

INTEGRASI EMPAT KETERAMPILAN BERDASARKAN LIMA PRINSIP PADA TES BAHASA INGGRIS DI SMP NEGERI 1 DOLOK PANRIBUAN TIGA DOLOK

Santun Saniora Tamba

santunsaniorat@gmail.com

Universitas HKBP Nommensen Pematangsiantar

Octania Rajagukguk

octaniaarajagukguk@gmail.com

Universitas HKBP Nommensen Pematangsiantar

Netania Diva Sincha Saragih

netaniasaragihppbiuhbnp@gmail.com

Universitas HKBP Nommensen Pematangsiantar

Junita Siska Nauli Saragih

saragihjunita549@gmail.com

Universitas HKBP Nommensen Pematangsiantar

Dumaris E Silalahi

Dumaris.silalahi@uhnp.ac.ad

Universitas HKBP Nommensen Pematangsiantar

ARTICLE INFO Article
History: Received Date: 10th
Oct 2025 Received in
Revised Form Date: 27th Oct
2024 Accepted Date: 29th
Oct 2025 Published online
Date 31th Oct 2025

Keyword: Empat
Keterampilan Bahasa
Inggris, Prinsip Penilaian
Bahasa Kepraktisan,
Reliabilita, Validitas

ABSTRACT

This study examines the design and implementation of an English language proficiency assessment that integrates the four core skills of listening, speaking, reading, and writing based on five basic principles of language assessment: practicality, reliability, validity, authenticity, and washback. The study was conducted at SMP Negeri 1 Dolok Panribuan, involving 23 eighth-grade students as participants in a classroom-based test development and evaluation project. The main objectives of this study were to assess the effectiveness and suitability of the test instrument with the Independent Curriculum and to promote meaningful and pedagogically relevant language learning.

The results of the study showed that students tended to achieve better results in receptive skills (listening and reading) compared to productive skills (speaking and writing). The average scores also indicated that the tasks given were realistically feasible (practical), produced consistent scores through inter-rater checking (reliability), aligned with curriculum objectives (validity), reflected real-world situations (authenticity), and provided feedback that influenced future teaching strategies (washback). Despite facing challenges such as limited facilities and varying student ability levels, the developed assessment model proved to be adaptive and beneficial for both students and teachers. This research emphasizes the importance of applying assessment principles to classroom learning practices and the need for teachers to be equipped with assessment literacy. The study demonstrates that high-quality language assessment can still be implemented effectively even without the support of advanced technology, through careful planning, collaboration between educators, and reflection on teaching practices.

PENDAHULUAN

In the ever-evolving field of English language education, assessment plays a crucial role. It not only measures students' language proficiency but also influences teaching methods, guides curriculum development, and enhances educational outcomes. Today, language assessment has moved beyond merely assigning grades; it has become an integral part of the teaching process, deeply connected to classroom practices. According to O'Malley and Pierce (1996), assessment serves as a tool for teachers to monitor student progress, identify strengths and weaknesses, and improve their teaching strategies based on observed performance. This function is especially vital in countries like Indonesia, where English is a foreign language and students have limited opportunities to use it outside the classroom. In such contexts, assessment is more than just a grading mechanism; it provides meaningful language practice to support communicative competence.

Page | 17

Indonesia's national curriculum, known as Kurikulum 2013, emphasizes communicative English teaching with a student-centered approach. It promotes the integrated use of the four essential language skills: listening, speaking, reading, and writing, while also aiming to equip students with real-life competencies. However, research indicates a gap between the curriculum's intentions and actual classroom practices. For instance, Suryati (2019) found that many Indonesian schools still prioritize grammar and vocabulary through multiple-choice tests. While such tests are convenient and easy to grade, they often fail to reflect students' true communicative abilities. This issue aligns with what McNamara (2000) terms "construct under-representation," where tests do not adequately capture the skills they are intended to measure. Consequently, these assessments may reflect students' test-taking strategies rather than their actual language proficiency.

To address these challenges, Brown and Abeywickrama (2019) propose five key principles for effective language assessment: practicality, reliability, validity, authenticity, and washback. Practicality involves designing assessments that are feasible in real classroom settings, considering time, resources, and usability. Reliability ensures that test results remain consistent across different conditions, graders, or times. Validity refers to the degree to which a test accurately measures the intended language skills. Authenticity means that test tasks should resemble real-world language use, while

washback emphasizes the positive influence of assessment on teaching and learning. When these principles are effectively applied, assessment can become a powerful tool to enhance both language development and teaching quality.

