IJOMER
(Indonesian Journal of Multidisciplinary Educational Research)
P-ISSN: 2987-7601 E-ISSN: 2987-0968
https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/ijomer
DOI: https://doi.org/10.30762/ijomer.v4i2.9259

Implementation of Non-Test Evaluation in Descriptive Text Learning in Class IX Science SMPS Sabilillah

***Rulisa Jovana Bahar** [1]
*Indonesian Language Tadris Study Program, Faculty of Tarbiyah and Teacher Training, UIN Madura, Indonesia
Email: 23381072061@student.uinmadura.ac.id
Liana Rocmatul Wachidah [2]
Indonesian Language Tadris Study Program, Faculty of Tarbiyah and Teacher Training, UIN Madura, Indonesia
Email: lianarwachidah@uinmadura.ac.id

(*) Email Correspondent: 23381072061@student.uinmadura.ac.id

Abstract

This article describes how non-test evaluation was carried out in descriptive text learning in Grade IX Science at SMPS Sabilillah. Non-test evaluation matters in authentic assessment because it can capture attitudes and skills that written tests rarely reveal. The study used a descriptive qualitative design involving 30 students and one Indonesian language teacher, selected through purposive sampling. Data were collected through participatory observation, semi-structured interviews, and document analysis, and were checked using source triangulation, technique triangulation, and member checking. The analysis followed the interactive model of Miles, Huberman, and Saldaña: data reduction, data display, and conclusion drawing and verification. The results show that the teacher combined four formative assessment components (oral, written, performance, and assignment) in a sequence that followed the flow of learning. Assignments and performance drew the most consistent student engagement, whereas the oral component caused speaking anxiety in some students. The teacher also reported a heavier workload in preparing rubrics and recording observations. Over the semester, students' descriptive texts became more detailed, more varied in vocabulary, and more personal. The study concludes that non-test evaluation deserves a place as a main assessment instrument alongside tests in descriptive text learning, provided that teachers receive adequate institutional support.

Keywords: non-test evaluation; descriptive text; authentic assessment; performance assessment; Merdeka Curriculum

Received:	25-06-2026	Revised:	28-10-2026	Accepted:	31-10-2026

Copyright: © 2022 Fakultas Tarbiyah, Institut Agama Islam Negeri Kediri. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution - ShareAlike 4.0 International License (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

INTRODUCTION

Evaluation is an inseparable part of education, because through evaluation, teachers can find out the extent to which students have understood the material being taught. The results of evaluation make the achievement of learning objectives clearly visible (Ali, 2000; Surur et al., 2026). In general, evaluation is understood as an assessment process to find out whether an activity has achieved its goal. Therefore, evaluation is not enough to be seen as a one-time measurement. It includes planning, data collection, reporting, and the use of the results to improve student learning achievement, so that in essence evaluation is an effort to collect information needed to make decisions (Azizah et al., 2025).

Tests are not the only way to assess learning outcomes. The non-test technique provides another path, namely assessment without test questions that are directed at the student's personality as a whole, such as attitudes, behaviors, character, social relationships, ways of expressing themselves, and life backgrounds, both in individuals and groups (Al Amien et al., 2026; Tamam et al., 2026). The implementation of the Independent Curriculum strengthens this direction at the junior high school level. Assessment no longer stops at the knowledge aspect, but is required to be more authentic and reach all student competencies. Majid (2022) emphasized that authentic assessment requires a variety of instruments, not just written tests, so that students' competencies in real learning situations can be recorded. This development is what encourages the wider use of non-test evaluations in junior high school classrooms.

In grade IX, the description text asks students to describe an object in detail in vivid language. The skills required, namely observation sharpness, accuracy in choosing words, and creativity in composing images, are difficult to measure through objective questions. Nurgiyantoro (2020) stated that students' creativity in writing description texts needs to be assessed with tools that are sensitive to the process, not just the final result. Thus, observation, performance assessment, and portfolio are reasonable choices for this material.

