

Developing a Context-Sensitive Teacher Self-Efficacy Assessment Instrument in an Indonesian Islamic Secondary School

Rismayani¹, Sahraini², Subekti Masri³

State Islamic University of Palopo, Indonesia

E-mail: 2205020011@uinpalopo.ac.id

(Received: February-2025; Reviewed: March-2025; Accepted: April-2025;

Available Online: May-2025; Published: May-2025)

ABSTRACT

Teacher self-efficacy is central to instructional decision-making, persistence, classroom management, and professional learning, yet its measurement should reflect the tasks and conditions teachers actually encounter. This study developed and preliminarily validated a context-sensitive teacher self-efficacy assessment instrument for Madrasah Aliyah Sampano, an Islamic secondary school in Luwu Regency, Indonesia. A research-and-development design adapted from the 4D model was applied through needs definition, instrument design, development, and limited dissemination. The needs analysis involved six school stakeholders; two experts reviewed content and language; five teachers participated in a limited pilot; and 21 teachers completed the field test. The final instrument contained 31 Likert-type items organized into seven operational domains. Expert review prompted clarification of ambiguous items, inclusion of educational technology, shorter sentences, and a definition of self-efficacy. All items showed significant item-total relationships in the thesis dataset, with representative coefficients of .784, .846, and .872, while Cronbach's alpha reached .892. These results provide promising preliminary evidence of internal consistency and contextual relevance. The instrument can support school-level self-reflection and professional-development planning, but larger multi-site studies are required to establish factorial, convergent, criterion, and test-retest validity.

Keywords: Teacher Self-Efficacy; Instrument Development; Islamic Secondary School; Professional Development; Preliminary Validation

INTRODUCTION

Teacher self-efficacy provides a critical psychological foundation for professional action because it concerns teachers' judgments about their capability to organize and execute the practices required in particular teaching situations. These judgments shape the effort teachers invest, the persistence they maintain when instruction becomes difficult, and the choices they make when managing classrooms or engaging reluctant learners (Bandura, 1997; Tschannen-Moran et al., 1998). A broad synthesis further indicates that teacher self-efficacy is associated with classroom processes, students' academic adjustment, and teachers' well-being, although the strength of these relationships varies across tasks and contexts (Zee & Koomen, 2016). Consequently, educational leaders need measurement tools that do more than produce a general confidence score; they need evidence that identifies the professional domains in which teachers feel capable and the conditions under which those beliefs weaken.

Teacher self-efficacy is task- and context-specific rather than a fixed personality trait. A teacher may feel highly capable of explaining subject matter but less capable of sustaining student engagement, integrating technology, responding to curriculum change, or communicating with families. The most influential conceptual models therefore emphasize an interaction between perceived teaching demands, personal competence, and contextual resources (Tschannen-Moran et al., 1998; Tschannen-Moran & Hoy, 2001). Cross-national evidence also shows that teacher-, classroom-, school-, and leadership-level conditions contribute to efficacy judgments, confirming that measurement should remain sensitive to organizational context (Fackler & Malmberg, 2016). This contextual character is especially important in schools where resources, professional-development opportunities, and community participation are uneven.

The Teachers' Sense of Efficacy Scale established three widely used domains: instructional strategies, classroom management, and student engagement (Tschannen-Moran & Hoy, 2001). These domains have provided a strong basis for teacher-efficacy research, but a scale designed for broad use cannot automatically represent every institutional setting. Measurement items must specify meaningful tasks, use language familiar to respondents, and avoid combining beliefs about capability with unrelated attitudes or general optimism (Bandura, 2006; DeVellis & Thorpe, 2021). The challenge is therefore not to replace established theory, but to translate it into locally intelligible indicators while retaining conceptual discipline.

Rigorous instrument development requires an iterative relationship among theory, respondent experience, expert judgment, pilot feedback, and statistical evaluation. Content definition and item generation should precede empirical testing, while expert review should examine relevance, clarity, representativeness, and possible ambiguity (American Educational Research Association et al., 2014; Hinkin, 1998). Pilot testing then reveals how respondents interpret the items, and larger field testing provides evidence concerning item performance and internal consistency (Boateng et al., 2018; Worthington & Whittaker, 2006). A high reliability coefficient is useful, but it does not by itself demonstrate that a scale is unidimensional or valid across populations; factorial, convergent, criterion, and temporal evidence remain necessary before an instrument can support high-stakes decisions (DeVellis & Thorpe, 2021).

