Child Education Journal

Child Education Journal Vol. 8 No. 2 (2026) 229-258

e-ISSN: 2685-1903 p-ISSN: 2087-0396

DOI: 10.33086/cej.v8i2.8830

THIS STUDY RESEARCH ARTICLE

Diagnostic Assessment in Pre-Service Language Teacher Education: Evidence from Prospective Teachers and Secondary-School Learners

Mukhzamilah[1*], and Shigenori Wakabayashi[2]

[1]Universitas Negeri Surabaya, Indonesia; ,[2]Chuo University, Japan

Correspondence: mukhzamilah@unesa.ac.id

Article History: Received: Mar 18, 2026 • Revised: Jul 25, 2026 • Accepted: Aug 15, 2026 • Published: Aug 30, 2026

ABSTRACT

Diagnostic assessment can play a dual role in pre-service teacher education by identifying prospective teachers' learning needs while strengthening their capacity to respond to language difficulties encountered in school settings. This study examined the use of diagnostic assessment in the language development of 30 second-semester students enrolled in the Indonesian Language and Literature Education Study Program at Universitas Negeri Surabaya, supported by an error analysis of 30 descriptive texts written by seventh-grade students at a junior high school in Surabaya. A descriptive qualitative design with purposive sampling was employed. Data were collected through diagnostic tests covering listening, speaking, reading, and writing, as well as classroom observation, interviews, and documentation, and were analyzed using the interactive model of Miles, Huberman, and Saldana. The findings identified four major areas of difficulty among the prospective teachers: standard spelling and grammar (43.3%), academic vocabulary (36.7%), paragraph coherence (50.0%), and speaking confidence (40.0%). Supplementary school-level analysis recorded 505 language errors, 87.9% of which involved spelling, with capitalization as the most frequent category. Diagnostic findings informed differentiated instruction, scaffolding, and formative feedback during a twelve-week instructional period. The mean post-test score increased by 18.4 points, while qualitative evidence indicated greater learning autonomy, speaking confidence, and metalinguistic awareness. These findings suggest that diagnostic assessment can support responsive instruction in pre-service language teacher education while strengthening prospective teachers' assessment literacy for addressing language-learning needs in secondary-school contexts.

Copyright and License
© 2026 The Author(s). Published by Universitas Nahdlatul Ulama Surabaya, Indonesia.
This article is distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).
License: https://creativecommons.org/licenses/by-sa/4.0/
Journal License Policy: https://journal2.unusa.ac.id/index.php/CEJ/License

ABSTRAK

Asesmen diagnostik dapat memainkan peran ganda dalam pendidikan calon guru, yaitu mengidentifikasi kebutuhan belajar calon guru sekaligus memperkuat kapasitas mereka dalam merespons kesulitan berbahasa yang ditemui di lingkungan sekolah. Penelitian ini mengkaji penggunaan asesmen diagnostik dalam pengembangan kemampuan berbahasa 30 mahasiswa semester dua Program Studi Pendidikan Bahasa dan Sastra Indonesia di Universitas Negeri Surabaya, yang didukung oleh analisis kesalahan terhadap 30 teks deskriptif karya siswa kelas tujuh di sebuah SMP di Surabaya. Penelitian ini menggunakan desain kualitatif deskriptif dengan teknik pengambilan sampel bertujuan (*purposive sampling*). Data dikumpulkan melalui tes diagnostik yang mencakup keterampilan menyimak, berbicara, membaca, dan menulis, serta melalui observasi kelas, wawancara, dan dokumentasi; data kemudian dianalisis menggunakan model interaktif Miles, Huberman, dan Saldana. Hasil penelitian mengidentifikasi empat bidang kesulitan utama pada calon guru: ejaan dan tata bahasa baku (43,3%), kosakata akademik (36,7%), koherensi paragraf (50,0%), dan kepercayaan diri dalam berbicara (40,0%). Analisis tambahan di tingkat sekolah mencatat 505 kesalahan berbahasa, dengan 87,9% di antaranya berkaitan dengan ejaan, di mana penggunaan huruf kapital merupakan kategori yang paling sering ditemukan. Temuan diagnostik menjadi dasar bagi penerapan pembelajaran berdiferensiasi, *scaffolding*, dan umpan balik formatif selama periode pembelajaran dua belas minggu. Skor rata-rata *post-test* meningkat sebesar 18,4 poin, sementara bukti kualitatif menunjukkan peningkatan otonomi belajar, kepercayaan diri dalam berbicara, dan kesadaran metalinguistik. Temuan ini mengindikasikan bahwa asesmen diagnostik dapat mendukung pembelajaran yang responsif dalam pendidikan calon guru bahasa sekaligus memperkuat literasi asesmen calon guru dalam menangani kebutuhan pembelajaran bahasa di konteks sekolah menengah.

How to cite: Mukhzamilah, M., & Wakabayashi, S. (2026). Diagnostic Assessment in Pre-Service Language Teacher Education: Evidence from Prospective Teachers and Secondary-School Learners. *Child Education Journal*, *8*(2), 229-.258 https://doi.org/10.33086/cej.v8i2.8830

Keywords: diagnostic assessment; pre-service teacher education; prospective language teachers; assessment literacy; secondary-school learners; responsive instruction

1. INTRODUCTION

Protracted Preparing prospective language teachers in contemporary teacher education requires more than strengthening their personal competence in listening (Cen et al., 2026), speaking (Hendratno et al., 2026), reading (Cui et al., 2026), and writing (Zhong & Hao, 2026); it also requires developing their capacity to interpret assessment evidence and translate it into pedagogically responsive decisions for future learners. Within international and Indonesian teacher-education contexts (Al Khaldy et al., 2026), language assessment literacy has therefore become an important component of professional preparation because teachers are expected not only to administer assessments but also to understand learners' strengths and difficulties (Mottafegh & Ahn, 2026), interpret results responsibly (Pastukh et al., 2026), provide meaningful feedback (Sepúlveda-Oviedo et al., 2025), and redesign instruction accordingly (Chen & Yang, 2026; Rasoolzadeh et al., 2025). Recent studies have shown that assessment literacy is multidimensional and context-dependent and that prospective teachers need opportunities to connect theoretical knowledge of assessment with authentic pedagogical practice. Indonesian studies likewise indicate that pre-service

language teachers may perceive themselves as reasonably prepared for classroom assessment while still demonstrating uneven knowledge and practical assessment competence, suggesting that assessment should be experienced as part of the learning process rather than treated only as a technical course component (Anam & Putri, 2021; Farmasari et al., 2023; Prastikawati et al., 2024). Diagnostic assessment is particularly relevant to this professional demand because it positions assessment as evidence for identifying learning needs before determining subsequent instruction. Accordingly, its significance extends beyond measuring current achievement toward supporting prospective teachers' development as reflective and responsive future educators (du Plessis, 2025; Hagen & Rydland, 2026; Spear et al., 2025).

Despite this importance, the major challenge lies in the persistent distance between assessment knowledge, actual language competence, and the ability to use assessment evidence pedagogically. Pre-service language teachers enter higher education with heterogeneous literacy experiences and proficiency profiles, while teacher-education programs must simultaneously develop their disciplinary competence and prepare them to recognize similar differences among school learners. Studies of Indonesian pre-service teachers have reported moderate or insufficient levels of language assessment literacy, difficulties translating assessment knowledge into classroom practice, and gaps between perceived preparedness and demonstrated assessment competence (Anam & Putri, 2021; Sulistio, 2023). The present manuscript reflects this challenge in its this study empirical context: preliminary observations among second-semester students of the Indonesian Language and Literature Education Study Program at Universitas Negeri Surabaya indicated differences in academic writing, oral presentation, vocabulary use, reading comprehension, and prior literacy experience. At the school level, diagnostic difficulties are also evident; recent Indonesian studies show that diagnostic assessment is needed to identify heterogeneous language-learning needs before teachers determine instructional responses, particularly under learner-centered and differentiated approaches (Refoina & Rustam, 2024; Suleman & Idayanti, 2024). Consequently, the central problem is not simply whether students can be assessed, but whether diagnostic evidence can be transformed into differentiated instruction, scaffolding, formative feedback, and other responsive pedagogical decisions appropriate to learners' actual needs.

Previous research has approached this issue through several related but still partially separated strands. Research on language assessment literacy among pre-service teachers has been conducted by Anam and Putri (2021), Öztürk (2022), Sulistio (2023), Farmasari et al. (2023), Alcazaren (2024), Moraes et al. (2024), and Prastikawati et al. (2024). Anam and Putri (2021) identified differences between Indonesian pre-service and in-service teachers' assessment literacy but primarily measured assessment knowledge rather than tracing a diagnosis–instruction cycle. Öztürk (2022) showed that a classroom-assessment course could change prospective teachers' conceptions and perceived competence, although its focus remained on the assessment course itself. Sulistio (2023) explored prospective teachers' perceived academic and professional readiness but relied primarily on interview evidence. Farmasari et al. (2023) examined preparedness following language-assessment coursework, whereas Alcazaren (2024) emphasized reflective experiences within a language-assessment course. Moraes et al. (2024) investigated which classroom tasks foster components of assessment literacy, while Prastikawati et al. (2024) demonstrated improvement in assessment literacy after teaching practicum. Collectively, these studies establish the importance of assessment literacy, yet they largely concentrate on teachers' knowledge (Bruyninckx et al., 2026), perceptions (Levinson et al., 2026), coursework (Duthois et al., 2026), or practicum readiness rather than examining how prospective teachers themselves experience diagnostic mapping of language difficulties followed by differentiated instructional responses. Thus, existing scholarship provides a strong foundation but leaves space for a more integrated examination of diagnosis, pedagogical follow-up, and teacher preparation.

Against this background, the novelty of the present study lies not in proposing an entirely new assessment construct, but in integrating diagnostic assessment, prospective teachers' language development, responsive instruction, and supplementary school-level evidence within one pre-service language teacher-education context. Previous studies have commonly examined assessment literacy as knowledge, perception, preparedness, classroom tasks, or practicum competence (Zapatero-Ayuso et al., 2026). Studies of diagnostic assessment in Indonesian schools, meanwhile, have generally focused directly on school learners and the use of diagnostic results for grouping, instructional adaptation, or improving understanding of subject matter (Refoina & Rustam, 2024; Suleman & Idayanti, 2024). The present study connects these two lines without merging their datasets: it examines diagnostic assessment among 30 second-semester prospective language teachers and uses an error analysis of 30 seventh-grade students' descriptive texts as complementary evidence illustrating language difficulties that future teachers may later encounter in school contexts. This arrangement preserves the methodological distinction between the university and school datasets while offering a pedagogically relevant bridge between teacher preparation and secondary-school language learning. The novelty therefore concerns the diagnose–respond–reflect linkage within pre-service teacher education, supported rather than replaced by school-level evidence. This formulation also avoids overstating causality and is consistent with the descriptive qualitative orientation of the this study research.

