

DIFFERENTIAL ENGLISH SUBSKILL GROWTH AMONG INDONESIAN PRE-SERVICE ELEMENTARY TEACHERS

Disa Evawani Lestari¹,

¹President University, Indonesia,

✉ (e-mail) disa.silaen@president.ac.id

Risa Fitria²,

²President University, Indonesia,

✉ (e-mail) disa.silaen@president.ac.id

Receive in	Revised in	Accepted in
13 May 2026	08 June 2026	13 June 2026

ABSTRACT

This study investigated the differential growth of English proficiency scores among Indonesian pre-service elementary teachers during their first year of study using an explanatory sequential mixed-methods design. The study involved 34 undergraduate students enrolled in an Elementary Teacher Education program utilizing full English-Medium Instruction (EMI) at a university in Indonesia. Quantitative data were collected through pre-test and post-test from TOEFL ITP practice sets. The tests were administered in the beginning of semester one and at the end of semester 2, while qualitative data were obtained through online questionnaires distributed after the post-test, exploring students' English use habits and learning experiences. The quantitative findings revealed that students' average English scores increased from 416.12 to 437.56, indicating a progression from CEFR A2 to B1 proficiency levels. However, the growth patterns across subskills were uneven, with Structure and Written Expression demonstrating the greatest growth while Listenign Comprehension showed the least improvement. Interestingly, the qualitative findings revealed an apparent discrepancy as many participants perceived listening as the skill they had improved the most. Qualitative results indicate that grammatical gains were driven by the students' emerging professional identity, an awareness of their future role as teachers in international schools. The study concludes that while current institutional and digital scaffolds enable students to navigate the linguistic demands of the program, sustainable growth in communicative subskills requires a pedagogical shift from passive input toward interactive classroom discussions and systematic lexical support.

Keywords: English subskill development, Pre-Service Elementary Teachers, English-Medium Instruction (EMI), Teacher Education.

INTRODUCTION

In the Indonesian higher education context, Elementary Teacher Education programs (traditionally known as *Pendidikan Guru Sekolah Dasar* or PGSD) face a unique professional challenge as graduates must operate as

generalists responsible for a broad range of subjects rather than as specialized subject-matter experts. As Indonesian elementary schools increasingly transition toward bilingual programs and English-Medium Instruction (EMI), the linguistic demands double for these pre-service teachers, who must navigate various pedagogical and linguistic hurdles (Sukmawati & Pujiani, 2024). They are now expected to demonstrate the adequate English competence necessary to effectively implement EMI practices in classroom instruction (Ratri et al., 2025). This is especially critical when implementing Content and Language Integrated Learning (CLIL), which requires teachers to integrate subject matter instruction with English use to scaffold learning for children whose cognitive and first-language development are still maturing (Khoiriyah et al., 2024). Therefore, teacher education programs must prepare future primary teachers with a balanced and well-rounded command of all language subskills to meet these evolving professional demands and effectively access global educational resources (Damayanti & Meisani, 2025).

Research in the Indonesian higher education context further underscores that English-Medium Instruction (EMI) environments can significantly facilitate

students' English proficiency improvement and broader academic development through sustained engagement with English-mediated learning materials (Lestari, 2020). Beyond individual skill gains, the implementation of EMI provides extensive language exposure that fundamentally reshapes students' learning experiences and institutional language practices (Al Hakim, 2021). However, while continuous exposure within an EMI setting offers a critical foundation for growth, existing literature suggests that academic exposure alone may not automatically guarantee balanced development across all language domains. Consequently, pedagogical interventions grounded in scientific findings are essential to supplement this classroom-based exposure, ensuring that teacher education programs provide the structured guidance required for pre-service teachers to effectively master the dual demands of content and language (Macaro et al., 2018).