However, although the principles of effective language assessment are clearly explained in theory, they are often not applied in practice in Indonesian junior high school classrooms. Many English teachers lack sufficient training in how to assess students properly. They also face several challenges, such as limited time, large class sizes, and a lack of appropriate tools for testing. Sukyadi and Mardiani (2011) noted that these issues frequently prevent teachers from using assessments that support communicative learning and align with curriculum goals. As a result, tests are often designed and administered based on convenience rather than educational effectiveness.

To address these problems, this study was conducted to design and implement an English language test that follows the five core principles of language assessment. The research took place at SMP Negeri 1 Dolok Panribuan, a public junior high school in North Sumatra, and involved 23 eighth-grade students. The study assessed students' performance in the four main language skills: listening, speaking, reading, and writing. The tests were carefully constructed to reflect real-life communicative situations, using both multiple-choice items and open-ended questions. Scoring was guided by detailed rubrics, and efforts were made to ensure fairness and consistency, following Weigle's (2002) recommendations for performance-based language assessment.

This study holds both theoretical and practical significance. Theoretically, it demonstrates that modern concepts of language assessment can be successfully applied in Indonesian EFL classrooms, and it contributes to the limited body of local research on communicative assessment methods in junior high schools. Practically, the study provides a concrete example of how teachers can design and use assessments that are meaningful and effective, even in contexts with limited resources. It illustrates that, with careful planning and adherence to assessment principles, teachers can create tests that not only measure learning outcomes but also support language development. Moreover, the findings emphasize the need for teacher training in assessment, so that educators can design tasks that both evaluate and enhance student learning. When teachers view

assessment as a tool for student engagement and language growth rather than merely a requirement they become more reflective and effective in their practice. This aligns with Popham's (2003) perspective that assessment should serve to improve teaching and empower learners.

In Conclusion, this research is grounded in the urgent need to bridge the gap between language assessment theories and their practical application in real classroom settings. The study emphasizes the importance of developing assessments that are not only valid, reliable, and authentic, but also practical and capable of enhancing the overall quality of education. By applying well-established theoretical frameworks in the context of an actual Indonesian junior high school, the research offers valuable insights for improving both English language teaching and assessment practices in similar educational environments

METODE

Analyzing fourth language skills tests into fifth principle, here the structure explanation.

A. Practicality

involves the ease with which an assessment can be implemented within the constraints of the educational environment. This includes considerations such as time, budget, materials, and administrative support. According to Suherdi and Marlina (2022), practicality ensures that assessments are accessible and manageable, especially in resource-limited settings. An assessment that is theoretically sound but difficult to administer or score due to logistical barriers fails to be practical. For example, a test requiring advanced technology or extensive teacher training may not be feasible in rural schools. Therefore, educators must strive to balance ideal assessment design with real-world classroom conditions.

1. Coast

Cost as term define that the test should not be too expensive to conduct, the cost of the test must stay within the budget and avoid conducting a test that requires

excessive budget. Noted: Students didn't have to pay for the tests; meanwhile, we, as the responsibility for the test should prepare printing of paper sheets.

2. Time

Time accessibility to test students begun in the first day on Monday until to Tuesday, it means took three days off to finish the process of tests. For the specific time start at 10:00-12.30 a.m. in this case students took three hours thirty minutes to complete speaking test.

Analysis from time management:

$150 \text{ minutes} \div 45 \text{ questions} = \pm 3,3 \text{ minutes per question}$ Multiple choice (30 questions): reasonable to give time 1,5–2 minutes/questions → required $\pm 45\text{--}60$ minutes

Essay test (15 questions: because the students low average in English subject, reasonable to give students time 4–5 minutes/questions → required $\pm 60\text{--}75$ minutes

The properly time: 45 to 60 minutes (MCQ) + 60 to 75 (essay) = 105 to 135 minutes or 1 hours 45 minutes – 2 hours 15 minutes so, totally fine took 2 hours 30 minute less and consistently reasonable , even it have more a space for students to ask something from test that cannot be understood, and help students not to rush.

3. Administration

In the context of administration, this term refers to the design of tests that are structured simply, with clear questions that are not overly complicated or complex to administer during classroom test preparation. This means that students have the right to feel comfortable throughout the learning process. Therefore, in this test project, students were given the freedom to express themselves and ask questions.