A more specific research gap emerges when the findings of these earlier studies are considered together. First, knowledge–practice gaps remain evident: pre-service teachers may possess conceptual understanding of assessment without being equally prepared to interpret evidence and use it for classroom decision-making (Anam & Putri, 2021; Farmasari et al., 2023; Prastikawati et al., 2024). Second, experience gaps remain because research frequently examines assessment courses, beliefs, perceptions, or practicum outcomes rather than prospective teachers' direct experience of being diagnostically assessed and subsequently receiving differentiated instruction, scaffolding, and formative feedback (Öztürk, 2022; Alcazaren, 2024; Moraes et al., 2024). Third, cross-context pedagogical gaps remain between research on assessment literacy in teacher education and diagnostic-assessment research involving school learners; studies such as Refoina and Rustam (2024) and Suleman and Idayanti (2024) demonstrate the usefulness of diagnostic assessment at school level but do not investigate how such evidence can inform the preparation of future language teachers. Fourth, much Indonesian pre-service language-assessment research has focused on English-language education, leaving comparatively less empirical attention to diagnostic assessment within Indonesian Language and Literature teacher education (Anam & Putri, 2021; Sulistio, 2023). The present study addresses these gaps by retaining prospective teachers as its principal participants while using secondary-school evidence only as contextual support for understanding their future assessment responsibilities.

The theoretical orientation of this study is grounded in diagnostic and formative assessment within the broader framework of language assessment literacy, rather than in a newly introduced grand theory that is absent from the this study research design. Diagnostic assessment conceptualizes assessment evidence as information for identifying learners' existing strengths, weaknesses, and learning needs before subsequent instructional decisions are made, whereas formative-oriented assessment emphasizes using evidence during learning to adjust teaching, feedback, and learner support. Contemporary teacher-education studies reinforce this orientation by conceptualizing assessment literacy as a combination of knowledge, skills, principles, interpretation, and the ability to translate evidence into pedagogical action (Atjonen, 2022; Öztürk, 2022). Empirical work with Indonesian pre-service teachers further shows that professional readiness requires more than familiarity with testing terminology; teachers must develop practical competence in administering, interpreting, scoring, communicating, and using assessment information (Anam & Putri, 2021; Prastikawati et al., 2024). Within the present study, this theoretical orientation is consistent with the this study use of diagnostic evaluation to map listening, speaking, reading, and writing

difficulties and to use these findings as the basis for instructional follow-up. The framework therefore links three processes already contained in the manuscript—diagnosis, pedagogical response, and learner development—without altering the this study participants, intervention, data, or analytical procedure.

Conceptually, the study employs several interconnected constructs already present in this study: diagnostic assessment, language proficiency, differentiated instruction, scaffolding, formative feedback, language assessment literacy, and error analysis. Diagnostic assessment refers to the systematic identification of learner strengths and difficulties, while language proficiency in this research is represented through listening, speaking, reading, and writing performance. Differentiated instruction denotes instructional responses adjusted to learners' identified profiles; scaffolding refers to structured support that is gradually reduced as learners become more autonomous; formative feedback provides information about strengths, weaknesses, and specific directions for improvement. Recent studies show that differentiated and formative approaches are most meaningful when assessment evidence is actively used to adapt instruction rather than merely record achievement (Kristiyuana et al., 2025; Asriadi et al., 2025). Language assessment literacy connects these processes to the professional capacity of prospective teachers to understand, design, interpret, and use assessment appropriately, while error analysis supplies complementary evidence regarding patterns of school learners' language difficulty (Anam & Putri, 2021; Farmasari et al., 2023). Importantly, these concepts are not treated as independent additions to the manuscript; rather, they articulate processes already represented in the this study data, including diagnostic tests, differentiated grouping, scaffolding activities, formative feedback, qualitative observation, and supplementary analysis of seventh-grade descriptive texts.

What makes the present research particularly important is the pedagogical position occupied by its principal participants. The 30 second-semester students are simultaneously language learners and prospective language teachers: their current difficulties in writing, speaking, vocabulary, and language conventions constitute immediate learning needs, while their future professional roles will require them to identify comparable difficulties among school learners. Contemporary research shows that early and practice-oriented exposure to assessment can strengthen prospective teachers' awareness of assessment purposes and help bridge the divide between theoretical preparation and classroom application (Öztürk, 2022; Farmasari et al., 2023). At the same time, diagnostic studies involving Indonesian school learners demonstrate that heterogeneous language difficulties require teachers to identify learner needs before selecting instructional responses (Refoina & Rustam, 2024; Suleman & Idayanti, 2024). The supplementary analysis of 30 seventh-grade descriptive texts in the present manuscript is therefore pedagogically meaningful because it documents 505 language errors, predominantly spelling-related, while remaining analytically separate from the university sample. The interest of the study lies in connecting these two educational realities without statistically equating them: prospective teachers experience diagnostic-responsive instruction in higher education, while school-level evidence illustrates authentic language difficulties for which those future teachers will eventually require assessment literacy and responsive pedagogical competence.

This study is situated within applied linguistics, particularly at the intersection of diagnostic assessment, language development, and pre-service teacher education. Diagnostic assessment is relevant not only because it identifies areas of linguistic difficulty, but also because it provides evidence that can be translated into pedagogical decisions concerning scaffolding, differentiated instruction, and formative feedback. In this sense, assessment is treated as part of the language-learning process rather than as a separate measurement activity. For prospective language teachers, this relationship has additional professional significance because they are simultaneously developing their own language competence and learning how assessment evidence can inform future classroom decisions. Recent research on language assessment

literacy has shown that pre-service teachers often possess uneven levels of assessment knowledge and preparedness, particularly when theoretical understanding is not accompanied by opportunities to interpret real learner performance and make instructional decisions (Anam & Putri, 2021; Atjonen et al., 2022; Farmasari et al., 2023; Alcazaren, 2024; Moraes et al., 2024; Prastikawati et al., 2024). The present study therefore examines diagnostic assessment as both a language-development process and a professional learning experience within pre-service language teacher education (Francisca Carocca, 2026; Hsin et al., 2025; Mullah & Wan Zainon, 2026)

2. RESEARCH METHODS

2.1. Research Design

This study employed a descriptive qualitative design to examine how diagnostic assessment informs responsive instructional practices in pre-service language teacher education (Creswell & Poth, 2018; Miles et al., 2014). The research followed three interconnected stages: diagnosis, instructional follow-up, and evaluation. Qualitative data were used to explore participants' language difficulties, instructional responses, classroom experiences, and perceived development, while diagnostic scores, difficulty percentages, and pre-test–post-test differences provided descriptive supporting evidence. As the study lacked random assignment and a control group, numerical changes were interpreted as observed progress rather than causal effects of diagnostic assessment. A supplementary school-level dataset contextualized language difficulties in secondary education but was analyzed separately from the university-level sample. The methodological sequence is presented in Figure 1.

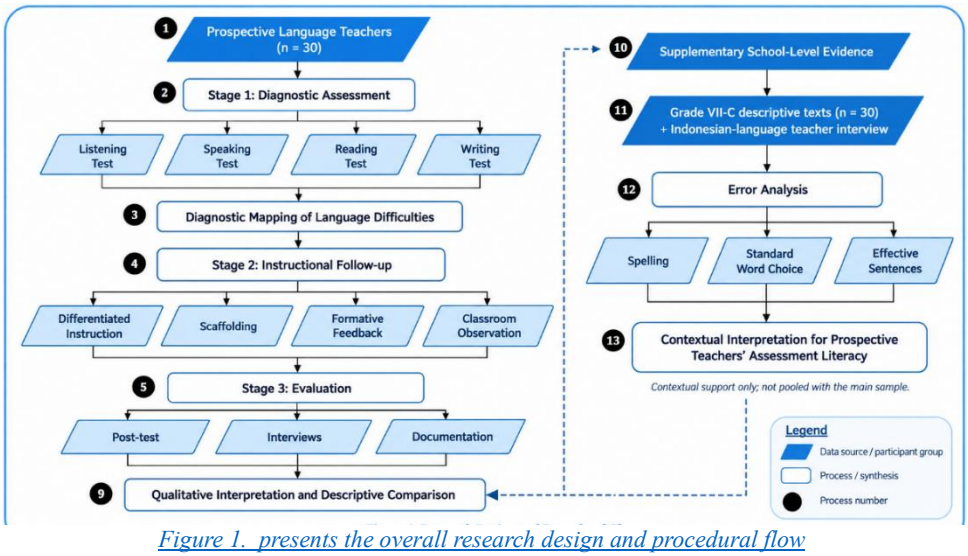


Figure 1. presents the overall research design and procedural flow

Source: Authors' research design based on the empirical procedures of the study, 2026.

As shown in Figure 1, the prospective-teacher dataset followed a complete diagnose–intervene–evaluate sequence, whereas the secondary-school dataset followed a separate error-analysis pathway. The two streams were connected only during pedagogical interpretation. Accordingly, the Grade VII data were not

treated as an intervention group, were not statistically pooled with the university data, and were not used for direct numerical comparison.

2.2. Research Participants and Setting

ThisThe principal participants were 30 second-semester students enrolled in the Indonesian Language and Literature Education Study Program, Faculty of Languages and Arts, Universitas Negeri Surabaya, during the 2025/2026 academic year. Within the context of this study, these participants were positioned as prospective language teachers because they were enrolled in a teacher-education program while simultaneously developing their own language competence. Purposive sampling was employed using three inclusion criteria: (1) active enrollment as second-semester students, (2) participation in the Academic Indonesian and Oral Language Skills courses, and (3) willingness to participate in all stages of the research. The study was conducted at the Lidah Wetan campus of Universitas Negeri Surabaya, Surabaya, East Java. Data collection and instructional follow-up were conducted over twelve weeks from February to April 2026.

In addition to the principal participants, supplementary school-level evidence was obtained from a separate study conducted by an undergraduate student under the supervision of the first author (Dewi, 2025). This dataset consisted of 30 descriptive texts written by Grade VII-C students at SMP Negeri 34 Surabaya during the odd semester of the 2025/2026 academic year, together with an interview with the class's Indonesian-language teacher. The class was selected purposively, and each of the 30 students contributed one descriptive text. The secondary-school participants differed from the university participants in educational level, task, instructional context, and time frame. Therefore, the school-level data were treated exclusively as supplementary contextual evidence. This distinction preserved the integrity of the two datasets while allowing the secondary-school findings to illuminate the types of language difficulties that prospective teachers may later encounter in their professional practice.

2.3 Research Procedures

The research procedures followed three sequential stages: diagnostic assessment, instructional follow-up, and evaluation. Each stage served a different function while remaining connected within the overall diagnostic-assessment cycle.

Stage 1: Diagnostic Assessment

At the beginning of the study, diagnostic assessment was conducted to identify participants' language-proficiency profiles across four domains: listening, speaking, reading, and writing. The listening test required participants to identify main ideas and detailed information from academic speech recordings. The speaking assessment consisted of a five-minute mini-presentation. The reading assessment measured comprehension of popular scientific texts, while the writing assessment required participants to produce a 500-word argumentative essay. The results of these assessments were organized into diagnostic profiles showing areas of relative strength and difficulty. The purpose of this stage was not simply to generate scores but to provide evidence for determining the types and levels of instructional support required by the participants.

Stage 2: Instructional Follow-up

The diagnostic findings were subsequently used to inform instructional follow-up through three main strategies: differentiated instruction, scaffolding, and formative feedback. Differentiated instruction involved adjusting learning activities and levels of instructional support according to participants' identified needs. Scaffolding included structured assistance such as paragraph models, cohesive-device lists,

presentation templates, guided exercises, and gradual withdrawal of assistance as participants developed greater independence. Formative feedback was provided continuously to help participants identify their strengths, recognize specific difficulties, understand areas requiring improvement, and determine subsequent learning targets. Classroom observation was conducted throughout the twelve instructional meetings to document the implementation of these strategies and participants' responses to them.

