JPPIPA 12(5) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Application of the Observational Rating Scale to PGMI Student Activity in Science Analysis Discussion

Destiana Puanda Ashari[1*], Ikhlasul Ardi Nugroho[1], Anita Ekantini[2], Uun Febriyani[3], Yanti Kumala Sembiring[4], Waode Sitti Fatimah[1], Annisa Rifah Rahman[1]

[1] Faculty of Education, Yogyakarta State University, Yogyakarta, Indonesia.
[2] Faculty of Tarbiyah and Teacher Training, UIN Sunan Kalijaga, Yogyakarta, Indonesia.
[3] Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia.
[4] Graduate School, Yogyakarta State University, Yogyakarta, Indonesia.

Received: November 24, 2025
Revised: April 22, 2026
Accepted: June 07, 2026
Published: June 07, 2026

Corresponding Author:
Destiana Puanda Ashari
destianapuanda.2024@student.uny.ac.id

DOI: 10.29303/jppipa.v12i5.13581

Abstract: This study aims to develop, validate, and examine the practicality of a rating-scale–based observation instrument designed to assess the activeness of PGMI students during science question analysis discussions. A quantitative descriptive approach was employed, involving 40 fifth-semester PGMI students. The instrument was developed based on learning activity theory and a multidimensional student engagement framework, consisting of 9 indicators and 27 items rated on a 1–4 scale. Content validity was evaluated by three experts using Aiken's V index. At the same time, instrument practicality was assessed by five trained observers acting as instrument users through a user satisfaction questionnaire based on a 1–5 Likert scale. The results indicated very high content validity, with an average Aiken's V value of 0.97, and excellent practicality, reflected by a mean user satisfaction score of 4.36 out of 5. The application of the instrument revealed that student activeness fell into the high category, with an average activity score of 89.70 (SD = 11.22), and 90% of students were classified as active to very active. These findings demonstrate that the developed observation instrument is valid and practical for evaluating student activity in discussion-based learning, particularly within teacher education contexts.

Keywords: Discussion-based learning; Observation instrument; PGMI students; Rating scale; Student engagement

Introduction

Improving the quality of higher education requires systematic evaluation of various learning components, one of which is students' active participation (Assefa et al., 2025; Bergdahl et al., 2024; Godsk & Møller, 2024; Heilporn et al., 2024; Johar et al., 2023; Waqar et al., 2025). Within the context of active learning, student activeness is not merely defined by physical attendance, but is reflected in behavioral engagement during discussions, cognitive engagement through expressing and processing ideas, and affective engagement manifested in interest, persistence, and commitment throughout the learning process. Numerous studies indicate that engagement across these three dimensions contributes significantly to academic achievement and the development of 21st-century competencies (Alshammari & Alrashidi, 2024; Bowden et al., 2021; Destiny & Lator, 2022; Heilporn et al., 2021; Laranjeira & Teixeira, 2024; Pearson, 2024; Safitri et al., 2024; Xu et al., 2023). Therefore, measuring student activeness requires instruments capable of capturing the complexity of learning engagement accurately and consistently, so that lecturers have a strong empirical basis for improving instructional practices.

How to Cite:
Ashari, D. P., Nugroho, I. A., Ekantini, A., Febriyani, U., Sembiring, Y. K., Fatimah, W. S., & Rahman, A. R. (2026). Application of the Observational Rating Scale to PGMI Student Activity in Science Analysis Discussion. *Jurnal Penelitian Pendidikan IPA*, *12*(5), 729–737. https://doi.org/10.29303/jppipa.v12i5.13581

The urgency of measuring student activeness becomes increasingly prominent when learning activities involve higher-order thinking processes, such as discussions focused on analyzing evaluation questions. Activities involving question analysis and discussion are essential components of teacher education, as they train students to understand the quality of assessment instruments, identify weaknesses in test items, and design improvements based on principles of validity and measurability. Learning through question-analysis discussions has been shown to deepen conceptual understanding while enhancing students' analytical and reflective skills (Ketonen & Nieminen, 2023; Shafii & Berger, 2025; Steenkamp & Brink, 2024; Xia & Xu, 2024). For students in the Madrasah Ibtidaiyah Teacher Education Program (PGMI), these activities are directly relevant to the professional competencies required of future teachers, particularly in designing and evaluating science assessment instruments. Nevertheless, the quality of question-analysis discussions is highly dependent on students' levels of activeness during learning, highlighting the need for measurement tools that can capture student engagement in a more specific and diagnostic manner.