The need for a locally responsive instrument was evident at Madrasah Aliyah Sampano. Preliminary observations and stakeholder interviews indicated that teachers encountered recurring problems involving student discipline, changing curricula, limited instructional resources, uneven access to educational technology, low student motivation, and insufficient family involvement. Professional development and collegial support were valued, but their frequency and intensity were considered inadequate. These conditions matter because efficacy beliefs develop through mastery experiences, observation of effective practice, social persuasion, and interpretations of emotional or physiological states (Bandura, 1997, 2006). Professional learning that combines knowledge with supported practice and coaching can therefore strengthen efficacy more effectively than information-only training (Tschannen-Moran & McMaster, 2009).

The madrasah context also connects teacher self-efficacy with educational administration and professional competence. Teachers are expected to master subject content, design effective instruction, manage learning environments, use technology, evaluate student progress, and participate in continuing professional development. Strong professional competence does not arise from confidence alone, yet confidence affects whether teachers mobilize knowledge and persist in applying it under pressure (Ilyas, 2022; Kırkıç & Çetinkaya, 2020). School leaders consequently need diagnostic information that can guide mentoring, workshops, peer collaboration, and resource allocation rather than relying on undifferentiated judgments about teacher performance.

A measurement gap emerged between established teacher-efficacy constructs and the operational concerns expressed by teachers at Madrasah Aliyah Sampano. The school required an instrument that represented instructional capability, classroom challenges, confidence, institutional support, facilities, collegial collaboration, and continuing professional growth. Existing theory supports the centrality of instructional strategies, classroom management, student engagement, and efficacy sources, but it also warns against treating contextual resources as identical to efficacy itself (Bandura, 2006; Klassen et al., 2011; Lazarides & Warner, 2020). Developing the instrument therefore required both contextual inclusion and later conceptual refinement.

This study aimed to develop a context-sensitive teacher self-efficacy assessment instrument, document the contextual factors incorporated into its design, and examine preliminary evidence of item functioning and internal consistency at Madrasah Aliyah Sampano. Its contribution is twofold. Practically, it offers a school-level diagnostic tool for reflective professional-development planning. Methodologically, it demonstrates how a locally developed instrument can be reported transparently while distinguishing preliminary item evidence from full construct validation.

METHOD

This study used a research-and-development design adapted from the Four-D model—define, design, develop, and disseminate—because it aimed to produce and refine an assessment instrument rather than merely describe teachers' existing self-efficacy levels (Thiagarajan et al., 1974). The define stage identified local needs and reviewed relevant theory; the design stage translated the construct and contextual requirements into an item blueprint; the develop stage included expert review, limited piloting, revision, and statistical testing; and the disseminate stage involved limited school-level distribution. The initial needs analysis and fieldwork were conducted from 6 November 2023 to 30 January 2024 at Madrasah Aliyah Sampano, Larompong Selatan District, Luwu Regency, South Sulawesi, Indonesia, while expert reviews were documented in August 2024. At the time of the study, the school served 111 students and employed 22 teachers. Because the development process was intentionally bounded to one institution, the items could reflect the school's language, organizational conditions, and professional concerns, although the resulting evidence was interpreted as preliminary and locally specific rather than generalizable to all Indonesian madrasahs.

The needs analysis involved six stakeholders represented in the results table: one madrasah principal, one vice principal for student affairs, and four subject teachers in biology, physics, chemistry, and Indonesian language. Their teaching experience ranged from approximately five years to more than twenty years, enabling the study to capture concerns from early-, mid-, and experienced-career educators. Participants discussed teaching challenges, knowledge of self-efficacy, perceived confidence, institutional support, participation in professional development, facilities, collegial collaboration, and expectations for the proposed instrument. The draft was subsequently reviewed by Dr. Dodi Ilham, M.Pd.I., as the management-education expert, and Dr. Hisbullah, M.Pd.I., as the language expert. The first review focused on item relevance, construct coverage, and the inclusion of educational technology, whereas the second examined clarity, sentence length, technical terminology, and consistency. A limited pilot involving five teachers was then conducted, followed by revision and a field test with 21 teachers. Participant names were omitted, and respondents received information about the research purpose before participation.