Building on these professional expectations, the transition to EMI-mediated learning creates a significant cognitive and pedagogical hurdle, as pre-service teachers must process high-level academic registers while simultaneously internalizing complex subject-matter concepts. Research identifies this overlap

as a primary source of difficulty in EMI contexts, where linguistic limitations can inadvertently compromise the conceptual clarity of the instructional material (Macaro et al., 2018). In the Indonesian context, these challenges are amplified by the rapid expansion of bilingual initiatives, which often outpace the development of the necessary linguistic and instructional support structures as the implementation of EMI continues to expand despite various pedagogical and linguistic challenges (Sukmawati & Pujiani, 2024). Addressing these pressures is essential for ensuring that future teachers do not merely rely on an initial 'willingness' to teach in English but actually possess the substantive linguistic competence required for effective, real-time classroom communication (Ratri et al., 2025). Investigating the specific growth trajectories of language subskills within this academic engagement model is therefore vital for identifying where these dual demands hinder progress, ultimately ensuring that teacher education programs provide the balanced proficiency necessary for contemporary educational demands.

To effectively address these linguistic hurdles, it is necessary to move beyond traditional assessment models that view standardized tests merely as

summative benchmarks. While institutions predominantly rely on aggregate proficiency scores as indicators of academic readiness, this approach creates a significant research gap by overlooking the diagnostic value of tracking specific language competencies over time. Research indicates that total score improvements can frequently mask important variations across language domains, as overall achievement may not reveal fundamental problems in specific receptive or productive skills (Bridgeman et al., 2016). Even when students meet established proficiency thresholds, they often continue to face significant, hidden challenges with specific receptive tasks, such as comprehending academic lectures or technical reading materials (Brown et al., 2022). The novelty of this study lies in its granular approach, which transitions from a generic benchmarking model toward a 'diagnostic-by-design' perspective to provide reliable information beyond a total score (Fisne et al., 2022). By investigating the individual trajectories of listening, structure, and reading comprehension, teacher education programs can move toward a diagnostic evaluation that identifies specific underlying difficulties and offers the targeted support necessary to ensure graduates are professionally ready for the

multidimensional demands of the EMI classroom (Brown et al., 2022).

To provide the empirical foundation necessary for such diagnostic evaluations and curricular revisions, the present study investigates the longitudinal proficiency development of Indonesian pre-service elementary teachers through an integrated analysis of standardized test performance and qualitative student accounts. By exploring the intersection between measurable linguistic growth and the EMI-mediated academic experiences that facilitate it, this research aims to identify the specific supports required for success in a teacher education environment. Accordingly, the study addresses the following research questions:

1. To what extent does the English proficiency of Indonesian pre-service elementary teachers change over one academic year?
2. Are there significant differences in the growth of TOEFL ITP subskills over one academic year?
3. Which TOEFL ITP subskill demonstrates the greatest and least improvement?
4. What are the challenges and the supports do students perceive as most critical for navigating the

linguistic demands of an EMI environment?

Ultimately, the findings of this study are expected to yield significant institutional benefits by providing an empirical blueprint for the systematic revision of the Elementary Teacher Education (PGSD) curriculum. By identifying specific linguistic bottlenecks—such as potential stagnations in auditory processing that may be masked by overall score gains—the institution can transition from a model of generic academic exposure toward a more pedagogically supportive EMI environment. Practically, this research informs the implementation of targeted instructional interventions, including the development of subject-specific vocabulary glossaries and the use of diagnostic feedback to provide remedial guidance (Brown et al., 2022). Such a shift is essential for ensuring that teacher education programs produce graduates who do not merely survive an EMI environment but possess the balanced, multidimensional proficiency required to effectively scaffold learning for children in future bilingual and CLIL classrooms (Fisne et al., 2022; Damayanti & Meisani, 2025).