Observation instruments based on rating scales are considered more appropriate for measuring student activeness in the context of question-analysis discussions than dichotomous yes/no instruments. Rating scales allow observers to assess student behaviors gradually and sensitively across varying intensities of behavioral, cognitive, and affective engagement (Buntins et al., 2021; Dhama et al., 2023; Fahruddin et al., 2024; Kim et al., 2022; Kohake, 2024; Papageorgiou et al., 2025; Pettersen et al., 2023; Setiawan et al., 2022; Siddiqi et al., 2021; Tannoubi et al., 2023). Through this approach, activities such as posing analytical questions, presenting concept-based arguments, responding to peers' opinions, and demonstrating persistence during discussions can be represented more comprehensively. Consequently, rating scales provide a more holistic depiction of students' participation patterns in higher-order thinking activities (Berlin et al., 2025; Fatih et al., 2024; Hendrowati et al., 2025; Irdalisa et al., 2024; Jia et al., 2024; Lestari et al., 2024; Pramitasari et al., 2025).

Previous studies have developed observational instruments to measure student participation or engagement (Ariyanti et al., 2022; Buntins et al., 2021; Hastomo & Septiyana, 2022; Jia et al., 2024; Qi et al., 2023; Sharif-Nia et al., 2024; Wolf et al., 2024). However, these instruments are generally designed for general learning contexts or for primary and secondary education levels, with indicators that are not specifically aligned with the higher-order cognitive demands of question-analysis discussions. Moreover, observational instruments that explicitly integrate behavioral, cognitive, and affective dimensions within a single measurement framework tailored to the curriculum of teacher education—particularly in the context of science assessment analysis—remain limited. This condition indicates the need for developing more contextual and relevant observational instruments that align with the characteristics of learning in PGMI programs.

The novelty of this study lies in the development of a rating scale–based observational instrument specifically designed to measure student activeness in science question-analysis discussions by integrating behavioral, cognitive, and affective indicators within a single observational framework. The development of the instrument indicators is grounded in active learning theory, which emphasizes student engagement through thinking, questioning, discussing, and reflecting (Fatih et al., 2024; Safitri et al., 2024), and is operationalized into observable behaviors during discussion activities. A multidimensional student engagement framework is employed to ensure that the instrument assesses not only the intensity of participation but also the quality of students' thinking processes and affective engagement, which are essential to the professional demands of future teachers.

Based on this background, the present study aims to develop a rating scale–based observational instrument designed to assess the activeness of students in the Madrasah Ibtidaiyah Teacher Education Program (PGMI) during science question-analysis discussions, as well as to examine the content validity and practicality of the developed instrument through expert evaluation and user trials. In addition, this study seeks to describe students' levels of activeness based on the measurement results obtained using the observational instrument within the learning context under investigation.

Method

Time and Place of Research

This study was conducted in the Elementary School Teacher Education Program (Pendidikan Guru Madrasah Ibtidaiyah/PGMI), Faculty of Tarbiyah and Educational Sciences, State Islamic University (UIN) Sunan Kalijaga Yogyakarta. The research was carried out from September to October 2025 and consisted of several main stages, including expert validation of the instrument, observer training, and the implementation of student activeness observation. The expert validation of the instrument was conducted in early September 2025. Observer training was carried out for two days in mid-October 2025. The main implementation of the study, namely the observation of student activeness, took place on October 24, 2025, coinciding with a discussion session on the analysis of science test instruments in the Learning Evaluation course. Therefore, October 24, 2025, represents the core observation phase, while the other stages functioned as preparatory and instrument development phases. The sequence of research stages is presented in Figure 1.

Figure 1. Research procedure flowchart

Research Design

This study employed a quantitative approach with a descriptive research design. This design was selected to systematically describe the level of student activeness during discussions on the analysis of science test instruments using an observation instrument based on a rating scale. In addition, the study aimed to describe the distribution patterns and characteristics of student activeness based on behavioral, cognitive, and affective indicators that emerged during the discussion process.

Research Subjects and Sampling Technique

The research subjects consisted of 40 fifth-semester students of the PGMI Study Program at UIN Sunan Kalijaga Yogyakarta who were enrolled in the Learning Evaluation course during the odd semester of the 2025/2026 academic year. The sampling technique used was purposive sampling, based on the consideration that all participating students met the research criteria, namely: being active fifth-semester students, officially enrolled in and attending the Learning Evaluation course, having completed the Science Learning for MI and Learning Strategies courses, and having no prior formal teaching experience in schools. A total of 40 students met these criteria, and all were included as research subjects.