Item construction integrated three conceptual sources. First, self-efficacy was defined as a belief in one's capability to perform specific professional tasks rather than as a general expression of self-esteem (Bandura, 1997, 2006). Second, instructional strategies, classroom management, and student engagement were adopted as the principal task domains of teacher efficacy (Tschannen-Moran & Hoy, 2001). Third, the needs analysis added contextually relevant areas involving curriculum adaptation, parent involvement, institutional support, facilities, technology, collaboration, and continuing professional growth. The resulting questionnaire used a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) and contained 31 items across seven operational domains: relevance to teaching experience and challenges, teacher confidence, institutional support and professional development, availability and need for facilities, collaboration and social support, practicality and ease of use, and continuous evaluation and development. Although this structure preserved the product described in the thesis, the interpretation distinguished domains that directly assessed capability beliefs from those measuring contextual enabling conditions or user perceptions of the instrument.

Data were collected through semi-structured interviews, classroom observation, document review, expert-review forms, pilot feedback, and the 31-item questionnaire. Interview and observation data were summarized to identify recurring needs and determine whether the proposed domains corresponded with school practice, while expert comments and pilot responses were examined qualitatively to identify items requiring clarification, shorter wording, examples, or stronger technological context. Field-test responses were analyzed using Pearson item-total correlations and Cronbach's alpha. The thesis reported that all 31 items were significantly associated with the total score and were therefore retained, with representative coefficients of .784 for Item 1, .846 for Item 3, and .872 for Item 5; internal consistency was reported as $\alpha = .892$. Because the sample consisted of only 21 teachers for 31 items, exploratory or confirmatory factor analysis was not undertaken. Accordingly, the statistical findings were interpreted as preliminary evidence of item coherence and internal consistency rather than definitive evidence of factorial construct validity (Boateng et al., 2018; Worthington & Whittaker, 2006).

RESULTS

Contextual Needs Identified at the Define Stage

The needs analysis showed that a locally useful instrument had to reflect both teachers' capability beliefs and the conditions that shaped those beliefs. Experienced teachers generally described greater confidence because repeated teaching situations had provided practical strategies, whereas less-experienced teachers reported a stronger need for structured support. Experience alone was not sufficient, however, because confidence weakened when curriculum expectations changed, resources were limited, or difficult classroom situations persisted. The instrument therefore needed to represent professional tasks and the contexts in which those tasks were performed.

Four recurring clusters shaped item development. The first concerned instructional adaptation and classroom management, including curriculum change, student discipline, low motivation, and the need for more varied strategies. The second concerned material and technological resources, particularly limited teaching aids and insufficient access to educational technology. The third concerned professional and organizational support, including the frequency of training, mentoring, and collaboration. The fourth concerned family and community involvement, especially communication with parents and shared responsibility for student learning. These findings established the contextual content that stakeholders expected the instrument to capture.

Table 1. Contextual Needs Used to Define the Instrument

Need cluster	Evidence from the school context	Implication for item development
--------------	----------------------------------	----------------------------------

Need cluster	Evidence from the school context	Implication for item development
Instruction and classroom demands	Curriculum changes, student discipline, low motivation, and difficulty sustaining active learning.	Include capability beliefs for planning, explanation, classroom management, motivation, and adaptation.
Resources and technology	Limited teaching aids, uneven facilities, and a need for greater educational-technology access.	Represent technology adaptation and record perceived enabling conditions.
Professional and organizational support	Training and workshops were valued but considered insufficient in frequency and intensity; collegial collaboration required strengthening.	Include professional learning, institutional support, feedback, and collaboration.
Family and community engagement	Teachers reported limited parent involvement and a need for more effective communication.	Include teacher capability and contextual items related to parent engagement.

Instrument Design and Expert Revision

The design stage produced a 31-item questionnaire distributed across seven operational domains. The largest domain contained seven items on teaching experience and instructional challenges, while each of the remaining domains contained four items. The sequence moved from core teaching tasks toward institutional context, collaboration, instrument use, and continuing development. This organization responded directly to stakeholders' request for an instrument that was comprehensive but still practical for routine reflection.