Initial observations at SMPS Sabilillah show that there is a gap between curriculum demands and classroom practice. Indonesian language teachers in this school are still quite dependent on written exams and daily tests, while the assessment of student attitudes, processes, and performance has not been

systematically organized. A similar trend was noted by Sudjana (2020), namely the practice of evaluation in schools which is dominated by formal tests so that the affective and psychomotor aspects of students are less well documented. In grade IX of Science, teachers began to combine several forms of non-test assessments in one semester, so this class is worth studying as an example of application.

This study is important to be carried out at SMPS Sabilillah for three reasons. First, the assessment of descriptive texts that rely on written exams risks hiding students' observation and writing skills, even though that ability is the core of this material. Second, students in the Science class are used to observing objects, so this class has the opportunity to show how observation skills can be assessed through text. Third, as an Islamic-based private school that implements the Independent Curriculum, SMPS Sabilillah needs an overview of assessment practices that can be used as material for reflection for teachers and schools in developing assessment policies.

A number of previous studies have touched on non-test evaluations in language learning. Wahyuni and Ibrahim (2021) reported that performance assessment in learning to write exposition texts increases student motivation and engagement. Minatik et al. (2026) found that portfolios provide a more complete picture of the development of writing skills compared to conventional tests. In another field, Cahayu et al. (2023) examined affective and psychomotor domain non-test instruments in 21st century skill-based science learning, while in the Indonesian language group, Sitorius et al. (2023) discussed the authentic assessment of writing descriptive texts in the Independent Curriculum at SMP Labschool UPI. However, research that specifically examines non-test evaluations of class IX descriptive texts, especially in religious-based schools, is still difficult to find. This gap is what this article aims to fill.

Based on this description, the formulation of this research problem is (1) how the pattern of implementation of non-test evaluation in the learning of descriptive texts in grade IX of Science at SMPS Sabilillah is concerned, (2) how is the involvement and response of students to each component of the evaluation, (3) what are the challenges faced by teachers and students, and (4) how is the impact of non-test evaluation on the learning process. In line with that, this study aims to describe these four things as well as provide a reflective record for the improvement of assessment practices in schools. The novelty of this article lies in its context, namely the non-test evaluation of descriptive texts in Islamic-based

schools that has not been widely discussed in the Indonesian language education literature.

METHOD

Research Design

This study uses a descriptive qualitative approach because the goal is to understand the implementation of non-test evaluations in the learning of descriptive texts as they take place in the classroom. According to Moleong (2021), qualitative research seeks to capture the meaning, process, and context of a symptom from the perspective of the perpetrators, so that the results are in the form of detailed and contextual descriptions, not just statistical numbers.

Population and Sample OR Subject

The research was carried out on June 22, 2026 at SMPS Sabilillah, a private junior high school based on Islamic teachings. This school was chosen because it has developed a number of assessment innovations that are in line with the Independent Curriculum, making it suitable for studying non-test evaluation practices. The research participants consisted of all 30 students in grade IX Science, all of whom were women according to school policy, and one Indonesian language teacher who taught the class. Participants were determined using the purposive sampling technique, which is the deliberate selection of subjects based on information needs and their suitability with the focus of the research (Sugiyono, 2022). Data collection was carried out through classroom observations, semi-structured interviews, and documentation. The number of meetings observed was adjusted to the implementation of observations recorded during the study.

Data Analysis

Data were collected using three complementary techniques. First, participatory observation: the researcher was present in person in class during the non-test assessment, observed teacher and student activities, recorded interactions, and wrote down class conditions in field notes. Second, semi-structured in-depth interviews with teachers to explore their views on the implementation, challenges, and meaning of non-test assessments, as well as with a number of students who were deliberately selected to find out their experiences when assessed with non-test instruments. Third, document analysis, namely the review of assessment rubrics, student writings, teachers' formative grade notes, and learning planning tools (RPP/Teaching Modules) to understand the design and implementation of non-test evaluations in the classroom.

The components of the assessment studied were assignments, performance assessments, oral assessments, and written assessments. All four were formative, process-focused, and assessed by rubrics. Therefore, the oral and written assessments in this study were different from the final exams and were still grouped as non-test evaluations.