Data Collection Techniques

Data were collected through direct observation of student activities during group discussions. The observations were conducted simultaneously by five observers who had previously undergone training to ensure more objective and consistent data collection. Observer training lasted for two days and included an introduction to the concepts of student engagement and active learning, an explanation of the indicators and behavioral descriptors in the observation instrument, observation simulations using recorded videos of student discussions, and discussions aimed at establishing shared understanding among observers. In addition to direct observation, the discussion sessions were video-recorded to support data verification and rechecking of the observation results.

Research Instrument

The research instrument was a student activeness observation instrument based on a rating scale, developed by adapting several theoretical frameworks, including motivation- and scaffolding-based learning engagement theory, classifications of learning activities encompassing behavioral, cognitive, and emotional aspects, and a multidimensional student engagement framework. The adaptation process resulted in nine indicators of student activeness: active participation, explanatory ability, relevance of arguments, collaboration in discussion, perseverance, consistency, initiative, leadership, and the ability to connect theory with practice.

Each indicator was elaborated into three observable behavioral descriptors, resulting in a total of 27 observation items. The instrument employed a 1–4 rating scale with the categories very inactive, inactive, active, and very active. The maximum score that could be obtained by each student was 108. In this study, student activeness was defined as the level of behavioral, cognitive, and affective engagement demonstrated by students during the discussion on test item analysis, as represented by the total score of the observation instrument, ranging from 27 to 108.

Instrument Validity and Practicality Testing

Content validity of the instrument was examined through expert judgment by three senior lecturers from UIN Sunan Kalijaga Yogyakarta with expertise in learning evaluation and science education. Validity analysis was conducted using Aiken's V index, with a minimum criterion of V ≥ 0.70 at a significance level of α = 0.05.

The practicality test of the instrument was conducted by five observers who evaluated the ease of use of the instrument based on five aspects: clarity of completion instructions, clarity of observation items, clarity of the rating scale, instrument layout format, and overall ease of use. The assessment used a 1–5 scale, with score interpretations ranging from very difficult to use to very easy to use. This study focused on testing content validity and instrument practicality. Inter-rater reliability testing was not conducted and is acknowledged as a limitation of the study, which may be addressed in future research.

Research Procedure

Observations were conducted while students engaged in group discussions consisting of eight members per group. Each group was asked to analyze seven multiple-choice science test items for fifth-grade elementary students, covering aspects of content validity, item difficulty level, discrimination index, and distractor quality. The discussion activities lasted for 90 minutes in a single session, and all student activities during the discussion were observed using the developed instrument.

Data Analysis Techniques

Observation data were analyzed using descriptive statistics, including the calculation of mean, median, mode, standard deviation, and frequency distribution. Student activeness scores were categorized into four levels: very inactive, inactive, active, and very active. Categorization was performed using the ideal interval method by proportionally dividing the total score range. This categorization was determined as a methodological decision by the researcher to facilitate the interpretation of student activeness levels within the context of active learning.

Result and Discussion

The content validity of the instrument was examined using Aiken's V index, with a minimum acceptable criterion of V ≥ 0.70 at a significance level of 0.05 for three validators using five relevance categories (Aiken, 1985). This approach was employed to ensure that each observation item was appropriately aligned with the construct of student activeness being measured.

The analysis results indicate that out of a total of 27 observation items, 25 items (92.6%) achieved the maximum Aiken's V value (V = 1.00), reflecting complete agreement among validators regarding the relevance and clarity of the items. The remaining two items obtained Aiken's V values ranging from 0.75 to 0.92, which still exceeded the established minimum validity threshold. These differences were primarily attributable to minor variations in expert judgments concerning sentence clarity rather than the substantive indicators being measured. Accordingly, all items were deemed valid and retained for use in the data collection stage.

Overall, these findings indicate that the instrument demonstrates very strong content validity, with a high level of expert agreement regarding indicator relevance, alignment with measurement objectives, and the clarity of observable and measurable behavioral descriptors. This result is consistent with principles of instrument development that emphasize the importance of content validity as a foundational aspect of observational instrument quality (Sawu et al., 2023; Weriza et al., 2025).

After establishing content validity, the practicality of the instrument was evaluated through a user satisfaction test involving five observers. The assessment employed a five-point Likert scale (1–5) across five aspects: clarity of usage instructions, clarity of observation items, clarity of the rating scale, instrument layout format, and overall ease of use in the context of classroom observation. Higher scores indicated greater satisfaction and ease of use.