Illustration: Students with lower proficiency in language skills often have no idea how to answer certain questions and end up asking the teacher. In such cases, students struggle to find the correct answers, which can lead to wasted time and disrupt time management during the test process.

As a solution, students should be guided and directed to develop a deeper understanding of the function of each language skill and to cultivate curiosity in learning.

4. Scoring and evaluation

The scoring and evaluation process must be aligned with the available time and allocation. This is especially important in evaluating multiple-choice and essay tests to ensure the results are obtained clearly and efficiently within the designated time frame. The assessor, as the person in charge, is expected to evaluate each section of the test for every language skill individually, accompanied by scoring rubrics and answer keys to support effective time management. For multiple-choice questions, each correct answer was awarded one point, resulting in a total score of 30 points for a perfect result. Meanwhile, the essay tests were scored using a range of 6 and 4 points, based on question complexity and following a holistic rubric. 6 points were given for well-developed answers to more complex questions, while four points were assigned for correct but simpler responses.

In the listening section, students were assessed based on their ability to receive and understand information from written scripts adapted from the textbook used in the Merdeka Curriculum for Grade VIIIB. In the speaking section, students were asked to answer questions based on printed dialogues. Their performance was evaluated on their ability to communicate, pronounce each word correctly, and deliver their answers clearly. However, most students faced significant difficulties in speaking English. Even so, being given the chance to try boosted their self-confidence.

As for reading and writing, the evaluation revealed that many students were still very weak in these areas. Instead of attempting to read the questions and write their answers, some students gave up quickly without making an effort. This indicates the need for further support and improvement in reading and writing instruction.

B. Validity

is regarded as the cornerstone of any effective assessment, as it reflects the degree to which the test measures what it is intended to measure. Setiyadi (2021) explains that validity encompasses several dimensions, including content validity (alignment with instructional objectives), construct validity (accurate representation of the targeted language skills), and criterion-related validity (correlation with external

measures of proficiency). Without validity, even a highly reliable and practical test lacks educational meaning. For instance, a vocabulary test that focuses only on spelling may not accurately reflect a student's lexical knowledge.

C. Authenticity

Page | 22

refers to how well an assessment replicates real-life language use. Authentic assessments present learners with tasks that resemble communicative situations outside the classroom, thereby making the learning experience more relevant and engaging. According to Rukmini and Arifin (2024), tasks such as composing emails, engaging in interviews, or interpreting news articles help bridge the gap between classroom learning and real-world language application. Authenticity enhances learner motivation and encourages the development of functional language skills, rather than isolated knowledge of grammar or vocabulary.

D. Washback

(or backwash) describes the influence that assessment has on teaching and learning behaviors. Positive washback encourages practices that align with meaningful learning, such as critical thinking, collaboration, and self-reflection. Conversely, negative washback may result in teaching to the test, surface learning, or test anxiety. Pratiwi and Suryana (2023) highlight that well-designed assessments contribute to constructive washback by promoting instructional alignment, enhancing teacher reflection, and guiding learners to focus on language development rather than mere test performance. For example, a writing test that values clarity and creativity can inspire students to write more thoughtfully in other contexts as well.

1. In conclusion, the five principles practicality, reliability, validity, authenticity, and washback form a comprehensive framework that underpins the design of effective language assessments. Each principle contributes to the educational value of the assessment process by ensuring that tests are not only accurate and fair but also meaningful and conducive to language learning. Recent research reaffirms the relevance of these principles in modern classrooms, particularly as

educators strive to adapt assessments to diverse learner needs and evolving pedagogical paradigms. By integrating these principles, language educators can construct assessment tools that enhance both the accuracy of language measurement and the quality of the teaching and learning process

HASIL DAN PEMBAHASAN

The findings of this research indicate that students' performance in assessing the four core English language skills listening, speaking, reading, and writing was largely below the desired proficiency level. Among the 28 students who took part, the majority earned low scores, with the weakest outcomes observed in the speaking and writing components. Many learners struggled to comprehend the instructions provided in the test and needed further clarification and support during the assessment process.

Each language skill was evaluated using a mix of 30 multiple-choice questions and 15 essay-based items, resulting in a total of 180 questions across all four skills. The best results were typically recorded in listening and reading sections, while performance in the productive skills speaking and writing was significantly lower. This poor performance is attributed to the students' lack of familiarity with test formats presented entirely in English, along with their limited practice in answering open-ended questions.