The validity of data is tested through triangulation, which is data examination by utilizing sources, techniques, or other points of view so that findings do not depend on only one perspective (Denzin, 1978; Patton, 1999; Surur et al., 2025). Denzin (1978) distinguishes several types of triangulation, namely data triangulation or sources, methods, researchers, and theories. This study applies two types that are most in line with the design and data sources, namely source triangulation and technique triangulation, as also explained by Moleong (2021) and Sugiyono (2022). Source triangulation is carried out by comparing information from teachers, students, and documents about the same thing. Technique triangulation is carried out by matching the results of observations, interviews, and document analysis. For example, the teacher's statement about the order of assessment is checked with the schedule that appears during observation and with the content of the teaching module. In addition, member checking is carried out by conveying findings to teachers and several students to ensure that the findings are in accordance with their experiences (Creswell, 2022).

Data analysis followed the interactive model of Miles et al. (2020) which consisted of three interrelated activity flows. Data reduction was carried out by sorting, focusing, and simplifying raw data from field notes, interview transcripts, and documents so that data was left in accordance with the focus of the research. Data presentation was carried out by arranging the selected data into narrative descriptions, tables, or charts so that conclusions were easier to draw. Conclusion drawing and verification was carried out by formulating temporary conclusions, then testing them continuously during the analysis until credible conclusions were obtained.

FINDINGS AND DISCUSSION

FINDINGS

Pattern of Non-Test Evaluation in the Classroom

Observation results

Classroom observation shows that teachers carry out four components of assessment, namely oral tests, written tests, performance, and assignments, in an integrated manner in the learning of descriptive texts in grade IX of Science at Sabilillah Junior High School. In the early stages of learning, students are given the task of observing objects and compiling an initial draft of the description text. This activity is part of the task assessment process as well as providing an initial overview of students' ability to develop observation results into written form. In the middle of the learning unit, the teacher conducts a performance assessment to observe students' involvement and abilities during the learning process. Furthermore, an oral assessment is carried out at the end of the unit to determine students' understanding of the material that has been learned. In addition, written assessment is used as one of the components to determine students' ability to compose descriptive texts. The series of activities shows that non-test evaluations are carried out gradually and integrated with the learning process, not just as a closing activity.

Interview results

The teacher's view of the order of assessment is seen in the following interview.

> *"I don't want to judge children just at the end. They need to know from the beginning what they are assessing, so that the learning process is directed and they have the opportunity to develop along the way." (G1)*

In addition to regulating the order of assessment, teachers also need to pay attention to the preparation of rubrics and the delivery of assessment criteria to students. Rubrics serve as guidelines in assessing learning processes and outcomes, while the delivery of criteria from the beginning helps students understand aspects that are of concern in completing assignments. To find out how this is applied in the learning of description texts in grade IX of Science at SMPS Sabilillah, the results of interviews with Indonesian teachers are examined based on explanations of the preparation of rubrics, determination of criteria, and how to communicate these criteria to students.

Documentation results

The Teaching Module examined showed the alignment between the learning objectives, learning activities, and non-test instruments used. This indicates that the teacher has designed the assessment in a directional manner from the planning stage.

The list of teacher grades for grade IX Science for the 2025/2026 academic year for the odd semester contains the first formative score in the description text material for 30 students in four components, namely oral tests, written tests, performance, and assignments. A summary of these grades is presented in Table 1.

Table 1. Summary of formative grades of class IX Science descriptive text (n = 30)

Components	Average	Lowest	Highest
Oral test	82,9	78	93
Written test	83,6	80	92
Performance	84,1	78	92
Duties	84,1	78	92

The average score of the four components was in the range of 83 to 84, with an average of 83.7 for all components. The performance and assignment components had the highest average (84.1), followed by the written test (83.6) and the oral test (82.9). The difference between the components was small, but the direction of the sequence was in line with the observational finding that tasks and performance were the strongest components of student involvement. The oral test had the lowest average, although the difference was thin. The highest score in each component reached 92 or 93, while the lowest score was 78 to 80. Individual student grades are displayed in the appendix of the teacher's grade list.

Student Engagement and Response to the Evaluation Component

Observations and interviews show that the level of student involvement differs in each component. The assignment and performance components show the strongest involvement. This can be seen from the students' enthusiasm when doing independent assignments, the courage to ask questions to the teacher, and the liveliness during the project presentation session.