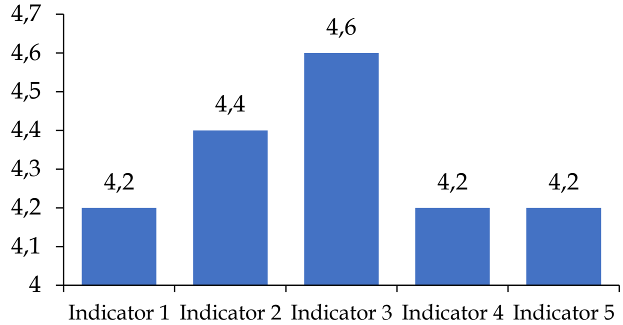

Figure 2. User satisfaction on each indicator

Based on Figure 2 it can be seen that the rating scale indicator received the highest average score (4.6), indicating that observers were highly satisfied with the rating scale because it was easy to understand and effectively differentiated levels of student activeness. The Observation Items and Layout Format indicators each received an average score of 4.2, suggesting that the items were considered representative in capturing variations in student activeness and that the instrument format was perceived as neat, systematic, and easy to complete during the observation process.

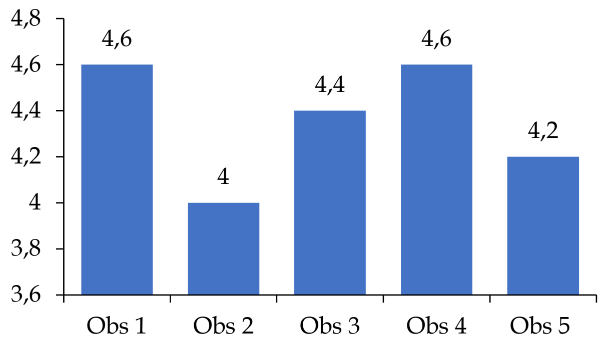

Figure 3. User satisfaction with each observer

User satisfaction results by individual observer (Figure 3) demonstrate a relatively consistent scoring pattern, with no extreme differences among observers. The overall average satisfaction score of 4.36 out of a maximum score of 5.0 indicates that the instrument possesses a very high level of practicality.

These findings are consistent with the study by Sholichin et al. (2022), which reported that practical and user-friendly instruments tend to be more readily accepted by users and support the collection of valid data. A high level of instrument practicality also reduces observers' cognitive load during observation, allowing them to focus more on accurately recording student behavior rather than interpreting the instrument itself. Such practicality is a crucial aspect of observational research, as it contributes to the consistency and accuracy of the collected data (Laranjeira & Teixeira, 2024).

To enhance measurement consistency, the five observers involved in this study participated in an intensive two-day training session (15–16 October 2025). The training included familiarization with activeness indicators, simulations of instrument use, and discussions aimed at aligning assessment perceptions. Nevertheless, this study did not quantitatively calculate inter-rater reliability using statistical coefficients such as Cohen's Kappa or Fleiss' Kappa. This represents a methodological limitation that should be addressed in future research.

Descriptive analysis was conducted on student activeness scores and activeness categories based on observation results from group discussion-based learning activities focused on science assessment item analysis. Observations were carried out with 40 PGMI students who actively participated in group discussions and class presentations.

Table 1. Descriptive statistics of research variables

Variable	N valid	N missing	Mean	Median	Std. Deviation	Min	Max
Activeness	40	1	89.70	92.00	11.22	58.00	106.00
Category	40	1	2.30	2.00	0.65	1.00	3.00

Source: Primary data processed 2025

Table 1 shows that student activeness scores ranged from a minimum of 58.00 to a maximum of 106.00, with a mean score of 89.70 and a median of 92.00. The relatively close proximity of the mean and median values indicates a fairly balanced distribution around the central tendency. A standard deviation of 11.22 suggests a moderate level of variation in student activeness, which remains within an acceptable range.

For the activeness category variable, the mean value of 2.30 indicates that, overall, students were classified within the Active to Very Active categories. A median value of 2.00 suggests that the central tendency lies within the Active category, while a standard deviation of 0.65 reflects low variability among students.

The frequency distribution of activeness scores presented in table 2 indicates that most students achieved high scores, particularly in the ranges of 90–99 (42.50%) and 100–106 (20.00%). In contrast, only a small proportion of students fell into the lower score ranges, namely 58–69 (7.50%) and 70–79 (5.00%). This pattern suggests that the majority of students demonstrated relatively high levels of activeness during discussion-based learning.