Expert review resulted in substantive and linguistic revision rather than wholesale item deletion. The management-education expert judged the content broadly relevant but requested examples for ambiguous statements, clearer wording, and an item addressing adaptation to educational technology. The language expert requested a short definition of self-efficacy, shorter sentences, and consistent terminology. These revisions improved interpretability and made the items more closely connected to recognizable teaching situations.

Table 2. Operational Domains of the 31-Item Instrument

No.	Operational domain	Items	Primary content
1	Teaching experience and challenges	7	Planning, explaining, classroom management, resource constraints, curriculum adaptation, student motivation, and parent involvement.
2	Teacher confidence	4	Understanding self-efficacy, perceived capability to teach effectively, confidence in teaching tasks, and handling classroom challenges.
3	Institutional support and professional development	4	School support, usefulness of workshops, participation in training, and perceived training benefits.
4	Facilities and development needs	4	Availability of resources and technology, need for teaching aids, and need for more intensive training.
5	Collaboration and social support	4	Sharing practice, collegial learning, support from colleagues and family, and coping with teaching challenges.
6	Practicality and ease of use	4	Ease, routine usability, coverage, and support for identifying areas for improvement.
7	Continuous evaluation and development	4	Self-evaluation, development planning, continued professional growth, and collaborative participation.

Pilot Testing and Item Revision

The five-teacher pilot confirmed that the questionnaire was generally acceptable but exposed interpretive problems. The statement about strategies for motivating students was considered too general, the phrase about parent involvement required a concrete communication context, and several respondents preferred a sequence that followed planning, implementation, and evaluation. The revisions added examples, clarified parent communication, and improved the logical order of the items.

One documented revision illustrates the process. The original item, "I have strategies to increase student motivation," was changed to "I have strategies to increase students' learning motivation, for example through interactive learning methods or a personal approach." The revised wording converted a broad declaration into a more recognizable professional task. The final product retained 31 items after the addition or reconfiguration of technology-related content.

Preliminary Item Performance and Internal Consistency

The field test involved 21 of the school's 22 teachers. The thesis reported significant item-total relationships for all 31 items and did not eliminate any item. Representative results were strong: Item 1 correlated .784 with the total score, Item 3 correlated .846, and Item 5 correlated .872, with reported significance values below .001. These coefficients indicate that the example items moved consistently with the overall score in this sample.

The 31-item instrument produced Cronbach's alpha of .892. This value indicates high internal consistency for the total score in the field-test sample. The result supports the use of the instrument for preliminary school-level reflection, but it does not resolve whether the seven domains form one construct or several related constructs. The small sample and the inclusion of contextual and usability items require a conservative interpretation.

Table 3. Summary of Development and Preliminary Validation Evidence

Stage	Evidence	Outcome
Needs analysis	Six stakeholders with varied roles and teaching experience.	Four main contextual clusters were incorporated into the blueprint.
Expert review	One management-education expert and one language expert.	Ambiguous wording was clarified; technology, definitions, brevity, and consistency were strengthened.
Limited pilot	Five teachers completed the instrument and gave feedback.	Examples and contextual cues were added; item order was improved.
Field test	Twenty-one teachers completed the 31-item form.	All items were retained based on reported item-total evidence.
Internal consistency	Cronbach's alpha for the total instrument.	$\alpha = .892$, indicating high internal consistency in the local sample.
Representative item-total coefficients	Item 1 = .784; Item 3 = .846; Item 5 = .872.	Promising item coherence, pending replication and factor analysis.

DISCUSSION

The study demonstrates that teacher self-efficacy measurement becomes more useful to educational management when it begins with the professional realities of a specific school. The needs analysis did not reveal an abstract demand for confidence; it revealed concrete problems involving curriculum adaptation, classroom discipline, motivation, resources, technology, family engagement, and professional support. Efficacy judgments are formed by interpreting demands in relation to perceived competence and available resources, so these local conditions can shape how teachers judge their own capability (Fackler & Malmberg, 2016; Tschannen-Moran et al., 1998). The contextual development process therefore strengthened the practical relevance of the instrument.