The overall test outcomes were then analyzed based on the five fundamental principles of language assessment as outlined by Bachman and Palmer, which include practicality, reliability, validity, authenticity, and washback.

a. Practicality

Practicality in language assessment refers to the ability to design, administer, and grade a test efficiently, without requiring excessive time, cost, or special equipment, while still fulfilling its educational purpose. According to Hughes (2003), a practical test should not be too expensive, complicated to administer, or overly time-consuming, yet it must effectively measure students' language abilities. This principle was applied in the

development and administration of the English test for Class VIII-B at SMP Negeri 1 Dolok Panribuan. The test assessed the four core language skills reading, listening, writing, and speaking and included an equal number of questions: 30 multiple-choice items and 15 essay questions for each skill. All materials were created using simple classroom tools, such as printed papers, pens, and standard school printers, without relying on advanced technology or internet access. This ensured the assessment was affordable and matched the school's available resources. A clear example of practicality was seen in the listening section. While many schools use audio recordings and sound systems for listening tests, this particular class lacked such equipment. Instead, the teacher read the listening passages aloud, ensuring all students heard the same material. This simple yet effective solution demonstrated that practical assessment can be achieved by adapting to classroom constraints without compromising the quality or fairness of the test.

The number of test items was carefully planned to fit within the regular class periods, allowing students to complete each section during standard lesson time. While most students finished the multiple-choice questions quickly, some found the essay sections particularly writing and speaking more time-consuming due to their lack of confidence in using English. These students often needed extra time to translate their thoughts or struggled to express ideas clearly, sometimes switching back to Bahasa Indonesia. Nevertheless, with guidance and support from the teacher, nearly all students were able to complete the exam on time. From the teacher's perspective, grading the test was straightforward and manageable. The multiple-choice questions were evaluated using an answer key, ensuring a quick and objective grading process. For the essay sections, a simple rubric was applied. In the reading and listening sections, students were expected to respond in English, and each correct short answer was awarded 4.66 points. However, in the writing and speaking sections, many students responded in Bahasa Indonesia due to limited English proficiency. To accommodate this, a simplified scoring system was used two points for full and relevant answers, and one point for partially correct or off-topic responses. This approach allowed for efficient and fair grading while also supporting students at varying levels of language proficiency. Furthermore, the test design is easy to replicate and adapt for other classes or future practice sessions. Teachers can maintain the same format while updating topics or texts to align with students' needs

and curriculum changes, making the test both practical and sustainable over time. Some students reported feeling fatigued during the test, especially when encountering unfamiliar vocabulary. This highlights the need for better test preparation, such as practicing integrated language skills and ensuring that test instructions are clearly explained beforehand. These steps can help reduce anxiety and improve student performance in future assessments.

In conclusion, the English test served as a strong example of practicality in assessment. It was designed to fit normal classroom conditions, requiring no advanced technology and minimal resources, while remaining efficient in both administration and grading. Despite these limitations, the test provided valuable insights that supported the teaching process and aligned with the school's context. This demonstrates that practical assessments must always take into account the available resources, the specific needs of students, time constraints, and the learning objectives being targeted.

b. Reliability

Reliability in assessment refers to the consistency of test results, which is especially important when grading subjective tasks such as essays and oral presentations. According to Hughes (2011), a reliable test should yield similar results regardless of when it is administered, who scores it, or the testing conditions, provided the test format and grading criteria remain the same. This ensures that a student's score accurately reflects their ability and is not influenced by external factors or random variations. In language assessment, reliability becomes particularly crucial for evaluating skills like writing and speaking, where open-ended responses require graders to make subjective judgments. To maintain fairness, it is important that different graders produce similar scores for the same performance a concept known as inter-rater reliability. This minimizes bias and ensures equitable treatment for all students.

In this study, inter-rater reliability was essential in evaluating the accuracy and objectivity of scores in the Writing and Speaking sections. Unlike Reading and Listening, which consisted of objective formats such as multiple-choice and short-answer questions that

could be marked using an answer key, Writing and Speaking involved performance tasks that demanded interpretation. Therefore, a detailed scoring plan was necessary to ensure consistency among graders. By applying well-structured rubrics, involving multiple trained raters, and establishing clear grading procedures, the assessment achieved a high level of reliability. The minimal variation in scores between different graders demonstrated that the system was fair and dependable. This not only enhanced the reliability of the assessment but also strengthened its credibility as an effective tool for measuring students' language proficiency.