Table 2. Distribution of frequency of student activity scores by value group

Value groups	Frequency	Percentage (%)	Cumulative percentage (%)
58 – 69	3	7.50	7.50
70 – 79	2	5.00	12.50
80 – 89	10	25.00	37.50
90 – 99	17	42.50	80.00
100 – 106	8	20.00	100.00
Total	40	100.00	

Source: Primary data processed 2025

Furthermore, Table 3 shows that 50.00% of students were classified as Active and 40.00% as Very Active, while only 10.00% were categorized as Less Active. Thus, a total of 90% of students were classified as Active or Very Active. The empirically derived category ranges were Less Active (58–69), Active (70–94), and Very Active (95–106), with category determination based on the equal-interval principle commonly applied in educational measurement.

The findings of this study indicate that PGMI students exhibited relatively high levels of activeness during discussion-based learning activities involving science assessment item analysis. Although the descriptive research design does not permit causal inference, the observation results suggest that the discussion-based learning format facilitates the emergence of various activeness indicators, including questioning, expressing opinions, responding, collaborating, and critical thinking.

Table 3. Frequency distribution of student activity categories

Category	Frequency	Percentage (%)	Valid percentage (%)	Cumulative percentage (%)
Less Active	4	9.80	10.00	10.00
Active	20	48.80	50.00	60.00
Highly Active	16	39.00	40.00	100.00
Total Valid	40	97.60	100.00	
Missing	1	2.40		
Total	41	100.00		

Source: Primary data processed 2025

Figure 4. Students engage in group discussions while analyzing science questions

As illustrated in Figure 4, students were actively engaged in both small-group discussions and plenary presentations. Group discussions provided a safer environment for students to share ideas and express opinions, while class presentations encouraged the development of communication skills, critical thinking, and self-confidence. The combination of these activities created a learning dynamic that supported student engagement across behavioral, affective, and cognitive dimensions.

These findings are consistent with previous studies reporting that discussion-based learning and guided inquiry approaches enhance active participation, self-confidence, and critical thinking among students. In addition, the use of a structured observation instrument enabled more objective and systematic identification of student participation patterns.

Nevertheless, the distribution of activeness categories indicates that the proportion of students classified as Active (50%) was higher than those classified as Very Active (40%). The relatively high mean score (89.70), which approaches the threshold for the Very Active category, suggests that a subset of students with very high scores contributed to elevating the overall mean, while most students remained within moderate to high levels of activeness. This variation is reflected in the standard deviation, indicating that student activeness levels were not entirely homogeneous.

Approximately 10% of students classified as Less Active require particular attention. Observer notes indicate that students in this group tended to be more passive and primarily assumed the role of listeners. This finding reinforces the view that student engagement is influenced by individual characteristics and contextual learning conditions. In this context, the developed observation instrument can serve as a diagnostic tool for lecturers to identify students who require additional support, such as scaffolding, strategic grouping, or adjustments to discussion task complexity.

Furthermore, the findings suggest that PGMI students, as prospective elementary school teachers, tend to demonstrate higher engagement when learning activities are directly related to their future professional tasks. Science assessment item analysis requires the integration of conceptual understanding and pedagogical reasoning, thereby encouraging more meaningful participation. This aligns with studies emphasizing the effectiveness of authentic and reflective learning in the professional development of pre-service teachers.

This study has several limitations, including the limited sample size and scope involving only one class within a single institution, a relatively short observation duration, the absence of quantitative inter-rater reliability analysis, and the use of descriptively defined activeness category thresholds. Therefore, generalization of the findings should be approached with caution.

Future research is recommended to involve larger and more diverse samples, employ longitudinal designs, calculate inter-rater reliability using statistical methods, and integrate quantitative and qualitative approaches. The development of a digital version of the observation instrument and the application of experimental research designs may also serve as valuable directions for future studies to strengthen empirical contributions to the study of student activeness.

Conclusion

The results indicate that the developed rating scale–based observation instrument meets the criteria of content validity and practicality based on expert judgment and user trials, and is therefore suitable for measuring the activeness of PGMI students during science question analysis discussions. The application of the instrument shows that the level of student activeness falls into the high category within the learning context examined, without implying the effectiveness of the discussion method. A limitation of this study is the absence of inter-rater reliability analysis; therefore, future research is recommended to incorporate inter-rater reliability testing and to develop a digital version of the instrument to enhance the efficiency of observation and data management.

Acknowledgments
Thank you to the expert validators from Yogyakarta State University and observers who have provided valuable support and contributions to the implementation of this research.