REVIEW OF RELATED LITERATURES

In contemporary academic assessment, English proficiency is defined not as a singular, unitary trait but as a multidimensional competence comprising distinct receptive and productive domains, such as listening, structural knowledge, and reading comprehension (Fisne et al., 2022). Research suggests that student success in English-Medium Instruction (EMI) environments is fundamentally determined by their command of these various linguistic sub-competencies. Because second language (L2) acquisition is influenced by an uneven distribution of instructional practices and learner agency, the development of these individual skills often occurs asymmetrically (Brown et al., 2022). This is particularly relevant in Indonesian elementary teacher education, where pre-service teachers are expected to operate as generalists who must eventually provide linguistic models across diverse subject areas using approaches like Content and Language Integrated Learning (CLIL) (Siregar et al., 2025). Successful integration of content and language requires a balanced proficiency to ensure that linguistic limitations do not compromise the conceptual clarity of instructional material (Siregar et al., 2025).

A critical technical gap identified in current language assessment research is the heavy reliance on aggregate proficiency scores, which frequently mask significant variations across individual language domains (Bridgeman et al., 2016). While institutions often prioritize total scores as indicators of academic readiness for admissions, this summative approach overlooks the diagnostic value of tracking specific sub-competencies over time (Brown et al., 2022). As Bridgeman et al. (2016) emphasize, it is necessary to "peel the onion"—moving beyond a single composite score to identify underlying, fundamental problems that students face during real-time academic tasks. Meta-analyses indicate that aggregate scores can hide critical disparities where high achievement in one area statistically compensates for stagnations in another, leading to a discrepancy between a student's "benchmark readiness" and their actual ability to navigate complex instructional environments (Brown et al., 2022).

In EMI settings, research identifies a recurring pattern of differential growth where formal linguistic competence often develops at a different trajectory than communicative auditory processing.

CEFR B2 level is widely viewed as the minimum necessary requirement for successful academic study in a second language. At these intermediate levels, students report "fundamental problems" with receptive tasks, often focusing so intently on linguistic decoding of lectures and readings that they are unable to process deeper subject-matter concepts. Furthermore, utilizing Lexile measures to compare student readiness with authentic course materials, Holzknecht et al. (2022) identified a significant "reading gap," finding that a substantial number of students operate below the complexity level of even the least difficult instructional texts used in their courses. This gap is compounded in the Indonesian context, where teachers in primary schools

often struggle to explain difficult concepts in English to students who have limited competencies (German et al., 2025).

To mitigate these linguistic hurdles, contemporary literature increasingly advocates for a shift from summative benchmarking toward a "diagnostic-by-design" perspective (Min & Cai, 2023). This model seeks to provide granular information beyond a total score to guide remedial learning and instruction (Min & Cai, 2023). By integrating these granular diagnostic evaluations with targeted interventions, institutions can ensure that graduates develop the balanced, multidimensional proficiency required to meet the communicative demands of the modern classroom (Fisne et al., 2022; Rachmajanti et al., 2023).

METHODS

Research Design

This study employed an explanatory sequential mixed-methods design, characterized by an initial phase of quantitative data collection and analysis followed by a qualitative phase to explain the statistical results. According to Creswell and Plano Clark (2018), this design is particularly effective when qualitative data are required to provide

deeper, contextualized explanations for identified statistical patterns. In the present study, the quantitative phase identified longitudinal growth patterns in TOEFL ITP subskills, while the qualitative phase explored the institutional supports and academic experiences that facilitated or hindered this development.

Participants and Sampling

The participants were 34 first-year undergraduate students from the 2025 cohort enrolled in the Elementary Teacher Education (PGSD) Study Program at an English-speaking university in Indonesia, consisting of 30 female students and 4 male students. The study utilized total sampling, including all students who completed both the pre-test and post-test. These pre-service teachers operate in an EMI-oriented environment where they regularly engage with English in both academic and non-academic settings. To evaluate the longitudinal effects of English-Medium Instruction (EMI) on linguistic development, the research participants were limited to first-year Elementary Teacher Education students to investigate how one year of consistent academic exposure in an English-mediated environment impacts their English proficiency growth and the resulting patterns of differential subskill development.