Show the Santun's assessment method result

Skill	Averages Score	Highest Score	Lowest Score
Reading	61.70	84	36
Listening	64.13	80	39
Writing	52.74	72	33
Speaking	66.00	89	39

The table summarizes the average score, highest score, and lowest score across the four core English language skills: Reading, Listening, Writing, and Speaking. This overview provides a clear picture of student performance patterns and highlights both strengths and areas needing improvement within the group.

Among the four skills, Speaking has the highest average score at 66.00, indicating that most students feel more confident and capable when expressing their thoughts verbally. This could be the result of interactive classroom practices such as discussions, presentations, or dialogue-based activities that allow students to practice spoken English more frequently. The highest Speaking score of 89 further confirms this strength, although the lowest score remains at 39, suggesting that not all students are equally confident or proficient in this area.

Listening follows closely with an average score of 64.13, which reflects a generally good comprehension of spoken English among the students. The highest Listening score is 80, showing that some students excel in this receptive skill. However, the

lowest score of 39 points to a group of learners who may need targeted listening practice, such as using audio materials with guided tasks.

Reading shows a slightly lower average at 61.70, but still indicates a moderate level of comprehension. The top score of 84 demonstrates that certain students have strong reading skills, likely aided by a wide vocabulary and the ability to understand context. Meanwhile, the lowest reading score of 36 suggests that some students struggle with reading texts in English and may require support in building reading strategies.

Writing has the lowest overall average at 52.74, marking it as the most challenging skill for the majority of students. This is further supported by the lowest individual score of 33, indicating significant difficulties in constructing written texts in English. Writing typically requires mastery of grammar, organization, vocabulary, and coherence, which may explain the overall lower performance in this skill area.

In conclusion, the data suggests that while students are generally stronger in Speaking and Listening, more attention needs to be given to improving Writing skills. It is important for instruction to be balanced and tailored to address the specific needs of learners in each skill domain, ensuring holistic development in English language proficiency.

c. Validity

Validity is very important when evaluating the quality of a language test. A valid test accurately measures what it is intended to measure namely, English language proficiency. Hughes (2011) defines validity as the extent to which test results are supported by evidence and theoretical justification for how the test is used. This research examined two main types of validity: content validity and construct validity, both essential to ensure that the test is accurate and meaningful. Content validity refers to how well the test items match the materials taught, particularly those outlined in the *Kurikulum 2013* for junior high schools. In this study, the reading sections included familiar texts such as announcements, stories, and invitations, commonly found in student textbooks. The listening sections used realistic content like short talks and announcements, read aloud by

the teacher due to limited audio equipment. Writing tasks asked students to describe daily routines or write simple messages, and the speaking tasks involved basic conversations and self-introductions appropriate for the junior high school level.

Construct validity, on the other hand, ensures that each part of the test actually measures the specific skill it is intended to assess. In this study, both raters, Santun and Junita, used standardized scoring rubrics to maintain fairness and consistency. Writing was evaluated based on content relevance, organization, and clarity, while speaking was assessed in terms of fluency, pronunciation, vocabulary, and language appropriateness. The results showed that students who demonstrated strong skills in identifying information in texts or writing clear sentences earned higher scores, while those who struggled scored lower. This confirmed that the test accurately measured the intended language skills. By using consistent scoring guides and fair assessment practices, the test ensured trustworthy and valid results, reinforcing its credibility as a reliable measure of students' English proficiency.

d. Authenticity

In language assessment, authenticity refers to how closely test tasks resemble real-life communication and situations that learners are likely to encounter. An assessment is considered authentic when it engages students in activities that reflect the actual use of language in everyday settings, such as at school, at home, or during social interactions. In this study, authentic elements were intentionally integrated into all four language skill areas. First, the tasks were based on real-life topics and familiar contexts that students could easily relate to. For instance, the Reading and Listening sections featured content such as school schedules and casual invitations. The Writing tasks asked students to describe daily routines or talk about their favorite foods, while the Speaking activities involved describing familiar people, like friends or family members. These tasks not only aligned with the goals of the curriculum but were also designed to be meaningful and personally relevant to students. Reading materials included formats like school announcements, birthday messages, and short narratives. Listening items mimicked natural conversations or typical school announcements. Writing prompts encouraged

students to reflect on and share their own experiences, while speaking prompts invited them to introduce themselves or discuss their hobbies and interests.