Author Contributions
Conceptualization, methodology, investigation, data curation, writing—original draft preparation, D.P.A.; resources, A.E.; writing—review and editing, Y.K.S., W.S.F., and A.R.R.; formal analysis, Y.K.S. and U.F.; validation, I.A.N. and W.S.F.; visualization, A.R.R.; supervision, project administration, I.A.N. All authors have read and agreed to the published version of the manuscript.

Funding
There is no external funding for this research.

Conflicts of Interest
The author states that there is no conflict of interest.

References

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Alshammari, S., & Alrashidi, O. (2024). The Effect of Students' Engagement on Their Learning Achievement in EFL Online Courses: *International Journal of Online Pedagogy and Course Design*, *14*, 1–18. https://doi.org/10.4018/IJOPCD.357875

Ariyanti, I., Fikrie, F., & Hariyono, D. (2022). Students' Engagement dalam Proses Pembelajaran Daring Melalui Lesson Study pada Mata Kuliah Kalkulus Integral. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *6*(1), 824–836. https://doi.org/10.31004/cendekia.v6i1.1287

Assefa, Y., Tilwani, S. A., Moges, B. T., & Majeed, H. (2025). Unpacking Student Engagement and Its Mediating Role in Predicting the Relationship between Self-Determination and Academic Satisfaction Among Undergraduate Students in Higher Education. *Social Sciences & Humanities Open*, *12*, 101797. https://doi.org/10.1016/j.ssaho.2025.101797

Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking Student Engagement in Higher Education Learning Analytics: A Systematic Review. *International Journal of Educational Technology in Higher Education*, *21*(1), 63. https://doi.org/10.1186/s41239-024-00493-y

Berlin, I., Nugroho, A., Patonah, S., & Hayat, M. (2025). Profile of Learning Outcomes and Student Activity in Learning Natural Sciences at SMPN 2 Karangawen. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 112–118. https://doi.org/10.29303/jppipa.v11i1.9808

Bowden, J. L.-H., Tickle, L., & Naumann, K. (2021). The Four Pillars of Tertiary Student Engagement and Success: A Holistic Measurement Approach. *Studies in Higher Education*, *46*(6), 1207–1224. https://doi.org/10.1080/03075079.2019.1672647

Buntins, K., Kerres, M., & Heinemann, A. (2021). A Scoping Review of Research Instruments for Measuring Student Engagement: In Need for Convergence. *International Journal of Educational Research Open*, *2*, 100099. https://doi.org/10.1016/j.ijedro.2021.100099

Destiny, E., & Lator, C. (2022). Curriculum Design and Student Social Engagement in Human Kinetics Programmes: Empirical Evidence from Nigerian Universities. *International Journal of Multidisciplinary Research and Growth Evaluation*, *3*, 897–909. https://doi.org/10.54660/.IJMRGE.2022.3.6.897-909

Dhama, K., Sharun, K., Gugjoo, M. B., Tiwari, R., Alagawany, M., Yatoo, M. I., … Farag, M. R.. (2023). A Comprehensive Review on Chemical Profile and Pharmacological Activities of Ocimum basilicum. *Food Reviews International*, *39*(1), 119–147. https://doi.org/10.1080/87559129.2021.1900230

Fahruddin, F., Nurgiansah, T., Setiawan, V., & Saefudin, A. (2024). Quantitative Measures of Engagement in History Classes: Analyzing the Efficacy of Interactive Pedagogies. *SAR Journal - Science and Research*, 241–248. https://doi.org/10.18421/SAR73-10

Fatih, M., Alfi, C., & Muqtafa, M. (2024). Science Learning Game (SLG) Based on Augmented Reality Enhances Science Literacy and Critical Thinking Students Skills. *Jurnal Penelitian Pendidikan IPA*, *10*(2), 973–981. https://doi.org/10.29303/jppipa.v10i2.6107

Godsk, M., & Møller, K. (2024). Engaging Students in Higher Education with Educational Technology. *Education and Information Technologies*, *30*, 2941–2976. https://doi.org/10.1007/s10639-024-12901-x

Hastomo, T., & Septiyana, L. (2022). The Investigation of Students' Engagement in Online Class During Pandemic Covid-19. *Jurnal Penelitian Ilmu Pendidikan*, *15*(2). https://doi.org/10.21831/jpipfip.v15i2.49512

Heilporn, G., Lakhal, S., & Bélisle, M. (2021). An Examination of Teachers' Strategies to Foster Student Engagement in Blended Learning in Higher Education. *International Journal of Educational Technology in Higher Education*, *18*(1), 25. https://doi.org/10.1186/s41239-021-00260-3