Stage 3: Evaluation

Following the instructional period, participants completed a post-test using instruments parallel to those administered during the initial diagnostic stage. The post-test covered the same four language domains: listening, speaking, reading, and writing. Pre-test and post-test scores were examined descriptively to identify patterns of change during the twelve-week period. The post-test evidence was interpreted alongside classroom observations, interviews, and documentation. Thus, the evaluation stage was designed to examine observed development within the instructional process rather than to establish experimental causality.

2.4 Data Collection

Data from the prospective language teachers were collected using four complementary techniques: diagnostic testing, participant observation, in-depth interviews, and documentation. First, diagnostic tests covered the four core language domains. The listening test assessed participants' ability to identify main ideas and detailed information from academic speech recordings. The speaking test required a five-minute mini-presentation. The reading test assessed comprehension of popular scientific articles, while the writing test required a 500-word argumentative essay. Second, participant observation was conducted throughout twelve instructional meetings. The observations focused on participants' responses to differentiated instruction, scaffolding, formative feedback, classroom interaction, learning autonomy, and difficulties encountered during the learning process. Third, in-depth interviews were conducted with ten purposively selected participants and two course lecturers. The interviews explored participants' and lecturers' perceptions of diagnostic assessment, language-learning difficulties, instructional follow-up, and perceived changes during the twelve-week period.

Fourth, documentation included test answer sheets, written assignments, presentation recordings, analytic scoring rubrics, and relevant instructional records. The combination of diagnostic tests, observation, interviews, and documentation enabled methodological and source triangulation. The supplementary school-level evidence was collected separately. Grade VII-C students produced a descriptive text during one 40-minute lesson period in the school computer laboratory. Before writing, they received a brief explanation of descriptive texts and standard spelling conventions. The writing topic was "a favorite tourist destination," selected because it was based on familiar personal experience. The source-study researcher then examined each text for deviations from standard spelling conventions (*Ejaan Yang Disempurnakan*), standard word choice according to the Indonesian dictionary (KBBI), and principles of effective sentence construction. An interview with the Indonesian-language teacher was also used to explore the perceived causes of students' language errors. These procedures remained analytically separate from the university-level data collection.

2.5 Research Instruments

The researcher functioned as the primary instrument in accordance with the qualitative design. Supporting instruments included diagnostic test sheets, structured observation guidelines, semi-structured interview guidelines, documentation protocols, and analytic scoring rubrics for listening, speaking, reading, and writing. The diagnostic test instruments were developed based on language-proficiency indicators used in

the Indonesian Merdeka Curriculum for higher education. The instruments were reviewed by two senior lecturers with expertise in language assessment. Expert review focused on content relevance, clarity, alignment between items and the intended language skills, and appropriateness for the participant group. The observation guidelines were used to record classroom behavior, instructional responses, participant engagement, identified learning difficulties, and changes observed during the instructional process. Semi-structured interview guidelines were used to maintain consistency across interviews while allowing participants and lecturers to elaborate on relevant experiences. Instrument credibility was strengthened through expert review, methodological triangulation, and source triangulation. The use of multiple forms of evidence allowed diagnostic scores to be interpreted alongside classroom observations, interview data, written products, and instructional documentation. Because the study was primarily qualitative, methodological rigor was not treated solely in terms of psychometric reliability but also through systematic documentation, triangulation, and verification procedures.

2.6 Data Analysis

University-level data were analyzed using the interactive model of Miles, Huberman, and Saldana (2014), consisting of data condensation, data display, and conclusion drawing and verification. During data condensation, information from diagnostic assessments, field notes, interview transcripts, written assignments, presentation records, and other documentation was selected, organized, focused, and categorized according to the research questions. During data display, the findings were organized in the form of proficiency-profile tables, thematic matrices, descriptive summaries, and narrative evidence. Conclusion drawing and verification were performed iteratively throughout the analytical process.

Quantitative information was analyzed descriptively. Mean scores, percentages, frequencies, and pre-test–post-test score differences were used to illustrate patterns in participants' language performance. No inferential statistical analysis was conducted. Accordingly, observed score gains were interpreted descriptively and were not used to claim that diagnostic assessment independently caused the changes. Qualitative trustworthiness was addressed using the criteria of credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1985). Credibility was supported through source triangulation, method triangulation, and member checking. Transferability was strengthened through contextual description of participants and research settings. Dependability was maintained through documentation of research procedures and an audit trail, while confirmability was supported through researcher reflexivity and verification across different forms of evidence.

The supplementary school-level texts were analyzed separately using descriptive content analysis based on error analysis. Language errors were identified, classified, and counted under nine categories grouped into three broader domains: spelling, standard word choice, and effective sentence construction. Frequencies and percentages were calculated from the source-study counts and rounded to one decimal place. The Indonesian-language teacher interview was analyzed using the same interactive qualitative procedure. The university-level and school-level datasets were not statistically combined because they differed in educational level, task type, measurement context, and instructional treatment. The supplementary school evidence was used only for contextual and pedagogical interpretation.

2.7 Analytical Integration of the Two Data Sources

Although this study incorporated evidence from both university and secondary-school contexts, the two datasets served different methodological functions and were therefore maintained as analytically distinct sources. The university-level dataset constituted the principal research evidence, involving 30 prospective

language teachers who participated in the complete diagnostic–instructional–evaluation cycle, including diagnostic assessment, diagnostic mapping, differentiated instruction, scaffolding, formative feedback, post-assessment, interviews, observations, and documentation. In contrast, the Grade VII dataset functioned as supplementary contextual evidence derived from 30 descriptive texts and an Indonesian-language teacher interview. It was used to identify patterns of language difficulty at the secondary-school level through error analysis rather than to represent a second intervention group. Maintaining this distinction is important because assessment-literacy research emphasizes that prospective teachers develop professional assessment competence through the interpretation of assessment evidence and its connection with authentic classroom needs rather than through decontextualized knowledge alone (Anam & Putri, 2021; Farmasari et al., 2023). Empirical studies have similarly shown that prospective teachers' assessment literacy becomes more meaningful when assessment knowledge is connected with practical teaching situations and learners' actual performance (Giraldo, 2021; Prastikawati et al., 2024).

The integration of the two datasets therefore occurred at the interpretive and pedagogical level rather than through statistical comparison. Findings from the university-level diagnostic assessment were interpreted to explain how identified language difficulties informed differentiated instruction, scaffolding, formative feedback, and subsequent learning evaluation among prospective language teachers. The secondary-school evidence, meanwhile, was examined to contextualize the kinds of language difficulties that future teachers may encounter in actual school settings. This approach is consistent with recent empirical research showing that diagnostic assessment at the secondary-school level can identify specific language-learning difficulties and provide information for designing more responsive instructional support (Refoina & Rustam, 2024; Kuswari et al., 2024). It is also consistent with studies of pre-service teachers demonstrating that exposure to authentic assessment tasks, practicum situations, and classroom evidence contributes to the development of language assessment literacy by connecting assessment knowledge with pedagogical decision-making (Farmasari et al., 2023; Moraes et al., 2024). Therefore, the school-level dataset was not used to infer developmental continuity between Grade VII students and university students, nor was it used to determine differences between the two groups; instead, it provided contextual evidence for interpreting the professional assessment competencies required of prospective language teachers

2.8 Research Question Alignment

To maintain consistency between the research title, introduction, methodological procedures, and analytical strategy, each research question was aligned with a specific data source and analytical procedure.

Table 1. Research Questions, Research Stages, Data Sources, and Analytical Procedures

Research Question	Stage	Data Sources	Analytical Procedure
What language-proficiency difficulties are identified through diagnostic assessment among prospective language teachers?	Diagnosis	Listening, speaking, reading, and writing diagnostic assessments	Descriptive analysis and qualitative interpretation
How are diagnostic findings translated into differentiated instruction, scaffolding, and formative feedback in pre-service language teacher education?	Instructional Follow-up	Classroom observations, interviews, instructional documentation	Thematic analysis
How does supplementary evidence from secondary-school learners contribute to understanding the assessment literacy required of prospective language teachers?	Contextualization	Grade VII descriptive texts and Indonesian-language teacher interview	Descriptive content/error analysis and qualitative contextual interpretation

Source: Researcher's analytical framework, 2026.

3. RESULTS

The findings are organized according to the empirical sequence of the study and the three research questions. The section first presents the diagnostic profile of the prospective language teachers, followed by the instructional responses developed from the diagnostic evidence, the descriptive changes observed after the twelve-week instructional period, and the qualitative patterns documented through observation, interviews, and research records. Supplementary secondary-school evidence is subsequently presented as a separate analytical stream before the two datasets are brought together at the level of pedagogical interpretation. Throughout this section, photographs are used only as contextual documentary evidence of activities that can be directly observed in the research records; they are not treated as independent measures of learning outcomes.

3.1 Diagnostic Profile of Prospective Language Teachers

The initial diagnostic assessment demonstrated substantial variation across the four language domains. Among the 30 prospective language teachers, writing showed the lowest mean score, followed by speaking, whereas reading produced the highest initial mean. The diagnostic findings therefore indicated that the participants did not display a uniform pattern of language difficulty. Instead, problems were concentrated more strongly in productive language skills. Table 2 summarizes the initial diagnostic profile across the four language skills, including the mean score, performance category, and proportion of participants at or below the passing threshold.

Table 2. Diagnostic Profile of Prospective Language Teachers

Language Skill	Mean Score	Category	Participants at or Below the Passing Threshold
Listening	72.3	Sufficient	33.3% (10 participants)
Speaking	68.5	Sufficient	40.0% (12 participants)
Reading	74.8	Good	23.3% (7 participants)
Writing	65.2	Insufficient	50.0% (15 participants)
Overall Mean	70.2	Sufficient	36.7%

Source: Diagnostic test results, February 2026.

As presented in Table 2, writing constituted the most prominent initial difficulty, with a mean score of 65.2 and 15 participants at or below the passing threshold. Speaking was the second most problematic skill, while listening and reading showed relatively stronger performance. A more detailed examination of the diagnostic evidence identified four recurring areas: standard spelling and grammar difficulties among 43.3% of participants, limited academic vocabulary among 36.7%, paragraph-coherence difficulties among 50.0%, and limited speaking confidence among 40.0%. These four areas became the basis for the instructional responses described in the following subsection.

3.2. Diagnostic-Informed Instructional Responses

Following the diagnostic mapping, instructional activities were organized around the specific language difficulties identified during the initial assessment. Three recurring responses were documented during the twelve-week period: differentiated instruction, scaffolding, and formative feedback. These responses were not implemented as isolated techniques but as instructional actions derived from the diagnostic profile.

Table 3 presents the correspondence between the major diagnosed difficulties, the instructional responses, and the activities documented during the learning process.

Table 3. Instructional Responses to Identified Language Difficulties

Diagnosed Difficulty	Instructional Response	Documented Activity
Standard spelling and grammar	Differentiated instruction	Editing exercises, grammar and spelling activities
Academic vocabulary	Lexical scaffolding	Academic word lists, vocabulary journals, guided paraphrasing
Paragraph coherence	Modeling and structured practice	Model-paragraph analysis, paragraph reconstruction, peer review
Speaking confidence	Graded oral practice	Small-group discussions, presentation rehearsal, reflection

Source: Classroom observation and instructional documentation, 2026.

