Digital-Based Supervision of 21st-Century Skills: An Integrated Evaluation–Supervision Model Using E-MESp 4Cs Mobile

Alya Dwi Arianty[1*], Eny Winaryati[1], Andari Puji Astuti[1]

[1] Chemistry Education, Universitas Muhammadiyah Semarang, Semarang, Indonesia.

Abstract: This study examines the implementation of a digital supervision system based on 21st-century skills using the E-MESp 4Cs Mobile platform as an integrated evaluation–supervision model in primary and secondary schools. The research employed an applied research design within a Research, Development, and Diffusion (RD&D) framework, with a specific focus on the implementation evaluation stage. The study involved teachers, principals, and supervisors from Muhammadiyah schools across four regions, comprising an initial pilot phase in Semarang City and an expanded implementation phase in Semarang Regency, Kendal Regency, and Salatiga City. Supervision was conducted through the integration of three core components: lesson planning, instructional implementation, and student learning products, aligned with the competencies of critical thinking, creativity, collaboration, and communication. The findings indicate a recurring gap between documented instructional planning and classroom practices, while instructional implementation and student learning products based on implementation evaluation data generally demonstrated stronger performance across regions. These results suggest that effective instructional practices are not always adequately represented in formal planning documents. The study concludes that digital, data-based supervision can provide a more comprehensive and reflective representation of instructional quality compared to conventional administrative supervision. Within the scope of this implementation-focused study, the E-MESp 4Cs Mobile platform shows potential to support professional development and evidence-based supervisory practices in 21st-century education.

Keywords: 21st-century skills; Digital supervision; Evidence-based supervision; Instructional evaluation; Teacher professional development

Introduction

Twenty-first-century education demands a paradigm shift from the mere acquisition of knowledge toward the development of complex competencies that encompass higher-order thinking skills, creativity, collaboration, and communication. Global educational frameworks consistently emphasize that critical thinking, creativity, collaboration, and communication (4Cs) constitute core competencies that must be systematically integrated into school learning processes to prepare learners for increasingly complex social, economic, and technological challenges (Griffin et al.,

2012; Voogt et al., 2013; OECD, 2018; OECD, 2019; OECD, 2023; UNESCO, 2021). Recent studies in science education further confirm that 21st-century competencies are increasingly used as indicators of instructional quality and learning effectiveness (Susanti, 2025; Saputra et al., 2025).

Despite the extensive development of conceptual frameworks supporting 4Cs-based learning, empirical evidence indicates that their implementation at the school level remains inconsistent and challenging. Teachers frequently experience difficulties in translating 4Cs competencies into instructional planning that is operational, measurable, and aligned with classroom

realities, resulting in gaps between intended learning objectives and enacted instructional practices (Fullan & Langworthy, 2014; Darling-Hammond et al., 2017; Guskey, 2014; Widoyoko, 2017). Studies published in national accredited journals also report that instructional documents often fail to capture students' actual critical and creative thinking performance (Saputra et al., 2025; Sudirman et al., 2025). This condition suggests that exclusive reliance on planning documents may lead to partial representations of instructional quality (Hill et al., 2013; Slameto, 2014; Mulyasa, 2015).

Instructional supervision plays a strategic role in addressing this misalignment between instructional planning and classroom implementation. Effective supervision is increasingly conceptualized not merely as an administrative compliance mechanism, but as a continuous professional learning process characterized by reflection, pedagogical dialogue, and data-informed feedback (Timperley, 2007; Kraft & Gilmour, 2016; Korthagen, 2017). Empirical findings from JPPIPA indicate that principal-led academic supervision contributes to improvements in teachers' pedagogical competence when it moves beyond document checking toward instructional observation and feedback (Windriati et al., 2025). Nevertheless, supervisory practices in many school contexts remain predominantly administrative, emphasizing document completeness rather than instructional processes and learning outcomes (Suyanto & Jihad, 2013; Zamroni, 2016).

From an instructional evaluation perspective, integrated approaches that simultaneously assess instructional planning, instructional implementation, and student learning products are considered more capable of providing a comprehensive and valid representation of instructional quality. Research on formative and programmatic assessment highlights the importance of examining instructional processes alongside learning outcomes within a continuous evaluation framework oriented toward professional improvement (Black & Wiliam, 1998; Shute, 2008; Van der Vleuten et al., 2012). Recent instrument-development studies published in JPPIPA further emphasize the importance of valid and reliable assessment tools to capture students' higher-order thinking and learning outcomes accurately (Ernasari et al., 2025; Saputra et al., 2025).