Heilporn, G., Raynault, A., & Frenette, É. (2024). Student Engagement in a Higher Education Course: A Multidimensional Scale for Different Course Modalities. *Social Sciences & Humanities Open*, *9*, 100794. https://doi.org/10.1016/j.ssaho.2023.100794

Hendrowati, T., Badrun, M., Siswoyo, S., & Istiani, A. (2025). The Impact of Hybrid Learning on Student Engagement and Academic Performance in Post-Pandemic Science Education. *Jurnal Penelitian Pendidikan IPA*, *11*(4), 154–165. https://doi.org/10.29303/jppipa.v11i4.10701

Irdalisa, I., Akbar, B., Marjan, T., Maesaroh, M., & Kartikawati, E. (2024). Ricosre Model with Question Formulation Technique (QFT): Enhancing Students' Higher Order Thinking Skills (HOTS) and Science LiteracyRicosre Model With Question Formulation Technique (QFT): Enhancing Students' Higher Order Thinking Skills (HOTS) and Science. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1175–1178. https://doi.org/10.29303/jppipa.v10i3.6764

Jia, Y., Spagnolo, A., Barrett, N., Murphy, A., Basto, P., Rothpletz-Puglia, P., & Luther, S. (2024). Validation of a Peer Observation and Evaluation Tool for Online Teaching in the U.S. *Educational Technology Research and Development*, *73*. https://doi.org/10.1007/s11423-024-10428-z

Johar, N. A., Kew, S., Tasir, Z., & Koh, E. (2023). Learning Analytics on Student Engagement to Enhance Students' Learning Performance: A Systematic Review. *Sustainability*, *15*, 7849. https://doi.org/10.3390/su15107849

Ketonen, L., & Nieminen, J. (2023). Supporting Student Teachers' Reflection Through Assessment: The Case of Reflective Podcasts. *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2023.104039

Kim, S., Westine, C., Wu, T., & Maher, D. (2022). Validation of the Higher Education Student Engagement Scale in use for Program Evaluation. *Journal of College Student Retention: Research, Theory & Practice*, *26*, 769–793. https://doi.org/10.1177/15210251221120908

Kohake, K. (2024). Systematic Observation to Measure Teaching Quality in Different Contexts: Insights from Science Lessons, Physical Education Lessons, and Sports Training Using the Classroom Assessment Scoring System. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1437996

Laranjeira, M., & Teixeira, M. (2024). Relationships between Engagement, Achievement and Well-Being: Validation of the Engagement in Higher Education Scale. *Studies in Higher Education*, *50*, 1–15. https://doi.org/10.1080/03075079.2024.2354903

Lestari, H., Rahmawati, Y., & Usman, H. (2024). STEM-PjBL Learning Model to Enhance Critical Thinking Skills of Students on Magnets, Electricity, and TechnologySTEM-PjBL Learning Model to Enhance Critical Thinking Skills of Students on Magnets, Electricity, and Technology. *Jurnal Penelitian Pendidikan IPA*(8), *10*(8), 6027–6037. https://doi.org/10.29303/jppipa.v10i8.8153

Papageorgiou, E., Wong, J., Liu, Q., Khalil, M., & Cabo, A. (2025). A Systematic Review on Student Engagement in Undergraduate Mathematics: Conceptualization, Measurement, and Learning Outcomes. *Educational Psychology Review*, *37*. https://doi.org/10.1007/s10648-025-10046-y

Pearson, W. (2024). Affective, Behavioural, and Cognitive Engagement with Written Feedback on Second Language Writing: A Systematic Methodological Review. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1285954

Pettersen, E., Ertesvåg, S., Pöysä, S., Vaaland, G., & Virtanen, T. (2023). Students' Situational Engagement and Its Association with Overall Engagement: The Application of the InSitu Instrument in the Context of a Norwegian Lower Secondary School. *Scandinavian Journal of Educational Research*, *68*, 1–14. https://doi.org/10.1080/00313831.2023.2175245

Pramitasari, N., Hastuti, P. W., Tyas, R., Anjarsari, P., & Roektiningroem, E. (2025). The Influence of Inquiry Learning Models Containing STEAM to Improve the Science Process Skills. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 11109–11120. https://doi.org/10.29303/jppipa.v10i12.8985

Qi, M., Liu, M., Yan, T., & Liu, B. (2023). Analysis of Students' Learning Activities Based on the ICAP Framework. *ICETM '22: Proceedings of the 2022 5th International Conference on Education Technology*