Instrumentation

The primary quantitative instrument was the TOEFL ITP practice test, consisting of Listening Comprehension (50 questions), Structure and Written Expression (40 questions), and Reading Comprehension (50 questions). The test

was administered twice: as a pre-test at the beginning of the academic year and a post-test at the end of the academic year. To answer the qualitative research questions, online questionnaires were distributed to all participants, focusing on learning experiences, language exposure habits, and the perceived effectiveness of institutional support systems.

Data Collection Procedures

Quantitative data were gathered through the two TOEFL ITP practice set administrations over one academic year to track longitudinal changes. The pre-test was administered in the beginning of semester one, while the post-test was at the end of semester two. After the post-test results were analyzed to identify the strongest and weakest growth patterns, the qualitative data collection commenced. An online questionnaire was distributed to identifying the specific pedagogical hurdles and strategies used by students or supports available to navigate the EMI environment.

Data Analysis

Quantitative data were analyzed using descriptive and inferential statistics. Paired-samples t-tests were conducted to

determine the significance of changes between pre-test and post-test scores for both aggregate and subskill results. Gain scores and effect sizes were computed to identify the magnitude of improvement in specific domains. Qualitative data from questionnaires were analyzed using

thematic analysis following the framework by Braun and Clarke (2006). Finally, the qualitative findings were integrated with the quantitative results to explain the observed patterns of differential TOEFL subskill growth.

RESULTS AND DISCUSSION

This study investigated the differential growth of TOEFL subskills among Indonesian pre-service elementary teachers over one academic year. Of the initial participants, complete pre-test and post-test TOEFL score data were obtained from 34 students and were therefore included in the quantitative analysis. The analysis focused on changes in overall TOEFL scores as well as the three TOEFL subskills: Listening Comprehension, Structure and Written Expression, and Reading Comprehension.

Overall English Proficiency Development

Table 1 presents the descriptive statistics of students' overall TOEFL scores in the pre-test and post-test. The quantitative analysis revealed that average TOEFL scores increased from 416.12 in the pre-test to 437.56 in the

post-test. This represents a mean gain score of 21.44 points over one academic year. According to the CEFR framework, these results indicate a general progression from basic user proficiency (A2) to independent user proficiency (B1).

Table 1

TOEFL Score Development

Test	Mean	Gain Score
Pre-test	416.12.0 0	
Post-test	437.56.0 0	+21.44

Differential Growth of TOEFL Subskills

English proficiency development did not occur uniformly across language domains. Table 2 shows while the overall score improved, the growth patterns for specific subskills were markedly uneven.

Table 2

Comparison of TOEFL Subskills

Subskill	Pre-test Mean	Post-test Mean	Gain Score
----------	---------------	----------------	------------

Listening Comprehension	19.44	19.59	+0.15
Structure & Written Expression	15.50	19.09	+3.59
Reading Comprehension	24.71	26.59.00	+1.88
Overall TOEFL Score	416.12.00	437.56.00	+21.44

The findings show uneven growth patterns across TOEFL subskills. Among the three subskills, Structure and Written Expression demonstrated the greatest improvement, with a mean gain score of +3.59 points. Meanwhile, Reading Comprehension showed moderate improvement (+1.88), whereas Listening Comprehension demonstrated only minimal improvement (+0.15).

Table 3

Paired-Sample T-Test Results

Domain	Pre-Test Mean	Post-Test Mean	Mean Gain	t-statistic	p-value
Overall TOEFL	416.12	437.56	+21.44	3.089	.0041
SWE	15.50	19.09	+3.59	4.265	.0002
RC	24.71	26.59	+1.88	1.177	.2474
LC	19.44	19.59	+0.15	0.145	.8859

To determine the statistical significance of the changes in English proficiency over the one-year study period, a paired-samples t-test was conducted for both the overall TOEFL scores and the individual language

subskills. As can be seen in Table 3, the paired-samples t-test results reveal statistically significant improvements in two areas: the overall TOEFL scores ($t = 3.089$, $p = .004$) and Structure and Written Expression ($t = 4.265$, $p < .001$). These findings indicate that the one-year EMI environment effectively fostered grammatical precision. Conversely, the gains in Reading Comprehension ($p = .247$) and Listening Comprehension ($p = .886$) were not statistically significant, highlighting a developmental asymmetry where independent auditory and reading subskills remained stagnant despite the students' overall proficiency growth.