Advancements in digital technology provide significant opportunities to strengthen instructional supervision and evaluation practices. The use of digital platforms enables systematic data collection, timely feedback, and evidence-based instructional decision-making (Earl & Timperley, 2008; Pellegrino & Quellmalz, 2010; Means et al., 2014). Evidence from JPPIPA demonstrates that digitally supported evaluation instruments and learning platforms enhance teacher reflection, instructional effectiveness, and student learning engagement (Hidayat & Saputra, 2025; Purba et al., 2025; Rahma et al., 2025). International frameworks also emphasize that the integration of digital technologies and artificial intelligence in education should be ethically grounded and pedagogically oriented, rather than merely serving administrative efficiency (Redecker, 2017; Holmes et al., 2019; OECD, 2023; UNESCO, 2021).

In this context, digital-based instructional supervision emerges as a relevant and necessary approach for strengthening data-informed professional development and instructional quality. Empirical evidence indicates that technology-supported supervision encourages deeper teacher reflection, improves the quality of supervisory feedback, and supports collaborative professional learning communities (Boud & Molloy, 2013; Darling-Hammond et al., 2017; Trust et al., 2016; Tondeur et al., 2017; Zhang et al., 2016). Studies published in JPPIPA consistently highlight the growing relevance of validated digital instruments, STEM-based learning products, and inquiry-oriented modules in improving instructional evaluation practices (Sari et al., 2025; Rahma et al., 2025; Purba et al., 2025).

Based on this background, the present study examines the implementation of 4Cs-based instructional supervision using the E-MESp 4Cs Mobile platform as an integrated evaluation–supervision system. The novelty of this study lies in its integrated digital supervision approach, which simultaneously examines instructional planning, instructional implementation, and student learning products within a single evaluation framework, rather than relying solely on administrative document review. By building upon recent findings reported in JPPIPA, including the development and validation of the E-MESp 4Cs Mobile instrument (Winaryati et al., 2025), this study aims to provide empirical evidence on how digital supervision can reveal instructional alignment gaps and pedagogical strengths that are not fully captured through conventional supervision practices. Accordingly, this study contributes to the growing body of research on data-based instructional supervision aligned with the demands of twenty-first-century education instructional quality, rather than as purely technical solutions (Holmes et al., 2019; OECD, 2019).

Method

Research Design

This study employed an applied research design using the Research, Development, and Diffusion (RD&D) approach, in which the Research stage involved

needs analysis of 4Cs-based instructional supervision practices, the Development stage focused on the design and validation of supervision rubrics embedded in the digital system, and the present study specifically emphasized the implementation evaluation within the Diffusion stage. The RD&D approach was selected because the study not only involved the development of an innovative educational product, but also examined its application in authentic educational settings to generate evidence-based recommendations for improvement (Griffin et al., 2012; Earl & Timperley, 2008). The primary focus of this research was on evaluating the implementation of 4Cs-based instructional supervision, rather than measuring long-term impacts on student learning outcomes.
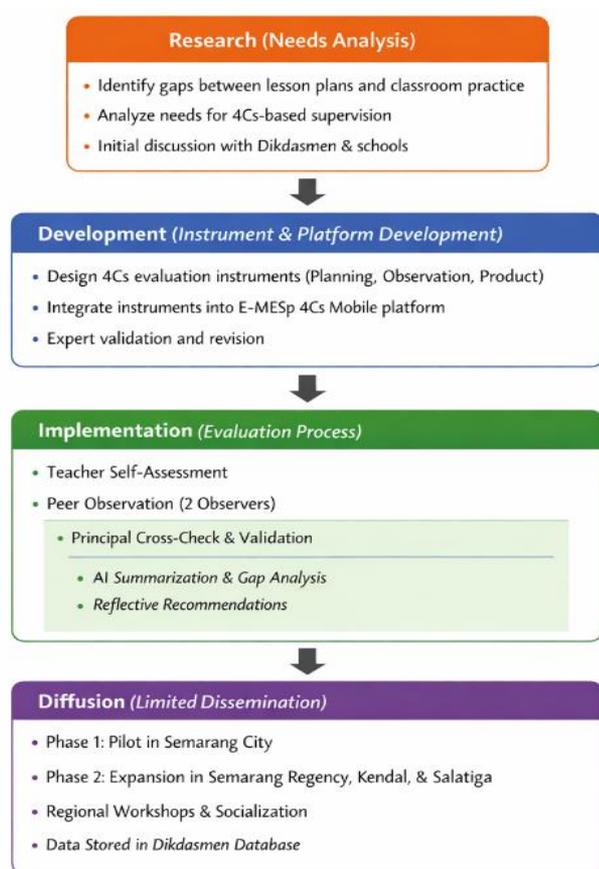
Flowchart of the Research, Development, and Diffusion (RD&D) process of the E-MESp 4Cs Mobile system focused on implementation evaluation.