Table 3 shows that each major diagnostic difficulty was followed by a corresponding instructional response. Writing difficulties, for example, were differentiated into spelling, vocabulary, and paragraph organization rather than addressed as a single general weakness. Similarly, speaking confidence was addressed through progressively more demanding oral activities instead of immediate whole-class performance. The classroom photographs provide contextual evidence of the learning environment in which these instructional responses were implemented. Figure 4 presents only activities that are visually observable in the research documentation and does not assign unverified pedagogical labels to the photographs.

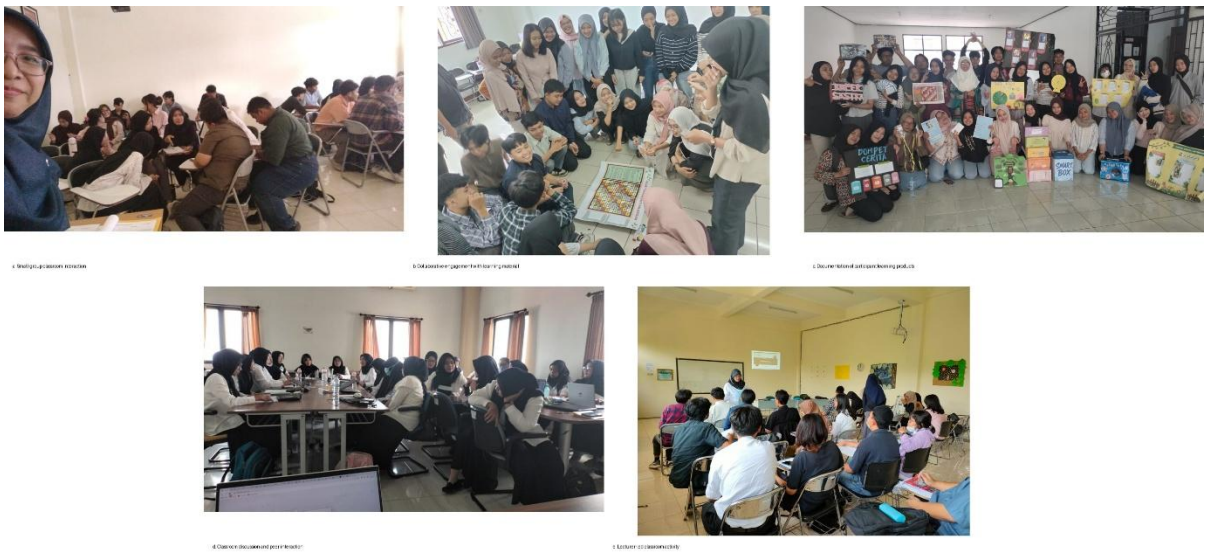


Figure 4. Documentary Evidence of Classroom and Collaborative Learning Activities
Source: Research documentation, 2026.

The photographs in Figure 4 support the observational evidence that instruction involved whole-class explanation, small-group interaction, collaborative activities, and production of learning materials. However, the photographs are treated as contextual documentation rather than as direct evidence of improvement in language ability, assessment literacy, or learning outcomes.

3.3. Descriptive Changes in Language Performance after Twelve Weeks

Following the twelve-week instructional period, participants completed a post-assessment using instruments parallel to the initial diagnostic assessment. Because the study did not include a control group,

the comparison is reported as descriptive change rather than evidence of a causal treatment effect. Table 4 presents the descriptive pre-test and post-test score patterns across the four language skills.

Table 4. Descriptive Comparison of Pre-test and Post-test Language Scores

Language Skill	Pre-test	Post-test	Observed Difference	Final Category
Listening	72.3	84.7	+12.4	Excellent
Speaking	68.5	88.1	+19.6	Excellent
Reading	74.8	91.5	+16.7	Excellent
Writing	65.2	89.9	+24.7	Excellent
Mean	70.2	88.6	+18.4	Excellent

Source: Pre-test and post-test data, February-April 2026.

The overall mean increased from 70.2 to 88.6, corresponding to an observed difference of 18.4 points. Writing showed the largest numerical increase, followed by speaking, reading, and listening. Notably, writing and speaking were also the two domains showing the greatest initial difficulty. This pattern establishes a descriptive relationship between the diagnostic profile and the later score pattern but does not establish that the instructional procedures independently caused the observed differences.

3.4. Observed Changes in Learning Behaviour

The numerical changes were accompanied by qualitative patterns documented through classroom observation, interviews, and instructional records. Three patterns appeared repeatedly: greater learning autonomy, increased participation in oral activities, and greater metalinguistic awareness. Participants increasingly used rubrics to revise their own work, took a more active role in group and classroom discussions, and identified language problems in peers' written work. Table 5 summarizes the qualitative patterns documented across the instructional period.

Table 5. Qualitative Patterns Documented during the Instructional Period

Observed Area	Initial Evidence	Subsequent Observed Behaviour
Learning autonomy	Dependence on lecturer guidance	Greater self-revision and use of rubrics
Speaking participation	Reluctance during oral performance	More active participation in discussions and presentations
Metalinguistic awareness	Difficulty identifying language problems	Greater ability to identify and discuss errors in peers' texts

Source: Classroom observations, interviews, and research documentation, 2026.

The convergence between observational, interview, and documentary evidence strengthens the interpretation that changes were not restricted to numerical scores. Nevertheless, these patterns are reported as observed developments during the study period rather than as experimentally verified causal outcomes.

3.5. Supplementary Secondary-School Evidence

The supplementary dataset consisted of 30 descriptive texts produced by Grade VII-C students at SMP Negeri 34 Surabaya. These data were analyzed separately from the prospective-teacher sample and were used exclusively to document language difficulties occurring in a secondary-school context. A total of 505 language errors were identified, corresponding to approximately 16.8 errors per text. Of these, 444 errors, or 87.9%, involved spelling conventions, while standard word choice accounted for 5.1% and ineffective sentence construction for 6.9%. Capitalization represented the most frequent individual error category,

accounting for 41.6% of all recorded errors. Table 6 presents the distribution of the identified error categories in the Grade VII descriptive texts.

Table 6. Distribution of Errors in Grade VII Descriptive Texts

Error Domain	Error Type	Frequency	Percentage
Spelling	Capital letters	210	41.6%
Spelling	Italics	16	3.2%
Spelling	Period	46	9.1%
Spelling	Comma	62	12.3%
Spelling	Affixation	17	3.4%
Spelling	Prepositions	21	4.2%
Spelling	Shortened words	72	14.3%
Spelling subtotal		444	87.9%
Standard word choice	Non-standard forms	26	5.1%
Effective sentences	Ineffective construction	35	6.9%
Total		505	100.0%

Source: Adapted from Dewi (2025); percentages calculated by the authors and rounded to one decimal place.

The predominance of spelling-related errors indicates that the school-level evidence was concentrated primarily in orthographic and convention-related language use. Because no instructional intervention or post-test was conducted within this supplementary dataset, these findings are used only for contextual interpretation.

3.6. Analytical Relationship between the Two Evidence Streams

The two datasets were not statistically compared because they involved different participants, educational levels, tasks, and measurement procedures. Nevertheless, they can be examined side by side to clarify their pedagogical relationship. The prospective-teacher dataset indicated that 43.3% of participants showed spelling and grammar difficulties, while the secondary-school dataset showed that 87.9% of recorded errors concerned spelling. These percentages represent different units of analysis and are therefore not directly comparable. Their relevance lies instead in the fact that both independently document difficulties involving written-language conventions. Table 7 presents the analytical relationship between the university-level and school-level evidence while explicitly preserving their methodological differences.

Table 7. Analytical Relationship between University-Level and School-Level Evidence

Area	Prospective Language Teachers	Secondary-School Learners	Interpretation
Language conventions	43.3% showed spelling/grammar difficulty	87.9% of errors were spelling-related	Contextually related, not statistically comparable
Written organization	50.0% showed paragraph-coherence difficulty	35 ineffective-sentence errors	Different units of analysis
Word use	36.7% showed limited academic vocabulary	26 non-standard word-form errors	Conceptually related language-use difficulty
Instructional response	Differentiated instruction and scaffolding implemented	No intervention evaluated	Different research stages
Outcome evidence	Pre/post and qualitative evidence available	No post-test	No outcome comparison

Source: University diagnostic data and supplementary secondary-school evidence, 2026.

Table 7 demonstrates why the school-level findings are relevant to pre-service language teacher education without treating the two populations as equivalent samples. The relationship is pedagogical and interpretive rather than statistical.

3.7. Triangulation of the Main Findings

The final analytical layer examined convergence across diagnostic tests, observations, interviews, and documentary evidence. The diagnostic assessments identified specific language difficulties; classroom observations documented participant behaviour during instructional activities; interviews provided participant and lecturer perspectives; documentation captured written work, presentations, learning products, and classroom activities; and the post-assessment provided descriptive evidence of numerical change. Table 8 summarizes how the main findings were supported across the different data sources.

Table 8. Cross-Source Triangulation of the Main Findings

Main Finding	Test Evidence	Observation Evidence	Interview Evidence	Documentation Evidence
Writing was initially weakest	Mean = 65.2	Writing difficulties observed	Limited prior writing practice reported	Written assignments
Speaking confidence was limited	Mean = 68.5	Initial reluctance observed	Fear of grammatical mistakes reported	Presentation activities
Instruction became differentiated	Different diagnostic needs identified	Different supports observed	Lecturer reported more focused teaching	Learning materials
Learning autonomy increased	—	Greater self-revision observed	Participant/lecturer perceptions	Revision records
Metalinguistic awareness increased	—	Peer-error identification observed	Classroom perceptions	Peer-review evidence
Scores increased descriptively	70.2 → 88.6	Greater later participation	Lecturer perceptions	Post-test records

Source: Diagnostic tests, classroom observations, interviews, and research documentation, 2026.

The triangulation indicates that the principal findings were supported by multiple types of evidence rather than a single measurement source. The photographs included in Figure 4 are therefore interpreted only as documentary support for the observed instructional context, while the claims regarding language performance and learning behaviour remain grounded in test, observation, interview, and documentary evidence.

4. DISCUSSION

4.1. Why Diagnostic Assessment Matters in Pre-Service Language Teacher Education

The first research objective asked what kinds of language-learning difficulties become visible when prospective language teachers are examined through diagnostic assessment rather than through conventional summative evaluation. The significance of the finding lies not in the numerical hierarchy of listening, speaking, reading, and writing scores reported in the Results section, but in the fact that the

participants did not constitute a pedagogically homogeneous group despite being enrolled in the same semester and study program. This heterogeneity is theoretically important because conventional course organization often assumes that students who enter the same curriculum possess sufficiently comparable linguistic foundations to receive essentially identical instruction. The present study problematizes that assumption. The diagnostic profile exposed different combinations of difficulties involving written conventions, academic vocabulary, paragraph organization, and oral confidence; consequently, the object of assessment shifted from determining who performed "well" or "poorly" toward identifying what each difficulty implied for subsequent teaching. This interpretation confirms the broader argument of Anam and Putri (2021) that assessment literacy among Indonesian pre-service teachers cannot be reduced to knowledge of testing procedures, yet it extends their work because their study primarily measured perceived and demonstrated assessment literacy, whereas the present research places prospective teachers themselves inside a diagnosis–instruction cycle. Öztürk (2022) similarly demonstrated that a classroom-assessment course can alter pre-service teachers' conceptions of assessment, while Farmasari et al. (2023) showed that satisfaction with assessment coursework does not necessarily guarantee preparedness for future assessment practice. Alcazaren (2024) further identified substantial development in pre-service teachers' understanding of assessment principles after a dedicated language-assessment course, and Moraes et al. (2024) showed that classroom tasks emphasize different components of language assessment literacy. Prastikawati et al. (2024), meanwhile, found that teaching practicum can raise assessment literacy from an initially insufficient level, while more recent Indonesian evidence suggests that pre-service teachers may express confidence while still displaying uneven understanding of core assessment principles (Putri et al., 2026). Taken together, these studies reveal a persistent gap between *knowing assessment* and *using assessment pedagogically*. The present study adds a different layer: the prospective teachers experienced what it means to be diagnosed, to have learning needs differentiated, and to receive instructional support based on evidence. The pedagogical meaning of the first finding is therefore experiential. Assessment literacy may be strengthened not only when future teachers study assessment frameworks, but also when they experience an assessment system that visibly changes subsequent instruction.