The instrument's strongest conceptual contribution lies in its task-oriented items. Statements about planning, explaining material, managing classrooms, motivating students, adapting to curriculum change, using technology, and involving parents ask teachers to evaluate recognizable professional actions. Task specificity is essential because self-efficacy scales should measure perceived capability rather than general confidence or personal worth (Bandura, 2006). These items also extend the established domains of instructional strategies, classroom management, and student engagement into the operational context of an Indonesian madrasah (Tschannen-Moran & Hoy, 2001).

The needs analysis also confirms that teacher self-efficacy is organizationally situated. Teachers identified collegial support, training, facilities, leadership support, and family participation as conditions that strengthened or constrained their work. Research across multiple countries similarly shows that school and leadership characteristics are associated with efficacy beliefs, while job resources and trust in colleagues contribute to teacher well-being through personal resources such as self-efficacy (Fackler & Malmberg, 2016; Huang et al., 2019). Professional development can be particularly influential when it generates mastery experiences, models effective practice, provides credible feedback, and includes follow-up support (Bandura, 1997; Tschannen-Moran & McMaster, 2009). These findings justify using the assessment results to guide school support rather than judging teachers in isolation.

At the same time, contextual responsiveness created a construct-boundary problem. Items concerning institutional support, available facilities, training frequency, social support, and instrument practicality do not all measure a teacher's belief that they can perform a task. Some measure antecedents of efficacy, some measure professional-development behavior, and others measure usability of the product. Scale-development standards require a clear distinction between the focal construct and variables that influence or result from it (American Educational Research Association et al., 2014; DeVellis & Thorpe, 2021). The present product is therefore best understood as a context-sensitive teacher self-efficacy assessment package with enabling-condition and usability components, rather than a fully established unidimensional efficacy scale.

The reported alpha of .892 is encouraging because the items showed substantial internal consistency in the local sample. Similarity among items and a relatively large number of items can elevate alpha, however, so reliability should not be treated as proof of dimensional validity (Boateng et al., 2018; Hinkin, 1998). The significant item-total relationships likewise indicate coherence with the total score, but they cannot determine whether the seven domains represent one factor, correlated factors, or a mixture of efficacy and contextual constructs. This distinction is especially important because the field test included 21 respondents for 31 items,

a ratio that is not adequate for stable exploratory or confirmatory factor analysis (Worthington & Whittaker, 2006).

The preliminary results still have direct value for school leadership when used appropriately. A school can examine domain profiles to identify whether teachers need support in instructional strategy, classroom management, technology adaptation, parent communication, peer collaboration, or reflective planning. The results can then inform mentoring, targeted workshops, peer observation, coaching, and resource allocation. Such use is consistent with evidence that self-efficacy relates to teaching attitudes, professional persistence, job satisfaction, and classroom processes (Kırkıç & Çetinkaya, 2020; Klassen & Chiu, 2010; Zee & Koomen, 2016). The instrument should support developmental conversations, not ranking, certification, promotion, or punitive evaluation.

The madrasah setting gives the study additional administrative relevance. Islamic secondary schools combine national curriculum responsibilities with institutional values, community expectations, and varying resource conditions. A locally understandable instrument can help leaders connect teacher reflection with continuing professional development and school improvement. Professional competence requires more than self-perception, so the questionnaire should be paired with classroom observation, teaching portfolios, student-learning evidence, and professional-development records (Ilyas, 2022). This triangulated approach prevents self-report scores from becoming the sole basis for conclusions about teacher quality.

Future refinement should separate the current item pool into conceptually cleaner components. A core teacher self-efficacy scale could retain capability statements related to instruction, classroom management, student engagement, technology, assessment, curriculum adaptation, and family communication. A second scale could assess enabling conditions such as leadership support, facilities, training, and collegial climate, while a separate short form could assess usability of the instrument. The revised measures should be administered across several madrasahs with a substantially larger sample, followed by exploratory and confirmatory factor analysis, convergent testing against an established efficacy measure, criterion testing against observed practice, and test-retest reliability (American Educational Research Association et al., 2014; Boateng et al., 2018; DeVellis & Thorpe, 2021).

The study's methodological value lies in making the limits of preliminary validation explicit. Locally grounded development can generate relevant content, but relevance must be followed by rigorous construct testing. The present instrument provides a credible starting point because it integrates stakeholder needs, expert judgment, pilot interpretation, and initial statistical evidence. Its next stage should convert that starting point into a more parsimonious and replicable measurement model.