Figure 1. Flowchart of the research, development, and diffusion (RD&D) process of the E-MESp 4Cs Mobile system focused on implementation evaluation

Research Context and Participants

The study was conducted in Muhammadiyah educational institutions across four regions serving as the evaluation context. The initial pilot implementation (Phase 1) was carried out in Semarang City, followed by an expanded implementation (Phase 2) in Semarang Regency, Kendal Regency, and Salatiga City. Research participants included subject teachers, school principals, and educational supervisors from elementary, junior high, and senior high schools who were directly involved in the instructional supervision process using the E-MESp 4Cs Mobile platform. Participants were selected through purposive sampling to represent variations in educational levels and school contexts, thereby enabling a comprehensive understanding of the implementation process (Suyanto & Jihad, 2013; Mulyasa, 2015).

Research Instruments and Platform

The primary research instrument consisted of standardized 4Cs-based supervision rubrics and observation sheets, while the E-MESp 4Cs Mobile platform functioned as a digital medium to facilitate data collection, documentation, and aggregation. The instruments integrate three core components of instructional supervision: instructional planning (antecedent), instructional implementation (transaction), and student learning products (interim product), all aligned with the indicators of 4Cs competencies. This approach is consistent with the principles of programmatic assessment and process-oriented instructional evaluation, which emphasize the integration of planning, practice, and learning outcomes within a coherent evaluation framework (Van der Vleuten et al., 2012; Black & Wiliam, 1998).

Data Collection Procedures

Data collection was conducted through several stages. First, teachers conducted structured self-assessment by uploading instructional planning documents to the platform for evaluation at the antecedent stage. Second, classroom instruction was observed by two human observers, consisting of peer teachers and/or supervisors, using the same 4Cs-based indicators of instructional implementation. When substantial discrepancies emerged between self-assessment and peer-observer evaluations, school principals assumed a mediating role to conduct cross-checking based on aggregated data. Third, student learning products were assessed as representations of interim achievements of 4Cs competencies using predefined rubrics. In addition to quantitative data in the form of evaluation scores for each component, qualitative data were also collected, including reflective notes and supervisory feedback to support contextual analysis (Timperley, 2007; Boud & Molloy, 2013).

Data Analysis Techniques

Quantitative data were analyzed using descriptive-comparative techniques by comparing mean evaluation

scores across antecedent, transaction, and interim product components to identify alignment patterns and implementation gaps across research regions. This analysis was used to map patterns of 4Cs-based instructional supervision implementation across different school contexts. Meanwhile, qualitative data were analyzed using thematic analysis to reveal implementation-related issues, challenges, and good practices in the application of digital instructional supervision (Miles et al., 2014; Widoyoko, 2017).

Trustworthiness and Research Ethics

Data trustworthiness was ensured through source triangulation by comparing teachers' self-assessment results, peer-observer evaluations, and principals' validation records. System-generated summaries functioned solely as decision-support tools to assist reflection and professional development planning, while all evaluative judgments remained under human authority. This approach was applied to enhance the credibility of the research findings and to minimize subjective bias (Creswell & Poth, 2018). From an ethical perspective, all research participants were involved voluntarily, and the collected data were used solely for the purpose of improving instructional quality and educational supervision, in accordance with principles of ethical and responsible use of educational technology (UNESCO, 2021).

Result and Discussion

Results
Overview of Implementation Phases

This study reports the results of implementing the E-MESp 4Cs Mobile integrated evaluation–supervision system across two sequential phases. Phase 1 focused on school-level pilot implementation in Semarang City, while Phase 2 involved expanded implementation across Semarang Regency, Kendal Regency, and Salatiga City. Instructional supervision was conducted through three integrated components: Antecedent (instructional planning), Transaction (instructional implementation), and Interim Product (student learning products).