In the performance component, some students said that the opportunity to read the description text in front of the class provided a memorable learning experience. One of the students said:

> *"When I read my text in front of my friends, I understand more which sentences are less clear. This is different from just handing over the writing." (Q1)*

This experience reflects the reflective value of performance assessment as stated by Nurgiyantoro (2020). When students display their abilities directly, they gain the opportunity to evaluate their own writing, especially in recognizing

parts of the text that are still unclear. These findings also show that performance assessment not only serves to measure students' abilities, but can be part of the learning process and improvement of work.

In addition to performance assessment, self-assignment also shows positive student involvement. Observation results indicate that students who tend to be passive in class discussions still have the opportunity to demonstrate their abilities through written work. Thus, self-assignment and performance assessment provide space for students to demonstrate competence in different ways.

The assignment component shows an interesting pattern. Students who appear passive in class discussions actually produce work that stands out on independent assignments. This finding supports the view of Minatik et al. (2026) that independent assignments provide a safer space for introverted students to demonstrate their competencies without social pressure.

Challenges Faced by Teachers and Students

The oral test component is one aspect of the evaluation that elicits diverse responses from students. The interview results show that some students experience anxiety when they have to answer the teacher's questions orally in front of the class. This condition shows that the implementation of the assessment is not only related to the mastery of the material, but can also be influenced by the psychological condition of the students at the time of the assessment. In line with the views of Magdalena et al. (2023), anxiety in assessment situations does not necessarily indicate a student's low ability, but needs to be understood by considering the conditions they experience when they have to give answers directly.

From the teacher's side, the main challenge conveyed in the interview is the need for greater time and energy to assess 30 students individually. This burden does not only arise during the assessment, but also in the stages before and after, namely compiling rubrics for each component, recording the results of daily observations, and processing and reporting grades one by one. This finding is in line with Aziz et al. (2020) who said that the administrative burden of authentic assessment, such as filling out rubrics, compiling portfolios, and reporting assessments, can consume teachers' time and energy for teaching preparation and risk reducing teaching effectiveness if not supported by an adequate system. Teachers said:

> *"Judging one by one is indeed more tiring than correcting multiple-choice questions. But I feel like the results are more honest, more depicting the real child." (G1)*

This statement illustrates a dilemma commonly experienced by teachers: assessment results are felt to be more meaningful, but to achieve them requires much greater readiness and perseverance than in conventional tests. Rosidah et al. (2021) found that teachers' readiness, both in understanding concepts, compiling instruments, and having experience in running them repeatedly, greatly determines the quality of the implementation of authentic assessments in the era of the Independent Curriculum. In this case, the readiness of Indonesian language teachers at SMPS Sabilillah actually grows through the process of trying and improving throughout the semester, not from formal training.

Another challenge is the potential for subjective bias in performance assessment, especially when teachers already have a certain impression of students. Purwanto (2020) reminded that the halo effect is a real risk of performance-based assessment and needs to be anticipated through periodic calibration of rubrics and teachers' self-reflection.

The Impact of Non-Test Evaluation on the Learning Process

Results of observations and interviews

Observations show that the application of non-test evaluations has a positive effect on class dynamics. When students realize that their learning is not only assessed through final exams, their behavior changes: they are more careful when observing the objects described, more careful in revising the writing, and more active in class discussions. Some students also state that assessments of performance and assignments make them more concerned about the quality of their writing.

These findings are in line with the intrinsic motivation theory of Deci and Ryan (in Haryati, 2022). Assessments that are designed as a meaningful learning experience, not just a moment of assessment, tend to encourage students to learn to develop, not just to graduate.

Documentation results

Analysis of student documents from the beginning to the end of the semester shows the skill gains seen in descriptions. Descriptions that were originally simple and general develop into more detailed, more varied in vocabulary, and more nuanced in personal nuance. This kind of development is

difficult to capture with conventional test tools, but it becomes apparent when student work is read as a series over the course of the semester.