CONCLUSION

This study developed a 31-item teacher self-efficacy assessment instrument for Madrasah Aliyah Sampano through needs definition, theoretical design, expert review, limited piloting, revision, and local field testing. The instrument addressed teaching experience and challenges, confidence, institutional support, facilities, collaboration, practicality, and continuous development. Expert feedback improved content relevance and readability, while pilot feedback made the items more concrete and context-sensitive. The field-test results provided promising preliminary evidence, with all items retained, representative item-total coefficients ranging from .784 to .872, and Cronbach's alpha of .892.

The instrument can assist madrasah leaders and teachers in mapping professional needs and planning reflective development, particularly in instruction, classroom management, technology, collaboration, and institutional support. Its use should remain formative and diagnostic. The study was limited to one school, 21 field-test respondents, self-report data, item-total analysis, and internal-consistency estimation; it did not establish factor structure, cross-school generalizability, criterion validity, or temporal stability. Future studies should test a refined core efficacy scale and separate contextual and usability components using larger multi-site samples, factor analysis, convergent and criterion validation, and test-retest procedures.

DECLARATION OF CONFLICTING INTEREST

The author declares that there is no conflict of interest in this work.

FUNDING ACKNOWLEDGEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Asmayawati, Yufiarti, & Yetti, E. (2024). Pedagogical innovation and curricular adaptation in enhancing digital literacy: A local wisdom approach for sustainable development in

- Indonesia context. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(2), 100233. <https://doi.org/10.1016/j.joitmc.2024.100233>
- Bachtiar, B., & Puspitasari, M. (2024). Barriers and strategies in implementing innovative and collaborative learning: A case of Indonesian English language teaching. *IJEE (Indonesian Journal of English Education)*, 11(1), 61–76. <https://doi.org/10.15408/ijee.v11i1.38434>
- Biesta, G., Priestley, M., & Robinson, S. (2015). The role of beliefs in teacher agency. *Teachers and Teaching*, 21(6), 624–640. <https://doi.org/10.1080/13540602.2015.1044325>
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 585–596. <https://doi.org/10.1080/01596306.2011.601555>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- de Jong, L., Meirink, J., & Admiraal, W. (2022). School-based collaboration as a learning context for teachers: A systematic review. *International Journal of Educational Research*, 112, 101927. <https://doi.org/10.1016/j.ijer.2022.101927>
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, 103(4), 962–1023. <https://doi.org/10.1086/231294>
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.
- Hakim, M. I., & Alfriti, B. (2025). The implementation of Kurikulum Merdeka in Islamic education in the 3T regions of North Lombok. *MODELING: Jurnal Program Studi PGMI*, 12(3), 269–293. <https://doi.org/10.69896/modeling.v12i3.2969>
- Iman, B. R. N., Nuris, M., & Syaripuddin, R. (2023). Integrated technology in learning viewed from the lens of 21st century education: A qualitative study on teacher's efforts in improving Islamic education quality. *Edukasi Islami: Jurnal Pendidikan Islam*, 12(1), 447–461. <https://doi.org/10.30868/ei.v12i01.3038>
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2024). Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 12 Tahun 2024 tentang kurikulum pada pendidikan anak usia dini, jenjang pendidikan dasar, dan jenjang pendidikan menengah.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler Publishers.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Publishing Company.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Nurlaili, Suranda, N., & Purwanto. (2024). Analisis inovasi Kurikulum Merdeka berdiferensiasi pada pembelajaran Pendidikan Agama Islam (PAI) di sekolah umum. *MODELING: Jurnal Program Studi PGMI*, 11(1), 821–831. <https://doi.org/10.69896/modeling.v11i1.2429>
- Pantić, N. (2015). A model for study of teacher agency for social justice. *Teachers and Teaching*, 21(6), 759–778. <https://doi.org/10.1080/13540602.2015.1044332>
- Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury Academic.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–431. <https://doi.org/10.3102/00346543072003387>
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.

- Vähäsantanen, K. (2015). Professional agency in the stream of change: Understanding educational change and teachers' professional identities. *Teaching and Teacher Education*, 47, 1–12. <https://doi.org/10.1016/j.tate.2014.11.006>
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, 15, 17–40. <https://doi.org/10.1016/j.edurev.2015.04.002>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.