The quantitative data revealed that Structure and Written Expression demonstrated the most substantial growth among all language domains, with a mean gain score of +3.59 points. Qualitative responses indicate that this growth is driven by the students' emerging professional identity as future teachers, which necessitates high grammatical precision. Students reported that the EMI environment forced them to "*thoroughly refine foundational English skills*" because, as future elementary educators, they cannot rely on casual fluency but must "*have a solid grasp of grammar to effectively break down and*

deliver lessons to young learners" [S12]. S26 also explained: *"I feel that mastering English grammar is a professional necessity for me; as a teacher at an international school, "I have to be able to construct simple and correct sentences so that when I deliver a lesson, my students can focus on the subject matter without being confused by my language use"*.

One student noted that their understanding of grammar evolved from simple rote memorization to practical application, stating, *"I have become more attentive to structural precision and the nuances of formal grammar... to provide clear and correct language models"* [S13]. Additionally, the frequency of academic writing assignments and the specialized instruction in Research Paper Writing courses were cited as instrumental, as these tasks moved beyond general communication to focus on the *"technical precision required to construct coherent and professionally structured academic arguments"* [S23].

In stark contrast to the development in grammatical structure, Listening Comprehension demonstrated the weakest quantitative growth, with a minimal gain of only +0.15 points. Interestingly, these results are in direct

opposition to students' self-perceptions in the questionnaire, where a majority of participants rated their improvement in understanding lectures very highly (often 4 or 5 on a 5-point scale).

Students frequently reported that their ability to follow English-mediated instruction had significantly improved, with one student remarking, *"I've become more used to listening to the lecturer's explanations, understanding academic terms, and following class discussions"* [S32]. Another student shared, *"I can handily understand the materials being delivered the majority of the time"* [S17].

This discrepancy suggests that while standardized listening (as measured by the one-off audio delivery of the TOEFL) remains stagnant, classroom listening has improved through interactional scaffolding. Unlike the TOEFL format, students in the EMI classroom manage comprehension hurdles by actively using coping strategies. For example, students noted they handle difficulties by proactively asking for clarification using phrases like *"Pardon?"* or *"Could you elaborate on what you mean by..."* [S13]. Others rely on social scaffolding, stating they *"ask friends to clarify what I missed"* or *"write down keywords and look up the meaning*

afterward" [S12, S32]. Consequently, students feel they are successfully navigating the lecture because they have the support of lecturers and peers to fill linguistic gaps, a luxury not afforded during standardized proficiency testing. The students demonstrating the most significant point gains in their overall TOEFL scores explained that the success is attributed to high learner agency, specific coping strategies, and significant time dedication.

The following results for Research Question 4 explain the observed differential TOEFL scores growth by triangulating the students' academic habits, linguistic challenges, and perceived support systems.

Perceived Challenges and Supports

The majority of pre-service teachers identify three primary hurdles that create a bottleneck in their academic engagement. Firstly, vocabulary load is the most pervasive challenge. Students (53%) frequently report that unfamiliar academic and technical terminology—specifically scientific and educational vocabulary—renders entire lectures or readings confusing. As S11 noted, *“a single unknown word can act as a barrier*

to the conceptual clarity of the entire subject”.

The second challenge is speed of instructional delivery. The rapid pace of lectures often leaves students no time to absorb the information. As S23 recounted, *“I often find it difficult to keep pace with the lectures because the speed of delivery is so fast; I get stuck translating the English explanations into Indonesian just to grasp the meaning, and by the time I understand it, the lecturer has already progressed much further ahead in the material”*.