Second, the assessment emphasized functional communication by encouraging students to use language with a clear communicative purpose such as describing, explaining, or expressing preferences. Even when students occasionally used their first language (Bahasa Indonesia), especially in writing and speaking, they still demonstrated authentic communication driven by a specific intention. For example, a student named Jogi delivered a speaking response in Bahasa Indonesia that reflected a real effort to introduce herself and describe personal interests: “Halo, nama saya Nayra. Umur saya 14 tahun. Saya suka makan nasi goreng. Hobi saya adalah memasak dan membaca. Saya orangnya cantik dan pintar.” Although the response was not entirely in English, it maintained authenticity by fulfilling the communicative goal of self-introduction. This illustrates that authenticity in assessment is not only about language form but also about engaging students in meaningful, purposeful communication.

e. Washback

Washback refers to how assessment affects teaching and learning, including how students study and how teachers plan their lessons. A good language test not only measures student performance but also provides useful feedback that can influence how they learn in the future. Positive washback can increase student motivation and engagement and help teachers improve their teaching. On the other hand, negative washback may lead to anxiety, lack of interest, or reluctance to learn.

In this study, the washback effect was observed during a practice test involving 23 eighth-grade students at SMP Negeri 1 Panribuan. The test assessed all four language skills reading, listening, writing, and speaking using multiple-choice and essay questions. Student reactions to the tasks varied. Some students were highly engaged and completed all parts of the test seriously, especially tasks related to personal and familiar topics like family, food, and hobbies. Even when they answered in Bahasa Indonesia, these students still showed a clear effort to express themselves and complete the tasks enthusiastically.

However, many students experienced difficulties such as confusion, lack of interest, and low motivation, often caused by limited English proficiency or unfamiliar vocabulary in the test. Many students asked for help translating words or understanding instructions, especially in the reading and writing sections. Additionally, the time limit and number of questions proved challenging, causing some students to be unable to finish all parts of the test.

Despite these issues, the test led to some positive learning outcomes. Some students became aware of areas they needed to improve, felt satisfied when they answered correctly, and attempted to complete written responses even if they used their first language. These actions indicate that the test encouraged students to reflect on their abilities and participate actively in the assessment, which are signs of positive washback.

From the teacher's perspective, the assessment was not only a tool for measuring student achievement but also for diagnosing strengths and weaknesses. The test results revealed several common issues, such as limited vocabulary, grammatical errors, and overreliance on Bahasa Indonesia to comprehend and express ideas. The assessment also showed that students need more exposure to English, particularly in listening, where they struggled the most.

The test data helped the teacher identify specific students who needed extra support, for example, with fluent writing and reading comprehension. This enabled the teacher to design more targeted lessons that addressed the real learning needs shown in the students' performance. The test generated both positive and negative washback. On the positive side, it raised student awareness of their language skills, motivated them to try hard, and gave the teacher a clearer picture of learning challenges at both the class and individual levels. This is helpful in creating more effective and responsive teaching.

However, the assessment also had some unintended negative effects. Some students lost interest because they found the tasks too difficult. Others felt anxious when they encountered unfamiliar vocabulary or couldn't finish all sections within the given time. This sometimes led them to depend too much on the teacher rather than learning to solve problems and use the language independently.

In conclusion, the test had both positive and negative washback, but overall, its impact was beneficial. It encouraged students to reflect on their learning and helped teachers become more aware of student needs. It also served as a strong foundation for adapting and improving classroom teaching based on students' real challenges.

To enhance positive washback in future assessments, several improvements can be made. Test formats should be more concise, with instructions that are simple and easy to understand. Teachers should use scaffolding techniques or provide examples before assessments, especially to support less proficient students. Classroom learning should also regularly feature real-life English contexts and situations relevant to students' daily lives. By applying these strategies, assessments can become more than just tools for grading they can support more meaningful and goal-oriented learning and teaching.