A second layer of interpretation concerns why productive language performance appeared more vulnerable than receptive performance within this local teacher-education context. A simplistic explanation would attribute the pattern to the general difficulty of speaking and writing, yet the field evidence suggests a more complex mechanism. Participants entered the program with unequal histories of literacy exposure, different levels of writing practice, and varying degrees of confidence in public oral performance. Such variation means that difficulty cannot be interpreted purely as a deficit in linguistic knowledge; it also reflects accumulated educational experience, familiarity with academic discourse, willingness to take linguistic risks, and the extent to which previous schooling required sustained production of formal language. This helps explain why the initial profile did not align neatly with a single cognitive account of proficiency. The finding therefore extends recent language-assessment literacy research by showing that diagnostic information in teacher education should incorporate the *conditions under which competence is expressed*, not merely the competence itself. Alcazaren (2024) emphasizes that assessment literacy requires understanding the purposes and principles of assessment, while Moraes et al. (2024) conceptualize it through knowledge, skills, and principles embedded in actual tasks. The current evidence suggests that these components need an additional pedagogical dimension: interpreting performance in relation to learners' prior learning trajectories and affective conditions. This point also complicates purely technocratic models of diagnostic assessment. A diagnostic instrument can identify a weak performance area, but without interview, observation, or contextual evidence, it may not distinguish between inadequate linguistic control, limited practice, low confidence, or unfamiliarity with academic conventions. The local significance of the study lies precisely in this triangulated interpretation. It suggests that teacher educators should resist the

temptation to convert diagnosis into fixed ability labels and instead treat it as a provisional account of learning needs. Such a position is compatible with the emerging literature showing that pre-service teachers themselves frequently require sustained practice before assessment knowledge becomes professionally usable (Farmasari et al., 2023; Prastikawati et al., 2024; Putri et al., 2026). In the longer term, this has implications for teacher-education curriculum design: diagnostic assessment should not be confined to one assessment course but embedded across language-development and teaching-methodology experiences so that prospective teachers learn to connect performance evidence, learner history, and instructional response.

4.2. From Diagnostic Evidence to Differentiated Instruction: Why the Follow-up Matters More Than the Test

The second research objective concerned how diagnostic findings were translated into differentiated instruction, scaffolding, and formative feedback. The critical issue is not simply that these strategies were used, but that each was attached to a specific type of diagnosed difficulty. This distinguishes the study from a procedural model in which diagnostic testing is completed as an administrative requirement but produces little change in classroom practice. Wardhani's (2023) study at an Indonesian vocational school demonstrated that diagnostic assessment can inform differentiation of content, process, and product, while Nazilah (2024) similarly argues that diagnostic information can guide appropriate differentiated responses to learner readiness and needs. Yet evidence from Indonesian teachers also reveals a structural problem: understanding diagnostic assessment does not always result in its systematic use for differentiation. Rohmana and colleagues' work with Indonesian EFL teachers found that although many teachers understood diagnostic assessment, a substantial proportion did not consistently use the results to differentiate instruction. This implementation gap is echoed by Musyarrofah et al. (2024), who found that pre-service teachers displayed lower capability than in-service teachers in applying differentiated instruction and required more practice in university coursework and practicum. Rahmani et al. (2024) adds another complication: pre-service teachers may adopt scaffolding, flexible grouping, and multiple assessment forms during teaching practice, yet encounter practical barriers involving time, resource constraints, learner resistance, engagement, and the challenge of evaluating different products. Against this literature, the present study occupies an intermediate position. It confirms that diagnosis can provide a workable basis for differentiated teaching, but it also extends previous work by documenting the relationship at the level of *specific pedagogical correspondence*: language-form problems were followed by editing and language-focused work, vocabulary limitations by lexical support, coherence difficulties by modeling and reconstruction, and oral confidence by graded speaking opportunities. Theoretically, this suggests that differentiation should be understood as an evidence-response architecture rather than as a collection of fashionable strategies. A differentiated classroom is not created merely by grouping learners or providing different worksheets; it emerges when the reason for variation in instruction can be traced to credible information about learner need. That distinction is crucial for pre-service education because it changes differentiation from a technique that student teachers imitate into a reasoning process they must learn to justify.

The study also reveals an important theoretical tension between differentiation as responsiveness and differentiation as potential fragmentation. If every diagnostic difference were converted into a separate instructional track, a classroom could become pedagogically unmanageable and socially segmented. The field evidence indicates a different solution: participants continued to engage with a common learning theme while receiving different forms or intensities of support. This pattern is analytically important because it preserves shared curricular participation while varying access to that curriculum. Recent Indonesian research on differentiated instruction helps clarify why this matters. Musyarrofah et al. (2024)

describe capability gaps and implementation barriers among pre-service and in-service teachers, whereas Rahmani et al. (2024) documents the practical difficulty of simultaneously addressing readiness, learning style, engagement, and classroom constraints. A 2026 study of junior-high English instruction likewise highlights teacher and student readiness as a condition shaping differentiated learning rather than assuming differentiation is automatically effective whenever introduced (Andriyani et al., 2026). The present study therefore complicates overly optimistic accounts of differentiation: diagnosis can make instructional adaptation more rational, but it does not eliminate logistical constraints, nor does it justify unlimited individualization. Its practical value lies in prioritization—determining which differences are instructionally consequential and what level of support is feasible. The same reasoning applies to scaffolding and formative feedback. The structured supports used in this study were not ends in themselves; their pedagogical legitimacy depended on their gradual withdrawal or adaptation as participants developed greater control. Formative feedback likewise mattered because it identified a next action, not because feedback was simply present. This interpretation expands language assessment literacy beyond the ability to construct and score assessments. An assessment-literate prospective teacher must be able to infer what pedagogical action is warranted by assessment evidence, recognize when support is no longer necessary, and avoid transforming diagnosis into a permanent label. Such reasoning is only partially captured by conventional assessment-literacy inventories. Moraes et al. (2024) emphasizes tasks that develop knowledge, skills, and principles, while Alcazaren (2024) demonstrates that coursework can reshape assessment conceptions; the present findings suggest that *instructional consequentiality*—the ability to make defensible teaching decisions from assessment evidence—should be treated as a central component of professional assessment literacy.

4.3. Interpreting Development after the Diagnostic–Instructional Cycle without Making Causal Claims

The third research objective concerns what can legitimately be inferred from the developmental pattern observed after the twelve-week instructional cycle. The key analytical challenge is to avoid converting temporal sequence into causal proof. The study recorded stronger performance at the later assessment point, particularly in domains that had been more problematic initially, but the design did not include random assignment, a control condition, or inferential procedures capable of isolating the effect of diagnostic assessment from other influences. A defensible interpretation therefore treats the pattern as *coherent with* diagnostic-responsive instruction rather than *caused by* it. This distinction separates the present study from Wardhani (2023), whose mixed-method design incorporated a paired statistical comparison and reported significant score differences after differentiated instruction. It also differs from Prastikawati et al. (2023), who used pre- and post-measures of language assessment literacy with a much larger group of pre-service EFL teachers and reported significant development after an assessment course, and from Prastikawati et al. (2024), whose mixed-method practicum study similarly examined development in language assessment literacy. Öztürk (2022), by contrast, focused on reflective reports and focus-group evidence of changed perspectives after a classroom-assessment course. Alcazaren (2024) used qualitative reflections to trace changes in conceptual understanding, while Putri et al. (2026) found continuing unevenness in assessment readiness even among participants who had completed practicum experience. These divergent designs matter because they demonstrate that "development" in teacher education is methodologically constructed differently across studies: some measure statistical change, some examine reflective transformation, others investigate demonstrated competence, and the current study combines descriptive performance trajectories with observational and interview evidence. Its contribution therefore does not lie in claiming superior effectiveness but in showing convergence across several kinds of evidence within a single instructional context.

The most analytically interesting pattern is not that later performance was higher, but that the domains initially requiring stronger instructional attention also displayed substantial later movement while qualitative observations documented shifts in autonomy, participation, and metalinguistic noticing. That combination invites a process explanation. Diagnostic instruction may have operated by increasing the *precision of practice*: participants did not merely receive more instruction but encountered practice more closely aligned with identified needs. Scaffolding reduced the immediate cognitive burden of difficult tasks, while formative feedback made the criteria for improvement more explicit. Repeated use of rubrics and peer examination may have externalized linguistic standards that had previously remained implicit, thereby enabling participants to monitor their own production more effectively. Yet alternative explanations remain plausible. Course progression itself, repeated exposure to similar task types, growing familiarity with lecturer expectations, peer interaction, or independent study could also contribute. This uncertainty is not a weakness to be concealed; it defines the epistemic boundary of the study and sharpens the next research agenda. A future quasi-experimental or longitudinal design could test whether diagnostic-responsive instruction produces advantages beyond ordinary instructional progression and whether gains remain stable during teaching practicum. The present evidence nevertheless extends the literature because it suggests that development in prospective teachers should not be judged solely through score improvement. Prastikawati et al. (2024) positions teaching practicum as a bridge into teachers' roles as assessors; Moraes et al. (2024) emphasizes that assessment literacy comprises knowledge, skills, and principles; Alcazaren (2024) shows that assessment courses can reshape how prospective teachers understand assessment. The current study adds *learner-side experience* to this picture. Participants encounter assessment not only as future assessors but as learners whose instruction changes in response to diagnostic evidence. Such experience may produce a more embodied understanding of why assessment consequences matter. The theoretical implication is that pre-service assessment literacy may develop through three interlocking positions: learning about assessment, practicing assessment, and experiencing assessment-informed pedagogy from the learner's side.