Management, 305 - 309. https://doi.org/10.1145/3582580.3582633

Safitri, W., Suyanto, S., & Prasetya, W. (2024). Influence of the STEM-Based Engineering Design Process Model on High School Students' Creative and Critical Thinking Abilities. *Jurnal Penelitian Pendidikan IPA*, *10*(2), 662–673. https://doi.org/10.29303/jppipa.v10i2.4765

Sawu, M., Sukarso, A., Ayu, T., & Handayani, B. (2023). Effect of STEM Learning in Building Creative Dispositions and Creative Thinking Skills of Junior High School Students. *Jurnal Penelitian Pendidikan IPA*, *9*(8), 6219–6229. https://doi.org/10.29303/jppipa.v9i8.4180

Setiawan, D., Hiroyuki, A., Syamsunarno, M. R. AA., Hartady, T., Lubis, A., & Widyastuti, R. (2022). Kondisi Fertilitas Mencit Jantan yang Diberi Ekstrak Etanol Akar Alang-alang (Imperata cylindrica). *Acta Veterinaria Indonesiana*, *10*(2), 142–147. https://doi.org/10.29244/avi.10.2.142-147

Shafii, R., & Berger, J.-L. (2025). Teacher Assessment Literacy, Formative Assessment Practices, and Their Perceived Efficacy in Tanzania: A Scoping Review. *Studies in Educational Evaluation*, *86*, 101496. https://doi.org/10.1016/j.stueduc.2025.101496

Sharif-Nia, H., Maroco, J., She, L., Narula, S., Ma, L., Kaveh, O., & Ghahrani, N. (2024). University Student Engagement Inventory: Validation in the Indian Online Learning Context. *Measurement and Evaluation in Counseling and Development*, *57*. https://doi.org/10.1080/07481756.2023.2301287

Sholichin, M., Razak, A., Lufri, L., & Irdawati, I. (2022). Validitas dan Praktikalitas E-Modul Berbasis Mobile Learning Berbantuan 3D Page Flip Professional pada Materi Ekologi dan Perubahan Lingkungan di Kelas X SMA. *Jurnal Penelitian Pendidikan IPA*, *8*(6), 3034–3043. https://doi.org/10.29303/jppipa.v8i6.2467

Siddiqi, A., Shabbir, M., Abbas, M., Mahmood, A., & Salman, R. (2021). Developing and Testing Student Engagement Scale for Higher Educational Students. *Journal of Applied Research in Higher Education*, *ahead-of-print*. https://doi.org/10.1108/JARHE-11-2020-0388

Steenkamp, G., & Brink, S. (2024). Students' Experiences of Peer Learning in an Accounting Research Module: Discussion Forums, Peer Review and Group Work. *The International Journal of Management Education*, *22*, 101057. https://doi.org/10.1016/j.ijme.2024.101057

Tannoubi, A., Quansah, F., Jnr, J. H., Srem-Sai, M., Bonsaksen, T., Chalghaf, N., Boussayala, G., Azaiez, C., Snani, H., & Azaiez, F. (2023). Adaptation and Validation of the Arabic Version of the University Student Engagement Inventory (A-USEI) Among Sport and Physical Education Students. *Psych*, *5*, 320–335. https://doi.org/10.3390/psych5020022

Waqar, Z., Sultana, S., & Ata, M. (2025). In-Class Active Engagement Strategies to Enhance Student Participation During Lectures : A Systematic Review. *Journal of Bahria University Medical and Dental College*, *15*, 146–158. https://doi.org/10.51985/JBUMDC2024486

Weriza, J., Ismail, I., Yustisia, H., Zulkifli, Z., & Asnur, L. (2025). Enhancing Student Participation through a Validated Cooperative Flipped Classroom Framework in Technology Education. *Jurnal Penelitian Pendidikan IPA*, *11*(7), 134–141. https://doi.org/10.29303/jppipa.v11i7.11445

Wolf, S., Jukes, M., Yoshikawa, H., Stern, J., & Kim, S. (2024). Examining the Validity of an Observational Tool of Classroom Support for Children's Engagement in Learning. *Early Childhood Education Journal*, *53*, 1325–1339. https://doi.org/10.1007/s10643-024-01731-8

Xia, J., & Xu, C. (2024). EFL Students' Critical Thinking in Asynchronous Online Discussions: A Perspective of Peer Facilitation. *Thinking Skills and Creativity*, *53*, 101581. https://doi.org/10.1016/j.tsc.2024.101581

Xu, B., Stephens, J., & Lee, K. (2023). Assessing Student Engagement in Collaborative Learning: Development and Validation of New Measure in China. *The Asia-Pacific Education Researcher*, *33*. https://doi.org/10.1007/s40299-023-00737-x