PENUTUP

This study found that it is both possible and important to assess the four English language skills listening, speaking, reading, and writing using the five basic principles of language assessment: practicality, reliability, validity, authenticity, and washback. The findings showed that applying these principles in real classroom settings helps improve students' language learning. Conducted with 23 eighth-grade students at SMP Negeri 1 Dolok Panribuan, the study revealed that each principle was effectively implemented. Practicality was achieved by using tests that matched the school's limited resources, such as teacher-led listening assessments and printed materials. Reliability was ensured through a consistent scoring system involving multiple raters, especially for writing and speaking tasks. Validity was maintained by aligning test items with the 2013 Curriculum and measuring genuine communication skills. Authenticity was reflected in tasks like writing messages or understanding announcements, which closely resembled real-life language use. Washback was evident as students became more aware of their strengths and areas for improvement, while teachers gained valuable feedback to inform future lesson planning. Score analysis showed that students performed well in speaking and listening but needed additional support in writing, particularly in structure and coherence. Overall, this integrated approach to assessing language skills proved effective and can

serve as a model for other schools facing similar resource constraints. Furthermore, the study emphasizes that assessment should be an ongoing part of the learning process, not a separate task, and should serve as a tool for reflection, feedback, and instructional development. Teachers are encouraged to create thoughtful assessments using detailed rubrics and real-world tasks. Students should view assessments as opportunities for growth and skill development. Future researchers and pre-service teachers are advised to apply theoretical knowledge in diverse contexts, gain practical experience in assessment design, and collaborate with mentors to deepen their assessment literacy. In conclusion, successful language assessment requires commitment and collaboration from all stakeholders to enhance teaching practices, empower learners, and improve assessment methods in Indonesian classrooms and beyond

REFERENSI

- Ahmadi, S., & Hashemi, M. (2021). *Assessment in language learning: Concepts and classroom practices*. Language Education Press.
- Rahimi, M. (2022). *Language evaluation in EFL contexts: A holistic approach to teaching and assessment*. Iranian Journal of Applied Linguistics, 45(2), 101–115. Page | 33
- Nabavi, R., & Yazdani, M. (2021). *Diagnostic assessment in English language learning: Tools for identifying learner needs*. Journal of Language Teaching and Research, 12(4), 635–642. https://doi.org/10.17507/jltr.1204.12
- Baleghizadeh, S., & Dadashi, M. (2023). *Accountability and assessment in educational systems: A policy analysis perspective*. Educational Measurement Quarterly, 9(1), 47–59.
- Brown, H. D. (2022). *Language assessment: Principles and classroom practices (2nd ed.)*. Pearson Education.
- Mohammadi, L. (2021). *Using diagnostic testing to identify reading difficulties among EFL learners*. International Journal of English Linguistics, 11(2), 85–92. https://doi.org/10.5539/ijel.v11n2p85
- Brown, H. D., & Abeywickrama, P. (2020). *Language assessment: Principles and classroom practices (2nd ed.)*. Pearson Education.
- Zarei, A. A., & Esfandiari, R. (2021). *The role of assessment in shaping student motivation and autonomy*. Journal of Educational Studies, 8(3), 134–149.
- Putra, A., & Hasanah, U. (2021). *Rethinking listening instruction in Indonesian EFL classrooms*. Indonesian Journal of English Language Teaching, 16(1), 22–33.
- Ningsih, T. R., & Sari, N. M. (2023). *Enhancing speaking skills through interactive classroom techniques*. Journal of English Language Pedagogy, 4(2), 55–66.
- Mulyadi, D., Suryani, E., & Fadhillah, I. (2022). *Building reading comprehension skills in EFL contexts*. Asian Journal of Language Education, 10(1), 98–110.
- Fitriyani, N., & Asri, D. (2021). *Addressing writing difficulties in junior high school EFL classrooms*. TEFLIN Journal, 32(2), 215–228. https://doi.org/10.15639/teflinjournal.v32i2/215-228
- Suherdi, D., & Marlina, R. (2022). *Practicality issues in English assessment in Indonesian schools*. Indonesian Journal of Applied Linguistics, 12(1), 45–59. https://doi.org/10.17509/ijal.v12i1.35512
- Fauzi, I., & Nurhayati, S. (2023). *Developing reliable assessment tools for EFL learners: A practical guide*. Journal of Language Assessment and Evaluation, 5(3), 89–102.
- Setiyadi, A. B. (2021). *Language assessment and validity in the EFL classroom*. Graha Ilmu.

- Rukmini, D., & Arifin, Z. (2024). *The role of authentic assessment in promoting communicative competence*. Journal of Language and Education Innovation, 7(1), 17–30.
- Pratiwi, N. D., & Suryana, Y. (2023). *Washback effect in English assessment: A case study of junior secondary schools*. Journal of Educational Research and Practice, 9(2), 119–135.