Phase 1 Results: School-Level Findings in Semarang City

Phase 1 involved four schools representing different educational levels (SMA, SMP, SMK). Table 1 presents mean scores (0–4 scale) from PRA (self-assessment & peer assessment) and classroom observation, allowing direct identification of implementation gaps.

Table 1. Phase 1 Results per School (Semarang City)

School	Antecedent (PRA → Obs)	Transaction (PRA → Obs)	Interim Product (PRA → Obs)	Dominant Pattern
SMA Muhammadiyah 1 Semarang	4.00 → 3.59	3.97 → 3.98	3.98 → 3.97	Strong planning, early execution gap
SMP Muhammadiyah 6 Semarang	3.89 → 3.90	3.95 → 3.96	3.98 → 4.00	High consistency
SMK Muhammadiyah 1 Semarang	2.27 → 3.70	3.65 → 3.76	2.27 → 3.76	Implementation exceeds planning
SMP Muhammadiyah 3 Semarang	3.84 → 3.95	3.90 → 3.95	3.97 → 4.00	Adaptive pedagogy

Figure 2. Documentation of phase 1 instructional supervision process in Semarang City

These results show that three of four schools demonstrated higher scores in Transaction and Interim Product compared to Antecedent, indicating that classroom practice and student outcomes often exceeded what was formally documented in lesson plans.

Phase 1 Regional Mean (Semarang City)

Table 2. Mean Scores Phase 1 (Semarang City)

Component	PRA Mean	Observation Mean	Gap Direction
Antecedent	3.50	3.51	Stable
Transaction	3.87	3.91	↑
Interim Product	3.55	3.93	↑ Significant

Phase 2 Results: Expanded Implementation Across Regions

Phase 2 involved ten schools across three regions, enabling cross-regional comparison and validation of Phase 1 patterns.

Table 3. Phase 2 Results per School (Semarang Regency)

School	Antecedent (PRA → Obs)	Transaction (PRA → Obs)	Interim Product (PRA → Obs)
MIS Muhammadiyah Suruh	3.70 → 3.72	3.70 → 3.75	3.26 → 3.80
SMK Muhammadiyah Sumowono	4.00 → 3.96	3.97 → 3.97	3.95 → 4.00
SMP Muhammadiyah Suruh	4.00 → 3.99	3.97 → 3.97	3.97 → 4.00

Table 4. Phase 2 Results per School (Kendal Regency)

School	Antecedent (PRA → Obs)	Transaction (PRA → Obs)	Interim Product (PRA → Obs)
SD Muhammadiyah Purin	3.52 → 3.55	3.48 → 3.58	3.50 → 3.47
SMK Muhammadiyah 1 Weleri	2.67 → 3.83	3.84 → 3.85	3.85 → 3.80
SMP Muhammadiyah 4 Sukorejo	3.63 → 3.51	3.49 → 3.59	3.27 → 3.53

Table 5. Phase 2 Results per School (Salatiga City)

School	Antecedent (PRA → Obs)	Transaction (PRA → Obs)	Interim Product (PRA → Obs)
SD Muhammadiyah Plus	3.89 → 3.90	3.88 → 3.92	3.87 → 4.00
SMA Muhammadiyah Plus	4.00 → 3.88	3.85 → 3.75	3.98 → 3.75
SMP Muhammadiyah Plus	4.00 → 3.88	3.84 → 3.83	3.88 → 4.00
SMP Muhammadiyah Salatiga	4.00 → 3.98	3.96 → 4.00	4.00 → 4.00

Table 6. Regional Mean Scores (Phase 2)

Region	Antecedent Mean	Transaction Mean	Interim Product Mean
Semarang Regency	3.90	3.93	3.93
Kendal Regency	3.63	3.67	3.61
Salatiga City	3.91	3.88	3.94

Figure 3. Documentation of phase 2 instructional supervision process

Discussion

The results consistently demonstrate a recurring implementation gap between instructional planning (Antecedent) and classroom-based outcomes (Transaction and Interim Product). Importantly, this gap is not unidirectional. While some schools showed a decline from planning to execution (e.g., SMA Muhammadiyah Plus Salatiga), the dominant pattern across regions indicates that student learning products frequently exceeded planned expectations, particularly in Kendal and Semarang Regency.

This phenomenon can be explained through the concept of tacit pedagogical knowledge, where teachers possess practical instructional competencies that are not fully articulated in formal planning documents. Teachers in schools such as SMK Muhammadiyah 1 Weleri and SMP Muhammadiyah 4 Sukorejo demonstrated strong classroom execution despite weak or conservative planning scores, suggesting that administrative documents underestimate actual pedagogical capacity.