DISCUSSION

The above findings show that non-test evaluation of descriptive text learning in grade IX of Science Sabilillah Junior High School not only functions as a measuring tool, but also becomes part of the learning process itself. The use of four components at once, namely oral tests, written tests, performance, and assignments, is in line with the principle of authentic assessment according to Kunandar (2020) that a valid and comprehensive assessment requires a variety of instruments so that various dimensions of student competence can be seen at the same time.

The high involvement of students in the task and performance components indicates that a well-designed non-test evaluation can help students demonstrate their best abilities. This is in line with the motivational theory of Deci and Ryan (in Haryati, 2022), which states that assessments designed as meaningful and authentic learning experiences tend to foster intrinsic motivation. Authentic performance assessments provide students with space to display abilities in a real-world context, as opposed to conventional tests that often trigger excessive anxiety.

Consistent involvement in the assignment component is also in line with Arifin's (2022) view that authentic and meaningful assignments are important indicators of student engagement. In project- and task-based learning such as in the Independent Curriculum, the ability to complete authentic assignments independently is worthy of being taken into account in assessment.

The quality of the learning experience in the performance component strengthens the opinion of Nurgiyantoro (2020) that junior high school students have great performance potential if given the opportunity to show their abilities directly and creatively. In the description text, performance assessment makes students not only write, but also present and defend their work in front of classmates, an experience that is very different from working on test questions. This finding supports the results of Wahyuni and Ibrahim (2021) that performance assessment in writing learning in junior high school is effective in improving the quality of student writing, and is in line with Sitorius et al. (2023) that authentic assessment of description text needs to touch on process aspects, not just final products.

Speaking anxiety in the oral test component is a finding that needs serious attention. Magdalena et al. (2023) note that assessment instruments need to be carefully designed to truly reflect student competence, not momentary psychological conditions. Therefore, teachers are advised to organize oral sessions in a more dialogical and supportive atmosphere, for example through small group discussions or paired presentations (Adelia et al., 2024). This method can reduce social pressure without sacrificing the validity of the assessment.

Subjective bias in performance appraisals also needs to be anticipated. As stated by Purwanto (2020), the halo effect can affect teachers' objectivity. Schools can help by providing rubric calibration training and encouraging teachers to conduct regular self-reflection.

The teacher's workload is another critical record. During the current semester, teachers must divide their attention between teaching, compiling and updating rubrics for four components, recording daily observations, and processing and reporting the scores of 30 students individually. This series of tasks is much more complex than correcting written tests. This finding corroborates Aziz et al. (2020) that the administrative burden of authentic assessments can reduce teaching preparation time if not supported by an adequate system. Even so, teachers still maintain non-test evaluations because they believe that the results more accurately describe students. This is a compromise between workload and assessment validity that ultimately depends on the readiness and commitment of teachers. Rosidah et al. (2021) show that readiness can grow from the experience of conducting authentic assessments repeatedly accompanied by reflection. Implicitly, school support in the form of reducing non-essential administrative burdens and providing collaboration time between teachers is just as important as technical training in instrument preparation.

The success of the implementation of non-test evaluations at SMPS Sabilillah can also be attributed to the character of Islamic-based schools that generally instill discipline, responsibility, and work ethic in students. These values can be seen from students' consistency in completing tasks and their activeness in performance assessments. Trianto (2021) is of the view that a conducive school environment and a strong value system are external factors that have a great influence on the effectiveness of authentic assessments.

Overall, these findings support the argument that non-test evaluations are relevant and effective for descriptive text learning in junior high school. Its ability to provide a comprehensive picture of students' knowledge, skills, and attitudes makes it an important complement to a comprehensive evaluation system. Indonesian language teachers are advised to continue to develop valid, reliable, and contextual non-test instruments as part of daily teaching practices (Kosasih, 2021).

CONCLUSION

Based on the results and discussion, three conclusions can be drawn. First, the non-test evaluation of descriptive text learning in grade IX of Science Sabilillah Junior High School takes place in a planned and integrated manner through four components, namely oral tests, written tests, performance, and assignments. Each component captures different competency dimensions and is arranged following the student's learning flow from the beginning to the end of the unit.

