evaluations, parallels findings by Topping, indicating that peer assessment may not always correlate with expert judgment [12]. This invites further inquiry into how mentorship feedback differs in focus from peer-based evaluations.

The clustering approach provided nuanced insights into student performance, supporting the value of data-driven methods for program evaluation and improvement [20], [21]. The study also reinforces the significance of comprehensive assessment strategies, including self, peer, and mentor evaluations, to capture a holistic view of teaching readiness [22], [23]. Furthermore, the results suggest that enhancing specific components—such as media development or portfolio preparation—could benefit particular groups of interns, echoing calls in the literature for tailored interventions and extended internship durations to maximize professional growth [16], [18], [19], [21]. The use of PCA for visualizing clustering results also echoes the methodological advances encouraged by Romero & Ventura, who advocate for learning analytics in evaluating educational interventions [13]. The effectiveness of silhouette and Davies-Bouldin scores as validation tools further supports their utility in educational data mining, echoing findings by Arbelaitz et al. [14].

Overall, this study not only confirms previous findings about the variability in student teaching performance but also extends them by offering a practical application of clustering for formative evaluation in teacher education programs.

Implications of the study

The findings of this study hold several implications for the design of internship preparation programs in teacher education. First, identifying performance-based clusters can inform differentiated pre-internship training. For instance, students projected to fall into Cluster 1 may benefit from intensified coaching in pedagogical planning and classroom confidence-building strategies. For prospective Cluster 3 students, greater emphasis on documentation and portfolio development should be incorporated into preparatory workshops.

Second, the clustering outcomes suggest the need for integrating peer feedback and self-assessment activities early in the teacher preparation curriculum to improve students' evaluative judgment. This approach aligns with growing advocacy for reflective practice and metacognitive skill development in preservice teacher education.

Finally, this study underscores the value of aligning mentor training with the clustering framework so that mentoring strategies can be better adapted to student needs. A more structured mentor orientation could support differentiated observation protocols and targeted feedback, particularly for students falling into less balanced profiles (e.g., Cluster 2 or 4). Thus, clustering-based analysis serves not only as an evaluative tool but also as a strategic planning guide for scaffolding future internship experiences.

Conclusion

This study concludes that clustering analysis provides a powerful lens for understanding the varied performance profiles of preservice teachers during their internships. The identification of four distinct clusters reveals meaningful differences in teaching ability, self and peer reflection, media development, and portfolio quality—dimensions that are crucial in evaluating professional readiness. These findings correspond closely to the study's initial objective of designing differentiated support mechanisms based on actual field performance.

Furthermore, this research pushes the boundary of current practices by demonstrating how clustering can inform both instructional planning and mentor feedback systems. These results pave the way for future research that could incorporate predictive models or longitudinal tracking of teacher development. In summary, this work not only enriches theoretical perspectives on preservice teacher evaluation but also offers practical, scalable tools for program designers. Future research is encouraged to validate the clustering model across diverse contexts and explore integration into real-time educational platforms to further enhance adaptive learning and teaching supervision systems.

References

[1] C. Kosnik and C. Beck, "The internship component of a teacher education program: Opportunities for learning," *The Teacher Educator*, vol. 39, pp. 18–34, 2003, doi: 10.1080/08878730309555327.

[2] S. Khajuria, "Perception of M.ed. Student-teachers Towards Internship as a Part of Teacher Training Programme," *International Journal For Multidisciplinary Research*, p., 2024, doi: 10.36948/ijfmr.2024.v06i01.14216.

[3] D. Kumari and R. Babu, "CONSTRUCTION AND VALIDATION OF INTERNSHIP PROGRAMME INTEREST INVENTORY," vol. 58, pp. 5777–5780, 2021, doi: 10.17762/PAE.V58I1.2213.

[4] H. Boholano, B. E. Jamon, B. Corpuz, and R. Bacaltos, "A Responsive Internship Program from the Viewpoint of Beginning Teachers," *Journal International Inspire Education Technology*, p., 2024, doi: 10.55849/jiiet.v3i2.641.

[5] S. Parveen and N. Mirza, "Internship Program in Education: Effectiveness, Problems and Prospects," *International Journal of Learning and Development*, vol. 2, no. 1, Jan. 2012, doi: 10.5296/ijld.v2i1.1471.