From an evaluation perspective, these findings reinforce arguments in programmatic and formative assessment literature that instructional quality cannot be validly inferred from planning documents alone (Black & Wiliam, 1998; Van der Vleuten et al., 2012). The E-MESp 4Cs Mobile platform enabled triangulation between self-assessment, peer observation, and principal validation, allowing discrepancies to be identified and discussed constructively.

At the governance level, the integration of AI-supported summaries and dashboards provided school principals and Dikdasmen authorities with aggregated, non-punitive evidence to support professional development decisions. This positions digital supervision not merely as an evaluative tool, but as a data-informed leadership instrument that supports reflective practice, peer mentoring, and policy alignment across schools.

By merging instructional findings with organizational and policy analysis, the results demonstrate that digital, integrated supervision can shift school culture from compliance-oriented administration toward reflective, evidence-based professional learning.

Conclusion

This study concludes that the implementation of 4Cs-based instructional supervision using the E-MESp 4Cs Mobile platform enables a more comprehensive, process-oriented, and data-driven representation of instructional quality compared to conventional supervision approaches that emphasize administrative document compliance alone. By integrating the

evaluation of instructional planning (antecedent), classroom implementation (transaction), and student learning products (interim product), the supervision system is able to empirically identify recurring implementation gaps between written lesson plans and enacted teaching practices, while simultaneously revealing teachers' tacit pedagogical knowledge that is often not fully articulated in formal planning documents. Across Phase 1 and Phase 2 implementations involving multiple school levels and regions, quantitative findings consistently show that instructional implementation and student learning products frequently achieve higher mean scores than instructional planning, indicating that active, adaptive teaching practices can effectively support the development of 21st-century competencies even when administrative readiness appears moderate. However, this study is limited to an implementation evaluation focus and does not measure long-term impacts on student learning outcomes, relies on descriptive–comparative analysis rather than inferential statistical testing, and is contextually bounded to Muhammadiyah schools, which may limit broader generalization. Therefore, future research is recommended to extend the supervision model through longitudinal designs, strengthen analytic rigor by incorporating inferential or mixed-method approaches, and test the platform across more diverse institutional contexts. At the practical level, findings suggest concrete follow-up actions, including the standardization of 4Cs assessment rubrics, structured peer-supervision calibration, and the systematic use of AI-generated analytic summaries to support principals' instructional decision-making, positioning digital supervision not as a control mechanism but as a sustainable professional learning and instructional governance tool aligned with the demands of 21st-century education.

Author Contributions
Conceptualization, A.D.A. and E.W.; methodology, A.D.A. and E.W.; software, E.W.; validation, E.W. and A.P.A.; formal analysis, A.D.A.; investigation, A.D.A.; resources, E.W.; data curation, A.D.A.; writing—original draft preparation, A.D.A.; writing—review and editing, E.W. and A.P.A.; visualization, A.D.A.; supervision, E.W. and A.P.A.; project administration, E.W.; funding acquisition, E.W. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The involvement of supervisors as co-authors reflects their academic supervision and scholarly contribution and does not constitute a conflict of interest.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Boud, D., & Molloy, E. (2013). *Feedback in higher and professional education: Understanding it and doing it well*. London: Routledge.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute.

Datnow, A., & Park, V. (2018). *Professional collaboration with purpose: Teacher learning for equitable and excellent schools*. New York: Routledge.

Dede, C. (2014). The role of digital technologies in deeper learning. *Students at the Center: Deeper Learning Research Series*. Boston: Jobs for the Future.

Earl, L. M., & Timperley, H. (2008). *Professional learning conversations: Challenges in using evidence for improvement*. Netherlands: Springer.

Ernasari, Dewi, N. M., Meilina, I. L., & Faresta, R. A. (2025). Development of an Assessment Instrument for Understanding Physics Concepts and Nationalism Attitudes of Learners on Newton's Laws Material. *Jurnal Penelitian Pendidikan IPA, 11*(11), 279–285. https://doi.org/10.29303/jppipa.v11i11.12735

Fullan, M., & Langworthy, M. (2014). *A rich seam: How new pedagogies find deep learning*. London: Pearson.

Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Netherlands: Springer.