Second, the assignment and performance components demonstrate the most consistent and active student engagement. Process-based and real-life assessments provide students with the opportunity to express their competencies more authentically than conventional tests. The oral component, on the other hand, raises speaking anxiety that needs to be addressed by organizing more interactive and supportive sessions. From the teacher's perspective, these assessments require greater time, effort, and readiness and require school support.

Third, non-test evaluations have a positive impact on class dynamics. Students are more motivated from within, more meticulous in observing and writing, and showing real progress on their work throughout the semester. This impact is difficult for conventional tests to reach, but can be observed through a qualitative approach that follows the learning process in the long term.

Thus, non-test evaluations play an important role in providing a comprehensive and authentic picture of the development of students' competencies in learning descriptive texts, and deserve to be a permanent part of the Indonesian language assessment system in junior high schools.

ACKNOWLEDGEMENTS

We would like to acknowledge the invaluable guidance, encouragement, and

support received throughout the completion. We are deeply grateful to our advisors, who provided constructive feedback, insightful suggestions, and continuous encouragement, which have greatly enhanced the quality of this work. Sincere gratitude is also expressed to our family for their endless love, prayers, and encouragement, as well as to friends and colleagues who provided moral support, assistance, and motivation during the research process. Their contributions have been invaluable in ensuring the successful completion.

REFERENCES

Adelia, N. P. C., Artika, I. W., & Wirahyuni, K. (2024). Penilaian Menulis Teks Deskripsi Dalam Penerapan Kurikulum Merdeka: Assessment Of Writing Descriptive Texts In The Implementation Of The Independent Curriculum. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia Undiksha*, *14*(2), 220-229. https://doi.org/10.23887/jpbsi.v14i2.86341

Al Amien, A. F., Farhani, Z. A., Siregar, A. F., Setiawan, M. F., Al Azhar, N. H., & Romelah, R. (2026). Non-Test Evaluation Framework Using the ADDIE Model to Enhance the Quality of Islamic Religious Education Learning. *Journal of Education Research*, *7*(3), 2310-2322. https://www.jer.or.id/index.php/jer/article/view/4059

Ali, Muhammad. (2000). Guru dalam proses belajar mengajar. Sinar Baru Algensindo.

Arifin, Z. (2022). Evaluasi pembelajaran: Prinsip, teknik, prosedur (Edisi ke-9). PT Remaja Rosdakarya.

Arikunto, S. (2021). Dasar-dasar evaluasi pendidikan (Edisi ke-3). Bumi Aksara.

Aziz, M. N. A., Yusoff, N. M., & Yaakob, M. F. M. (2020). Challenges in Using Authentic Assessment in 21st Century ESL Classrooms. *International Journal of Evaluation and Research in Education*, *9*(3), 759-768. https://eric.ed.gov/?id=EJ1274845

Azizah, A. N., Prayoga, B. T., Mandhasihanni, I. N., Damayanti, V. S., & Inayati, N. L. (2025). Implementasi evaluasi non tes pada pembelajaran akhlak di SMP 2 Muhammadiyah Surakarta. Educational Journal: General and Specific Research, 5(1), 68–77. https://felifa.net/index.php/DIDACTIC/article/view/96

Cahayu, S. A., Sampurna, R., Nadira, & Risnita. (2023). Instrument evaluasi non-tes ranah afektif dan psikomotorik pembelajaran IPA sinkronisasi berbasis keterampilan abad 21 di SMP Negeri 6 Sungai Penuh. EDU-BIO: Jurnal Pendidikan Biologi, 6(2), 60–72. https://www.edubio.ftk.uinjambi.ac.id/index.php/edubio/article/view/53

Creswell, J. W. (2022). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Denzin, N. K. (1978). The research act: A theoretical introduction to sociological methods (2nd ed.). McGraw-Hill.

Haryati, M. (2022). Model dan teknik penilaian pada tingkat satuan pendidikan. Gaung Persada Press.

Kemendikbud. (2018). Buku guru Bahasa Indonesia SMP/MTs kelas VII. Kementerian Pendidikan dan Kebudayaan.

Kosasih, E. (2021). Bahasa Indonesia untuk SMP/MTs kelas IX. Erlangga.