4.4. From Language Development to Emerging Assessment Literacy

A deeper interpretation of the qualitative findings suggests that the study is not only about language development but about the beginnings of professional assessment literacy. Increased use of rubrics for self-revision, greater willingness to participate orally, and more frequent identification of errors in peers' writing are significant because each involves a change in how participants relate to evidence about language performance. Self-revision indicates that assessment criteria began to function as regulatory tools rather than external judgments; peer-error identification requires noticing and articulating language problems; increased oral participation suggests that assessment contexts became less exclusively associated with threat and more with iterative learning. This interpretation confirms aspects of Öztürk's (2022) finding that assessment coursework can transform pre-service teachers' perceptions of assessment from negative or restrictive meanings toward greater professional confidence. It also resonates with Alcazaren's (2024) observation that pre-service teachers' reflections changed substantially across a language-assessment course and with Moraes et al. (2024), who argue that classroom tasks can selectively activate different components of assessment literacy. Yet the present case extends those studies because the participants' developing assessment awareness emerged through engagement with their *own* linguistic performance rather than exclusively through the design of assessments for hypothetical or practicum learners. Farmasari et al. (2023) demonstrates that preparedness is not guaranteed by having completed a language-assessment course; Putri et al. (2026) similarly reports uneven understanding among Indonesian pre-service English teachers who had completed practicum. These studies expose a recurring weakness in teacher preparation: declarative confidence can exceed practical interpretive capacity. The current evidence suggests one route

for addressing that mismatch. When prospective teachers repeatedly use criteria to diagnose, revise, and explain language performance, assessment literacy becomes situated within language learning itself.

The anomaly, however, is that the participants were not explicitly being trained in this study as assessors of school pupils; the main intervention focused on their own language development. Any claim that their future assessment practice has already improved would therefore exceed the evidence. This boundary is especially important given the literature on transfer from university coursework to practicum. Prastikawati et al. (2024) found that practicum experience was associated with improved language assessment literacy but still described assessment competence as developmental rather than complete. Musyarrofah et al. (2024) similarly reports that pre-service teachers remain less capable than in-service teachers in differentiated-instruction implementation, and Rahmani et al. (2024) shows that pre-service teachers encounter substantial barriers when attempting differentiation in actual school environments. These studies caution against assuming that successful participation in a university-based diagnostic cycle automatically produces successful classroom assessment later. What the present study more defensibly demonstrates is the creation of a *professional learning resource*: participants have experienced a model in which diagnosis precedes instructional adaptation, criteria support revision, and evidence is used to reconsider teaching decisions. Whether that resource becomes durable professional practice remains an empirical question. This distinction generates a useful theoretical contribution. Assessment literacy should not be conceptualized only as an endpoint competence measured through inventories; it can also be examined as a trajectory in which experiences progressively furnish prospective teachers with interpretive repertoires. In practical terms, teacher-education programs could deliberately connect language-development courses, assessment courses, and practicum rather than treating them as separate curricular silos. The same diagnostic cycle could first be experienced by prospective teachers as learners, then analyzed theoretically in assessment coursework, and finally adapted for school learners during practicum. Such sequencing would directly address the knowledge–practice gap documented by Anam and Putri (2021), Farmasari et al. (2023), Prastikawati et al. (2024), and Putri et al. (2026), while avoiding the assumption that one isolated assessment course can produce fully assessment-literate teachers.

4.5. What the Secondary-School Error Evidence Adds—and What It Does Not Permit Us to Claim

The supplementary Grade VII evidence should be interpreted as a contextual lens rather than as a second experimental dataset. Its analytical value lies in demonstrating that prospective teachers are preparing to enter classrooms where language problems are concrete, recurrent, and fine-grained: learners may struggle with orthographic conventions, standard word forms, punctuation, and effective sentence construction even when they can communicate the intended meaning. This contextual evidence strengthens the professional relevance of diagnostic assessment because it illustrates what future teachers may actually have to notice in students' texts. The comparison must nevertheless remain conceptual rather than numerical. The university data identify the proportion of participants exhibiting particular areas of difficulty, whereas the school dataset counts error instances in written products. The units of analysis, tasks, educational stages, and assessment conditions differ; any claim that the school pupils and prospective teachers exhibit "the same level" of difficulty would therefore be invalid. This methodological boundary is one of the most important revisions prompted by the reviewers. The more defensible conclusion is that both datasets independently expose vulnerabilities in written-language conventions and that the school evidence gives those vulnerabilities professional relevance. Recent diagnostic-assessment research in Indonesian schools supports the broader importance of using assessment to identify heterogeneous learner needs. Wardhani (2023) found that diagnostic assessment could be translated into differentiated content, process, and product in secondary English learning, while Nazilah (2024) argues that diagnostic information can help teachers

tailor differentiated responses. Rohmana's work on Indonesian EFL teachers adds a critical counterpoint: understanding diagnostic assessment does not ensure that results are actually used to differentiate instruction. The school-level evidence in the current manuscript therefore does more than illustrate learner error; it dramatizes the professional problem awaiting prospective teachers—diagnosis is valuable only when it informs action.

The teacher interview in the supplementary dataset also highlights a structural issue that individual assessment literacy alone cannot solve. The teacher identified limited mastery of standard writing conventions and described additional support that was not uniformly accessed by all learners. This suggests that persistent language difficulties may be reproduced not simply because teachers fail to identify them but because instructional time, access to remediation, participation patterns, and school-level arrangements constrain follow-up. Such constraints are consistent with Rahmani et al. (2024), who found that pre-service teachers implementing differentiated instruction encountered time-management, resource, engagement, and assessment difficulties, and with Musyarrofah et al. (2024), who documented barriers to differentiated practice among Indonesian teachers. The implication is theoretically significant: a diagnostic model that focuses only on the individual teacher risks underestimating the institutional conditions needed for responsive instruction. Prospective teachers need assessment literacy, but assessment literacy operates inside schedules, curricula, class sizes, resource availability, and school expectations. The present study therefore extends a purely competence-based account toward a more ecological interpretation. The school dataset demonstrates what diagnosis can reveal; the university dataset demonstrates one form of responsive follow-up; the gap between them raises the question of whether future teachers will have sufficient structural space to apply such responsiveness consistently. This is also where differentiated-instruction research becomes relevant. Andriyani et al. (2026) emphasize readiness among both teachers and students, while Rahmani et al. (2024) and Musyarrofah et al. (2024) show that differentiated practice is constrained by real classroom conditions. Long-term policy implications should therefore include not only assessment training but also curricular flexibility, time for diagnostic follow-up, access to formative tools, and practicum environments in which prospective teachers can experiment with evidence-based differentiation under supervision. The professional goal is not to produce teachers who can merely detect errors, but teachers who can make contextually realistic decisions about which errors require intervention, what kind of support is feasible, and how to distribute that support equitably.

4.6. An Applied Linguistics Perspective on Diagnostic Assessment and Language Development

The findings can be interpreted productively from an applied linguistics perspective because the study sits at the point where learner language, assessment evidence, instructional response, and teacher preparation intersect. Applied linguistics is concerned not only with describing language systems but also with understanding how language is learned, assessed, taught, and used within real educational contexts. In the present study, diagnostic assessment functioned as a mechanism for making learner-language difficulties visible and pedagogically interpretable. This is important because language performance is rarely homogeneous across domains; learners may display relatively strong receptive performance while continuing to experience difficulty in productive language use, academic vocabulary, discourse organization, or confidence in oral communication. Such patterns reinforce the need to view language competence as multidimensional and developmentally uneven rather than as a single stable ability.

This interpretation extends previous work on language assessment literacy and teacher preparation. Anam and Putri (2021) showed that Indonesian pre-service and in-service teachers demonstrate varying degrees of assessment literacy, while Atjonen et al. (2022) emphasized that assessment knowledge must be connected with teacher learning and conceptions of assessment. Farmasari et al. (2023) further reported that

satisfaction with assessment preparation does not necessarily indicate readiness to implement assessment effectively, and Alcazaren (2024) documented how pre-service teachers' understanding of assessment develops through explicit coursework. Moraes et al. (2024) similarly demonstrated that assessment tasks can activate different components of assessment literacy, whereas Prastikawati et al. (2024) found that teaching practicum contributes to the development of practical assessment competence. The current study adds a complementary dimension by showing that prospective teachers can also learn about assessment through experiencing diagnostic assessment as learners. This learner-side experience may deepen understanding of how assessment evidence influences pedagogy because participants encounter the consequences of diagnosis directly through differentiated support, scaffolding, and feedback.

From the perspective of language development, the study also raises questions about the relationship between error, awareness, and instructional mediation. Diagnostic identification of language difficulties does not automatically produce development; what matters is how those difficulties are interpreted and transformed into learning opportunities. The observed use of targeted support suggests that diagnostic assessment may be most valuable when it helps teachers distinguish between different forms of difficulty and match them with appropriate forms of assistance. This is consistent with sociocultural perspectives that emphasize mediated learning and scaffolded development, particularly where support is adjusted in response to learner needs rather than delivered uniformly (Lantolf & Poehner, 2014). The contribution of the present study is therefore not the claim that diagnostic assessment is a new concept, but that it can function as a bridge between applied linguistic analysis of learner language and responsive instructional decision-making within teacher education.

The school-level evidence strengthens this interpretation because it illustrates the kinds of language difficulties that prospective teachers may later encounter in professional practice. The presence of recurrent problems in spelling, standard word choice, and sentence construction among secondary-school learners suggests that future language teachers need more than theoretical familiarity with grammatical rules; they need the ability to interpret learner production diagnostically and decide what type of instructional response is appropriate. From an applied linguistics perspective, this represents a shift from language knowledge as declarative content toward language knowledge as professional judgment. Prospective teachers must be able to identify what a learner's error reveals, distinguish recurring patterns from incidental mistakes, and determine how feedback or scaffolding should be adapted to the learner's needs. This position places diagnostic assessment within a broader continuum of learner-language analysis, classroom assessment, and pedagogical reasoning.

4.7. Theoretical Contribution: Diagnostic Assessment as Experienced Assessment Literacy

The theoretical contribution of this study lies in extending the concept of language assessment literacy from knowledge about assessment toward experience with the pedagogical consequences of assessment. Much of the recent literature conceptualizes assessment literacy through teachers' knowledge, skills, principles, beliefs, preparedness, or ability to design and interpret assessment tasks. Anam and Putri (2021) documented differences in assessment literacy among Indonesian teachers, Atjonen et al. (2022) emphasized the relationship between assessment knowledge and teacher learning, and Farmasari et al. (2023) highlighted the gap between perceived preparedness and actual readiness. Alcazaren (2024) and Moraes et al. (2024) further demonstrated that assessment understanding develops through explicit coursework and structured assessment tasks, while Prastikawati et al. (2024) showed that practicum experience can strengthen practical assessment competence. These studies collectively suggest that assessment literacy is developmental, but they largely examine prospective teachers as learners of assessment concepts or as novice assessors.

The present study extends that perspective by positioning prospective teachers as recipients of diagnostic assessment whose instruction changes in response to assessment evidence. This is theoretically meaningful because it allows assessment literacy to be understood not only as the ability to design, administer, and interpret assessments, but also as an understanding of how assessment decisions affect learners. When prospective teachers experience diagnostic identification, differentiated support, formative feedback, and reflective use of assessment criteria, they encounter assessment as a pedagogical process rather than as a technical procedure. Such experience may contribute to a more situated understanding of why assessment matters and how assessment information should be used responsibly.

This contribution can be framed as a developmental sequence in which prospective teachers first encounter diagnostic assessment as learners, then interpret how assessment evidence informs pedagogical decisions, and later may apply similar reasoning in their own teaching practice. The study does not establish that this transfer necessarily occurs, because participants were not followed into practicum or professional teaching. However, it provides a theoretically plausible basis for linking learner experience with future assessment practice. This extends existing assessment-literacy models by emphasizing experiential understanding as a possible precursor to professional judgment.