[6] S. Murtiningsih, R. N. Swastika, E. Puspitasari, and A. W. D. Putri, "Being an effective English teacher through internship: Voices from the involved parties," *Englisia: Journal of Language, Education, and Humanities*, p., 2024, doi: 10.22373/ej.v11i2.20447.

[7] Yudhister, "A Study of Attitude of Student Teacher towards Internship Programme in B.Ed Course," *International Journal for Research Publication and Seminar*, p., 2024, doi: 10.36676/jrps.v15.i2.1625.

[8] N. D. S. Areekkuzhıyıl, "A Phenomenological Study on the Lived Experiences of Student Teachers during their School Internship," *International Journal of Research*, vol. 11, pp. 18–32, 2020, [Online].

[9] L. Darling-Hammond, *Powerful Teacher Education: Lessons from Exemplary Programs*. Jossey-Bass,

2006.
[10] K. M. Zeichner, “Rethinking the connections between campus courses and field experiences in college- and university-based teacher education,” *J Teach Educ*, vol. 61, no. 1–2, pp. 89–99, 2010.
[11] E. Panadero and A. Jonsson, “The use of scoring rubrics for formative assessment purposes revisited: A review,” *Educ Res Rev*, vol. 9, pp. 129–144, 2013, doi: https://doi.org/10.1016/j.edurev.2013.01.002.
[12] K. J. Topping, “Peer assessment,” *Theory Pract*, vol. 48, no. 1, pp. 20–27, 2009.
[13] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, 2010.
[14] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, “An extensive comparative study of cluster validity indices,” *Pattern Recognit*, vol. 46, no. 1, pp. 243–256, 2013.
[15] C. A. Tomlinson, *The differentiated classroom: Responding to the needs of all learners*. Alexandria, USA: ASCD, 2014.
[16] A. C. K. Cheung, K. L. Wong, H. F. Wang, and J. B. Dai, “Effect of a student teaching internship program on the self-efficacy of pre-service teachers in rural China,” *International Journal of Educational Management*, vol. 37, no. 2, pp. 373–392, Jan. 2023, doi: 10.1108/IJEM-03-2021-0081.
[17] J. Lobo, “Pre-service teachers’ evaluation of the student internship program: Basis for design improvement,” *Journal for Educators, Teachers and Trainers*, vol. 13, no. 5, Jan. 2022, doi: 10.47750/jett.2022.13.05.005.
[18] S. Li, “Research on the Effectiveness of Educational Internship——A Case Study on Primary Education Major,” vol. 4, p., 2020, doi: 10.18686/ahe.v4i9.2629.
[19] A. Bastian, J. König, J. Weyers, H.-S. Siller, and G. Kaiser, “Effects of teaching internships on preservice teachers’ noticing in secondary mathematics education,” *Front Educ (Lausanne)*, p., 2024, doi: 10.3389/feduc.2024.1360315.
[20] X. Feng, Y. Chang, and Y. Wang, “The implementation of evidence-based medicine and the Problem-Based Learning (PBL) teaching approach in the clinical practice education of the gastroenterology department,” *Region - Educational Research and Reviews*, p., 2024, doi: 10.32629/rerr.v5i6.1547.
[21] A. R. S. Tuasikal, S. Hartoto, B. Prakoso, D. Kartiko, and A. Hariyanto, “The analysis on teaching skills and learning effectiveness of internship students,” *Jurnal Cakrawala Pendidikan*, p., 2021, doi: 10.21831/CP.V40I3.40466.
[22] J. T. Lobo, “An inquiry on the effectiveness of the teaching internship program based on pre-service teachers’ appraisal,” *Journal of Research, Policy & Practice of Teachers & Teacher Education*, p., 2023, doi: 10.37134/jrpptte.vol13.1.2.2023.
[23] K. Michos, A. Cantieni, R. Schmid, L. Müller, and D. Petko, “Examining the relationship between internship experiences, teaching enthusiasm, and teacher self-efficacy when using a mobile portfolio app,” *Teach Teach Educ*, vol. 109, p. 103570, 2022.