Kunandar. (2020). Penilaian autentik: Penilaian hasil belajar peserta didik berdasarkan Kurikulum 2013. PT Raja Grafindo Persada.

Magdalena, I., Ramadhan, G., Wahyuni, H. D., & Safitri, N. D. (2023). Pentingnya proses evaluasi dalam pembelajaran di sekolah dasar. Ta'rim: Jurnal Pendidikan dan Anak Usia Dini, 4(3), 167–176. https://journal.staiypiqbaubau.ac.id/index.php/Tarim/article/view/220

Majid, A. (2022). Penilaian autentik: Proses dan hasil belajar (Edisi revisi). PT Remaja Rosdakarya.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). Qualitative data analysis: A methods sourcebook (4th ed.). SAGE Publications.

Minatik, A., Fadhilah, A. N., Sidik, H., Pratama, R. Y., Adillah, Z., Izzah, V. I., ... & Risqi, S. (2026). Mahārah Qirā'ah Dalam Pendidikan Bahasa Arab: Teori, Strategi, Teknologi, dan Evaluasi Pembelajaran. Inti Litera

Moleong, L. J. (2021). Metodologi penelitian kualitatif (Edisi revisi). PT Remaja Rosdakarya.

Nurgiyantoro, B. (2020). Penilaian pembelajaran bahasa berbasis kompetensi (Edisi ke-3). BPFE Yogyakarta.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Services Research, 34(5, Pt. 2), 1189–1208.

Permendikbudristek Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah.

Priyatni, E. T. (2022). Desain pembelajaran bahasa Indonesia dalam Kurikulum 2013. Bumi Aksara.

Purwanto, N. (2020). Prinsip-prinsip dan teknik evaluasi pengajaran. PT Remaja Rosdakarya.

Rosidah, C. T., Pramulia, P., & Susiloningsih, W. (2021). Analisis kesiapan guru mengimplementasikan asesmen autentik dalam kurikulum merdeka belajar. Jurnal Pendidikan Dasar, 12(1), 87–103. https://doi.org/10.21009/jpd.v12i01.21159

Sitorius, H. J. C., dkk. (2023). Penilaian autentik pada Kurikulum Merdeka Belajar materi menulis teks deskripsi di SMP Labschool UPI. Asas: Jurnal Sastra, 12(1), 200–210. https://pdfs.semanticscholar.org/8e10/34ffe1ac0c3d33f69827e3cbb1c9393a196b.pdf

Sudjana, N. (2020). Penilaian hasil proses belajar mengajar. PT Remaja Rosdakarya.

Sugiyono. (2022). Metode penelitian kuantitatif, kualitatif, dan R&D (Edisi ke-2). Alfabeta.

Surur, A. M., Günerhan, H., Muzaiyin, A. M., Rais, P., Subandono, A., & Pujilestari, S. (2025). An Integrative Model of Music Royalty Determination: Mathematical and Sharia Economic Perspectives. *Fenomena*, *24*(2), 179-192. https://fenomena.lppmuinkhas.com/fenomena/article/view/593

Surur, A. M., Pujilestari, S., Rosli, M. S., & Rabiu, A. A. (2026). Numeracy literacy in mathematics education through murabahah-based contexts. *LINEAR: Journal of Mathematics Education*, *7*(2), 133-156. https://e-journal.metrouniv.ac.id/linear/article/view/13498

Tamam, M. I., Aji, H. B., Rukmini, N. V. N., Qolbiyah, S., Mufidah, N., & Surur, A. M. (2026). Communicative Approach for Arabic Speaking Skill Based on Pesantren Needs. *At-Ta'dib*, *21*(1), 126-139. https://tadib.journal.unida.gontor.ac.id/index.php/tadib/article/view/20

Trianto, I. B. A. (2021). Mendesain model pembelajaran inovatif, progresif, dan kontekstual. Prenada Media Group.

Wahyuni, S., & Ibrahim, A. S. (2021). Asesmen pembelajaran bahasa (Edisi revisi). PT Refika Aditama.

Wirawan. (2020). Evaluasi: Teori, model, metodologi, standar, aplikasi dan profesi. PT Raja Grafindo Persada.