A further contribution concerns the integration of assessment literacy with learner-language analysis. The school-level evidence demonstrates that future teachers will encounter language difficulties that require more than mechanical error correction. They must decide whether a language form represents a recurring pattern, an emerging developmental feature, a register issue, or a context-specific deviation. This requires both linguistic knowledge and interpretive judgment. From an applied linguistics perspective, assessment literacy therefore intersects with knowledge of learner language, discourse, language development, and pedagogical response. The present study highlights this intersection and suggests that assessment literacy in language teacher education should be treated as a multidimensional professional competence rather than a discrete assessment skill.

4.8. Practical Implications for Language Teacher Education

The practical implications of the study concern how language teacher education programs structure the relationship among language development, assessment coursework, and practicum experience. The findings suggest that assessment literacy is unlikely to develop fully when assessment is treated as an isolated course detached from prospective teachers' own language-learning experiences. A more integrated model would allow students to experience diagnostic assessment in language-development courses, examine the theoretical basis of those practices in assessment courses, and later apply diagnostic reasoning to school learners during practicum. Such integration would respond to the persistent knowledge-practice gap identified in recent research. Farmasari et al. (2023) showed that pre-service teachers may feel satisfied with assessment preparation while remaining insufficiently prepared for classroom assessment. Prastikawati et al. (2024) found that practicum experience contributes to assessment-literacy development, while Musyarrofah et al. (2024) demonstrated that pre-service teachers continue to face difficulties when implementing differentiated instruction in practice. These studies indicate that professional competence cannot be developed through theory alone. Prospective teachers need repeated opportunities to interpret learner evidence, make instructional decisions, receive feedback on those decisions, and reflect on the consequences of their choices.

For language teacher educators, the present findings suggest that diagnostic assessment should be integrated into coursework as an instructional reasoning process. Students should not only learn how to construct

diagnostic instruments but also how to interpret learner responses, identify patterns of difficulty, select appropriate scaffolding, and evaluate whether instructional follow-up addresses the diagnosed need. This approach would shift assessment education from procedural competence toward pedagogical judgment.

The implications also extend to practicum design. Prospective teachers could be required to conduct small-scale diagnostic activities, document patterns of learner difficulty, design targeted instructional responses, and reflect on how their decisions affected classroom participation and learner performance. Such practicum tasks would create continuity between university-based assessment learning and real classroom practice. They would also provide stronger evidence of whether diagnostic-assessment experience is transferred into teaching. A longer-term implication concerns curriculum coordination within teacher-education programs. Language proficiency courses, assessment courses, methodology courses, and practicum are often taught as separate components. The current findings suggest that stronger integration among these areas could help prospective teachers develop a more coherent understanding of assessment as part of language pedagogy. In this model, language development, assessment literacy, and instructional responsiveness would be treated as interconnected dimensions of professional preparation rather than independent curricular requirements.

5. CONCLUSION

5.1 Conclusion

The most significant finding of this study is that diagnostic assessment serves two interconnected functions in pre-service language teacher education: identifying learners' specific language difficulties and providing prospective teachers with experiential opportunities to understand how assessment evidence informs instructional decisions. Beyond identifying language weaknesses, diagnostic assessment offers a foundation for differentiated instruction, targeted scaffolding, and formative feedback. More importantly, the findings highlight that the educational value of diagnostic assessment lies not merely in measuring language performance but in translating diagnostic information into responsive pedagogical practices. This dual function positions prospective teachers simultaneously as language learners receiving instructional support and future educators developing the capacity to interpret learner-language evidence. Although descriptive changes in language performance were observed during the instructional period, these improvements cannot be attributed exclusively to diagnostic assessment because the study did not employ random assignment or a control group.

The principal theoretical contribution of this study lies in connecting diagnostic assessment, responsive language instruction, and Language Assessment Literacy within pre-service teacher education. While existing approaches recognize diagnostic assessment as a means of identifying learning needs, this study offers a complementary perspective by emphasizing its experiential role in developing prospective teachers' understanding of assessment-informed instruction. From an Applied Linguistics perspective, the findings suggest that assessment literacy should extend beyond knowledge of assessment instruments and procedures to include the interpretation of learner-language evidence and its translation into pedagogical action. The supplementary secondary-school evidence further illustrates the professional relevance of this preparation by identifying language difficulties that prospective teachers may encounter in future classroom settings. However, these supplementary findings provide contextual support rather than direct evidence of professional competence or instructional effectiveness.

Several limitations should be acknowledged. First, the descriptive qualitative design, supported by numerical indicators, does not establish causal relationships between diagnostic assessment and observed language development. Second, the absence of random assignment and a control group limits conclusions regarding instructional effectiveness. Third, the university-level and supplementary school-level datasets represent distinct educational contexts and cannot be interpreted as a single integrated sample. Finally, the study did not longitudinally examine whether prospective teachers transferred their diagnostic-assessment experiences into independent assessment practices during teaching practicum or professional employment. Consequently, the findings primarily reflect assessment-informed instructional experiences and observed developmental patterns within the investigated contexts.

Future research should employ longitudinal and mixed-method designs to investigate how diagnostic-assessment experiences influence prospective teachers' Language Assessment Literacy, classroom decision-making, feedback practices, and differentiated instruction during practicum and early-career teaching. Comparative or quasi-experimental studies could provide stronger evidence concerning the contribution of diagnostic assessment to language development and instructional responsiveness. Further investigations should also examine how prospective teachers interpret diagnostic evidence independently, adapt instructional strategies to diverse learner needs, and sustain assessment-informed practices across different educational settings. Such research would strengthen the empirical foundation for integrating diagnostic assessment into language teacher education and clarify its longer-term contribution to professional teaching competence.

5.2 Limitations and Future Research Directions

Language teacher-education programs are encouraged to integrate diagnostic assessment at the beginning of relevant courses so that instructional planning can be based on identified learner needs rather than uniform assumptions. Diagnostic findings should be followed by differentiated instruction, scaffolding, and formative feedback so that assessment functions as part of the learning process rather than solely as a mechanism for final grading.

At the program level, further development of standardized and validated diagnostic instruments is recommended to improve consistency across classes and strengthen the interpretation of learners' language profiles. Particular attention should be given to writing, academic vocabulary, paragraph organization, and oral confidence, as these areas were prominent in the present study.

For secondary-school settings, Indonesian-language teachers may use systematic error analysis of students' own texts as a practical diagnostic tool, especially for identifying recurring problems in capitalization, punctuation, shortened informal forms, word choice, and sentence construction. However, diagnostic identification should be followed by structured instructional support rather than limited to error classification alone.

Future research should examine the diagnose–intervene–evaluate cycle using stronger longitudinal or comparative designs, particularly in secondary-school contexts. Studies involving control or comparison groups, repeated measurement, practicum settings, or technology-supported adaptive diagnostic assessment would provide stronger evidence regarding the relationship between diagnostic assessment, instructional responsiveness, and learner development. Further research may also investigate how prospective teachers transfer assessment literacy from university-based preparation into actual school teaching practice.

6. ACKNOWLEDGMENT

The authors gratefully acknowledge Universitas Negeri Surabaya, the participating students and lecturers, and SMP Negeri 34 Surabaya for their support and contributions to the data collection and implementation of this study.

Ethical & Author Statements

CRediT Statement: Mukhzamilah: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – This study Draft, Project Administration. Shigenori Wakabayashi: Theoretical Interpretation, Applied Linguistics Perspective, Writing – Review & Editing, Critical Revision. All authors reviewed and approved the final version of the manuscript.

Conflict of Interest: The authors declare no conflict of interest in this research.

Funding: This research received no external funding from any agency, commercial, or non-profit sectors.

Data Availability Statement (Data Policy): The data analyzed in this study were obtained from diagnostic assessments, classroom observations, interviews, instructional documentation involving prospective language teachers, and supplementary descriptive texts from Grade VII students. Due to participant confidentiality, ethical considerations, and the educational nature of the data, the raw dataset is not publicly available. Supporting data may be made available by the corresponding author upon reasonable request and subject to applicable confidentiality and institutional requirements.

AI Policy: Artificial Intelligence (AI) tools, including large language models, were used solely for linguistic consistency, language refinement, grammar improvement, and editorial support. All research design decisions, participant selection, data collection, data analysis, interpretation of findings, theoretical reasoning, and conclusions remain the responsibility of the authors. The authors reviewed and approved the final manuscript and take full responsibility for its accuracy, this studyity, integrity, and scholarly content.

7. REFERENCES

Ababneh, O. M. A. (2021). How do green HRM practices affect employees' green behaviors? The role of employee engagement and personality attributes. *Journal of Environmental Planning and Management, 64*(7), 1204–1226. https://doi.org/10.1080/09640568.2020.1814708

Al Khaldy, M., Gheraibia, Y., & Shaheen, A. (2026). Multilingual large language models for low-resource languages: Challenges and opportunities. In *2026 2nd International Conference on Computational Intelligence Approaches and Applications (ICCIAA 2026) – Proceedings*. https://doi.org/10.1109/ICCIAA68481.2026.11543773

Alcazaren, H. K. G. (2024). Developing language assessment literacy of Filipino pre-service teachers: Reflections from a language assessment course. *TEFLIN Journal, 35*(2), 216–231. https://doi.org/10.15639/teflinjournal.v35i2/216-231

Alfin, J. (2022). *Analisis kesalahan berbahasa Indonesia* [Analysis of Indonesian language errors]. Nasmedia.

Anam, S., & Putri, N. V. W. (2021). How literate am I about assessment: Evidence from Indonesian EFL pre-service and in-service teachers. *English Review: Journal of English Education, 9*(2), 151–162. https://doi.org/10.25134/erjee.v9i2.4374

Anderson, L. W., & Krathwohl, D. R. (2017). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (A. Prihantoro, Trans.). Pustaka Pelajar.

Andriyani, S., Mahendra, D., Saefudin, A., Sa'idah, N., Yulistianti, H. D., & Haq, D. A. F. F. (2026). Readiness for differentiated instruction in English language learning: Perceptions of junior high school teachers and students. *Sapienza: International Journal of Interdisciplinary Studies, 7*(1), e26002. https://doi.org/10.51798/sijis.v7i1.1267

Atjonen, P., Pöntinen, S., Kontkanen, S., & Ruotsalainen, P. (2022). Enhancing preservice teachers' assessment literacy: Focus on knowledge base, conceptions of assessment, and teacher learning. *Frontiers in Education, 7*, 891391. https://doi.org/10.3389/feduc.2022.891391

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Bruyninckx, S., Stevens, A.-L., Van Oss, V., Taelman, H., & Spilt, J. L. (2026). How do early childhood education teachers support language? Frequency and variation in practices across activities. *Early Education and Development*. https://doi.org/10.1080/10409289.2026.2663287

Cen, J.-P., Ge, S.-D., Zhou, Y.-S., Li, Y.-Z., Guo, Z.-F., Peng, R., Liu, Y.-G., Zhou, S., Xu, S.-Y., Zhao, S.-C., & Liu, D. (2026). General pathologists achieve near-specialist diagnostic performance using deep learning–based virtual staining for donor kidney assessment: A retrospective-prospective diagnostic concordance study. *Laboratory Investigation, 106*(4). https://doi.org/10.1016/j.labinv.2026.106077

Chen, H., & Yang, H. (2026). Explainable machine learning for predicting heritage-language adhesion from language-attitude survey data: A case study in Guangxi, China. In *Proceedings of the 2026 5th International Conference on Big Data, Information and Computer Network (BDICN 2026)* (pp. 174–179). https://doi.org/10.1145/3801228.3801255

Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833200

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Cui, Q., Hussin, S., & Wong, W. L. (2026). Reducing speaking anxiety through AI interaction: ChatGPT-assisted English-speaking practice and language proficiency development in Chinese universities. *World Journal of English Language, 16*(6), 524–537. https://doi.org/10.5430/wjel.v16n5p524

Dewi, F. A. (2025). *Analisis kesalahan penggunaan bahasa baku dalam teks deskripsi siswa kelas VII SMP Negeri 34 Surabaya* [Analysis of errors in standard-language use in descriptive texts by seventh-grade students of SMP Negeri 34 Surabaya] [Unpublished undergraduate thesis]. Universitas Negeri Surabaya.

du Plessis, T. (2025). Post-2023 language planning in South Africa: Implications for the historically marginalised indigenous languages, South African Sign Language and Afrikaans. *Tydskrif vir Geesteswetenskappe, 65*(1), 188–212. https://doi.org/10.17159/2224-7912/2025/v65n1a9

Duthois, T., Vanderlinde, R., Van Avermaet, P., & Montero-Perez, M. (2026). Can we professionally develop teachers' visual attention? Video-based professional development of early childhood teachers' in-action noticing during language-stimulating interactions: A mobile eye-tracking study. *Learning and Instruction, 105*, 102396. https://doi.org/10.1016/j.learninstruc.2026.102396

Farmasari, S., Wardana, L. A., Baharudin, B., Herayana, D., & Suryaningsih, H. (2023). Pre-service EFL teachers' language assessment literacy satisfaction and assessment preparedness. *International Journal of Language Education, 7*(4), 646–660. https://doi.org/10.26858/ijole.v7i4.47445

Francisca Carocca, P. (2026). Beyond translation: Epistemic justice in the adaptation of the "Coding as Another Language" curriculum. *Revista Iberoamericana de Tecnologias del Aprendizaje, 21*, 339–345. https://doi.org/10.1109/RITA.2026.3683518

Hagen, Å. M., & Rydland, V. (2026). Preserving pedagogy, promoting language: Implementation challenges and possibilities for structured language interventions in Nordic early childhood education. *Early Childhood Education Journal, 54*(4), 3057–3069. https://doi.org/10.1007/s10643-025-02076-6

Hendratno, Subrata, H., Istiq'faroh, N., Saraswati, L., Fakhrudin, A., Kharisma, N. V. E., Haq, M. S., & Adebisi, A. A. (2026). The effect of AI-assisted Bing Image Creator on elementary students' self-confidence and speaking: Fostering SDGs through education. In S. N., B. K. P., M. S., L. L., & G. M. A. (Eds.), *E3S Web of Conferences* (Vol. 696). EDP Sciences. https://doi.org/10.1051/e3sconf/202669602011

Hidayat, L. E., Amalia, S. N., Rohmana, W. I. M., Virdyna, N. K., Ardi, P., & Suryati, N. (2025). EFL teachers' understanding and practices of classroom-based diagnostic assessment in differentiated instruction. In *Differentiated instruction, equity, and inclusion in language education*. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-6497-0.ch007

Hsin, C.-T., Hsieh, M.-F., Tsao, Y.-L., Chang, W.-C., & Compton-Lilly, C. (2025). Teaching and learning indigenous languages: An Atayal book project. *Journal of Early Childhood Literacy, 25*(3), 556–586. https://doi.org/10.1177/14687984251320869

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108635547

Indihadi, D. (2021). Analisis kesalahan berbahasa. *Aphorisme: Journal of Language, Literature, and Education, 2*(1), 68–91. https://doi.org/10.37680/aphorisme.v2i1.671

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. Routledge.

Lee, I. (2015). Student teachers' changing beliefs on a pre-service teacher education course in Hong Kong. In T. Wright & M. Beaumont (Eds.), *Experiences of second language teacher education* (pp. 31–53). Palgrave Macmillan. https://doi.org/10.1057/9781137440068_3

Levinson, T., Carocca P., F., & Bers, M. (2026). Coding as another language: An early childhood programming curriculum in Argentina. *Computer Science Education, 36*(2), 278–298. https://doi.org/10.1080/08993408.2025.2455339

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.

Moraes, I. T., Furtoso, V. A. B., & Giraldo, F. (2024). Developing pre-service teachers' language assessment literacy: An analysis of classroom tasks. *Ilha do Desterro, 77*, e96898. https://doi.org/10.5007/2175-8026.2024.e96898

Mottafegh, A., & Ahn, G.-N. (2026). Adaptive human-in-the-loop optimization using language-guided priors for chemical experiments. *Journal of Chemical Information and Modeling, 66*(15), 8833–8847. https://doi.org/10.1021/acs.jcim.6c00976

Mullah, N. S., & Wan Zainon, W. M. N. (2026). AI and low-resource languages: Bridging the linguistic divide. In *Reshaping language and cognition in education through AI* (pp. 77–127). IGI Global. https://doi.org/10.4018/979-8-3373-4397-6.ch003

Musyarrofah, H., Kurniawati, A. A., Asad, A. S. A., & Fadlillah, A. (2024). Differentiated instruction in ELT: Indonesian pre-service and in-service teachers' capabilities, strategies, and barriers. *JEELS (Journal of English Education and Linguistics Studies), 11*(2), 897–926. https://doi.org/10.30762/jeels.v11i2.2847

Nazilah, A. S. (2024). Diagnostic assessment in differentiated learning: Supporting learners' needs for improved learning achievement. *Research Journal on Teacher Professional Development, 2*(1), 87–92. https://doi.org/10.21580/rjtpd.v2i1.20662

Nurgiyantoro, B. (2018). *Penilaian pembelajaran bahasa berbasis kompetensi* [Competency-based language learning assessment] (2nd ed.). BPFE Yogyakarta.

Ölmezer Öztürk, E. (2022). Developing language assessment literacy of EFL pre-service teachers through classroom assessment course. *International Journal of Contemporary Educational Research, 8*(3), 13–22. https://doi.org/10.33200/ijcer.932721

Pastukh, V., Beshley, P., Mokhun, S., Fedchyshyn, O., & Brych, L. (2026). Functional generative system architecture for natural language-driven control of cyber-physical systems. In *2026 IEEE 18th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2026) – Proceedings*. https://doi.org/10.1109/TCSET65181.2026.11460442

Prastikawati, E. F., Mujiyanto, J., Saleh, M., & Fitriati, S. W. (2023). Language assessment course: Impact on pre-service EFL teachers' language assessment literacy. *Proceedings of International Conference on Science, Education, and Technology, 7*(1), 160–167.

Prastikawati, E. F., Mujiyanto, Y., Saleh, M., & Fitriati, S. W. (2024). Teaching practicum: Does it impact to pre-service EFL teachers' language assessment literacy? *World Journal of English Language, 14*(3), 242. https://doi.org/10.5430/wjel.v14n3p242

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal, 100*(S1), 190–208. https://doi.org/10.1111/modl.12308

Putri, N. V. W., Fitri, A., Suwandi, E., Nadif, B., Diana, R. N., & A'yun, M. Q. (2026). Exploring language assessment literacy: Indonesian pre-service English teachers' perceived readiness for classroom assessment. *Journal of Research on English and Language Learning (J-REaLL), 7*(1), 212–228. https://doi.org/10.33474/j-reall.v7i1.24610

Rahmani, E. F., Riyanti, D., & Negara, K. Y. E. P. (2024). EFL pre-service teachers' experiences in implementing differentiated instruction during teaching practice program. *Jurnal Pendidikan Progresif, 14*(2). https://doi.org/10.23960/jpp.v14.i2.202490

Rasoolzadeh, N., Zhang, T., Gao, Y., van Dijk, J. M., Yang, Q., Tan, T., & Mann, R. M. (2025). Multimodal breast MRI language-image pretraining (MLIP): An exploration of a breast MRI foundation model. In R. M. Mann, T. Zhang, L. Han, G. Litjens, T. Tan, D. Truhn, S. Li, Y. Gao, S. Doyle, M. M. R., J. N. Kather, K. Pinker-Domenig, & S. Wu (Eds.), *Lecture Notes in Computer Science* (Vol. 15451, pp. 42–53). Springer. https://doi.org/10.1007/978-3-031-77789-9_5

Refoina, R., & Rustam. (2024). The diagnostic assessment strategy in the Merdeka Curriculum for Indonesian language learning in 7th grade at junior high school. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12*(2). https://doi.org/10.24256/ideas.v12i2.5572

Sepúlveda-Oviedo, E. H., Travé-Massuyès, L., Subias, A., Pavlov, M., & Alonso, C. (2025). Artificial intelligence in photovoltaic fault diagnosis: A natural language-based Topic-tSNE fusion analysis. *Energy and AI, 21*, 100558. https://doi.org/10.1016/j.egyai.2025.100558

Setiyadi, A. B. (2020). *Teaching English as a foreign language* (2nd ed.). Graha Ilmu.

Spear, C. F., Piasta, S. B., Lo, M.-T., Justice, L. M., & O'Connell, A. A. (2025). Examining naturally occurring language- and literacy-learning opportunities for children with disabilities across ECE/ECSE contexts. *Topics in Early Childhood Special Education*. https://doi.org/10.1177/02711214251372539

Sugiyono. (2019). *Metode penelitian kualitatif* [Qualitative research methods]. Alfabeta.

Suleman, M. A., & Idayanti, Z. (2024). Implementation of diagnostic assessment in Grade VI Teaching at the Right Level approach in Indonesian language lessons in elementary school. *Ibriez: Jurnal Kependidikan Dasar Islam Berbasis Sains, 9*(2), 137–148. https://doi.org/10.21154/ibriez.v9i2.603

Sulistio, P. H. (2023). Language assessment literacy: Cases of pre-service English teachers. *IJLECR (International Journal of Language Education and Cultural Review), 9*(1), 1–10. https://doi.org/10.21009/ijlecr.v9i1.33032

Susanto, G., Pickus, D., Espree-Conaway, D., Suparmi, Rusiandi, A., & Noviya, H. (2024). Indonesian language policy and perspectives on its implementation in promoting Bahasa Indonesia as an international language. *Cogent Arts & Humanities, 11*(1), 2364511. https://doi.org/10.1080/23311983.2024.2364511

Tarigan, H. G. (2015). *Menulis sebagai suatu keterampilan berbahasa* [Writing as a language skill] (Rev. ed.). Angkasa.

Universitas Negeri Surabaya. (2024). *Academic guidelines of the Indonesian Language and Literature Education Study Program*. Unesa Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wardhani, A. (2023). The implementation of diagnostic assessment to support differentiated instruction in English learning at SMK N 1 Pringapus. *English Education Journal, 13*(4), 586–595. https://doi.org/10.15294/8x9cmt48

Zapatero-Ayuso, J.-A., Díaz-Carretero, A., Gil-García, M.-I., & Gil-Madrona, P. (2026). Competence development of pre-service teachers during practicum: A qualitative study from the perspective of its protagonists. *Journal of New Approaches in Educational Research, 15*(1). https://doi.org/10.1007/s44322-026-00057-0

Zhong, X., & Hao, H. (2026). Cohesive gestures in discourse by Chinese-speaking individuals with aphasia. *Journal of Communication Disorders, 123*, 106665. https://doi.org/10.1016/j.jcomdis.2026.106665