Guskey, T. R. (2014). Planning professional learning. *Educational Leadership, 71*(8), 10–16. Retrieved from https://govboard.tusd1.org/Portals/TUSD1/GovBoard/docs/packets/packet09-23-14/9-23-14-BAI7-PD-PlanningProfessionalLearning.pdf

Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement. *Educational Management Administration & Leadership, 38*(6), 654–678. https://doi.org/10.1177/1741143210379060

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York: Teachers College Press.

Hidayat, S., & Saputra, R. D. A. (2025). Validation of a Google Sites-Based LMS Evaluation Instrument with the Importance Performance Analysis Approach in Sustainable Learning. *Jurnal Penelitian Pendidikan IPA, 11*(11), 317–322. https://doi.org/10.29303/jppipa.v11i11.12901

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research. *Educational Researcher, 42*(9), 476–487. https://doi.org/10.3102/0013189X13512674

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Boston: Center for Curriculum Redesign.

Korthagen, F. A. J. (2017). Inconvenient truths about teacher learning. *Teachers and Teaching, 23*(4), 387–405. https://doi.org/10.1080/13540602.2016.1211523

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development? *Educational Evaluation and Policy Analysis, 38*(4), 787–814. https://doi.org/10.3102/0162373716668029

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. New York: Routledge.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks: Sage.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684

Mulyasa, E. (2015). *Menjadi guru profesional*. Bandung: Remaja Rosdakarya.

Organisation for Economic Co-operation and Development (OECD). (2018). *The future of education and skills: Education 2030*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2019). *OECD learning compass 2030*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2023). *Education at a glance 2023*. Paris: OECD Publishing. https://doi.org/10.1787/69096873-en

Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134. https://doi.org/10.1080/15391523.2010.10782565

Purba, R., Fadilah, M., Helendra, & Arsih, F. (2025). Development of Local Wisdom-Based Inquiry E-Module on Anaerobic Respiration to Improve Students' Science Process and Collaboration Skills in Phase F, State Senior High School 2 Plus Panyabungan. *Jurnal Penelitian Pendidikan IPA, 11*(11), 509–515. https://doi.org/10.29303/jppipa.v11i11.13335

Rahma, M., Helmi, & Arafah, K. (2025). Development of STEM-Based Physics Learning Modules and Their Impact on Students' Scientific Literacy Skills. *Jurnal Penelitian Pendidikan IPA, 11*(11), 560–572. https://doi.org/10.29303/jppipa.v11i11.13014

Saputra, J., Rosidin, U., Yulianti, D., & Widodo, M. (2025). Development of an Instrument to Assess Critical and Creative Thinking Skills in Fourth-Grade Elementary Students. *Jurnal Penelitian Pendidikan IPA, 11*(11), 909–918. https://doi.org/10.29303/jppipa.v11i11.11994

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Timperley, H. (2007). *Teacher professional learning and development*. Wellington: Ministry of Education.

Tondeur, J., Van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2017). Teachers' pedagogical beliefs and technology use. *Computers & Education, 98*, 79-90. https://doi.org/10.1016/j.compedu.2016.02.014

Trust, T., Krutka, D. G., & Carpenter, J. P. (2016). Together we are better. *Computers & Education, 102*, 15–34. https://doi.org/10.1016/j.compedu.2016.06.001

UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. Paris: UNESCO.

van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical teacher, 34*(3), 205-214. https://doi.org/10.3109/0142159X.2012.652239

Windriati, Senowarsito, & Maryanto. (2025). Implementation of Principal Academic Supervision to Improve the Pedagogical Competence of Elementary School Teachers. *Jurnal*

Penelitian Pendidikan IPA, 11(11), 831–838. https://doi.org/10.29303/jppipa.v11i11.13186

Winaryati, E., Munsarif, M., Junaedi, I., Utomo, Arianty, A. D., Khoirunisa, A., & Ibrahim, K. M. (2025). Analysis of the Integrated E-Mesp 4CS Mobile Instrument. *Jurnal Penelitian Pendidikan IPA, 11*(11), 1021–1028. https://doi.org/10.29303/jppipa.v11i11.12980

Zamroni. (2016). *Supervisi pendidikan*. Yogyakarta: Ombak.

Zhang, J., Hong, H. Y., Scardamalia, M., Teo, C. L., & Morley, E. A. (2016). Sustaining knowledge building. *Journal of the Learning Sciences, 25*(4), 598–626. https://doi.org/10.1080/10508406.2016.1141838