Thirdly, some students reported that a unique challenge often arises during collaborative tasks due to unfamiliar accents; students find it difficult to comprehend the diverse accents of international classmates, which hinder *“effective knowledge transfer and brainstorming”* during classroom discussion [S33]. Lastly, A significant finding in the questionnaire is the students' critique of current instructional methods and tasks. Many participants explicitly argued that their real-time listening comprehension would improve more effectively if the classroom instructions moved away from one-way delivery. Students noted that sessions *“dominated by lecture session or*

reading tasks followed by individual/group presentation" are often too one-way, leaving less opportunities for real-time language processing [S17]. Participants perceived that interactive classroom discussions provide a more effective platform for listening growth than passive lecture attendance or presentation task. Participants perceived mandatory classroom discussions as valuable opportunities to develop their oral proficiency. They explained that being required to contribute to discussions encouraged them to process language more actively and refine their spoken English through real-time interaction. As one participant noted, *"Class discussions make me use English whether I feel ready or not. Being required to participate gives me more chances to practice speaking, and I learn to respond naturally during conversations"* (S14).

To overcome academic challenges, students rely on a combination of interactional, social, and digital scaffolding.

First, students perceive simplification strategies by lecturers as the most critical factor for classroom success. They highlight the effectiveness of instructors who explain concepts using *"simple and*

clear words step-by-step" [S29]. Additionally, the ability to proactively ask for real-time clarification *"Could you elaborate?"*) is seen as an essential lifeline for survival [S13, S4, S7].

In addition to that, peer-mediated support is highly valued. The BUDDY program and informal group discussions provide a more relaxed setting to negotiate meaning. As S32 explained, *"I believe that practicing how to express academic ideas in English with my peers in a low-pressure setting would help a lot; it makes me feel more confident to participate in class and manage my assignments"*. Students often use translanguaging with peers to ensure they understand complex topics. S12 recounted, *"the support from my peers is very important; being able to discuss materials together in a mix of Indonesian and English makes it much easier to understand complex topics without feeling overwhelmed by the language barrier"*. Furthermore, many of them also admitted that the presence of international peers in class is stimulating, increasing their motivation to keep improving their English proficiency and intercultural understanding. S8 argued, *"seeing my friends from other countries use English so fluently and confidently*

makes me want to try much harder to improve my own proficiency and understanding their different accent".

Lastly, digital and AI mediation emerges as a pervasive and sophisticated coping strategy used by pre-service teachers to bridge the gap between their current proficiency levels (typically CEFR A2/B1) and the high-level academic demands of an EMI environment. Rather than merely serving as a crutch, these tools function as a form of self-directed scaffolding that enables students to manage cognitive load and refine their professional output. A significant shift in this student cohort is the move from basic translation to using Generative AI for linguistic refinement and error correction. Some of them feel a professional urgency to provide correct language models as future teachers. Therefore, some initiatives like “*speaking with AI*” and “*discussing with a chatbot*” appear frequently in the questionnaire response [S10, S8, S5, S17].

Interestingly, for most high-performing students, digital mediation is used as a strategic tool for triangulation. They explained that when encountering dense scholarly articles, their primary strategy is leveraging various artificial intelligence tools to cross-reference

explanations. As S23 explained: “*When I encounter dense scholarly articles that are hard to grasp, my primary strategy is leveraging various artificial intelligence tools to cross-reference different explanations and ensure I have a consistent understanding of the content*”. This allows them to simplify complex academic jargon into more digestible concepts, ensuring that conceptual clarity is not lost to linguistic complexity. Furthermore, participants highlighted the critical role of teacher-led instruction in this process, noting they felt “*lucky to be introduced to AI-powered tools—such as Gemini and NotebookLM—by their lecturers* [S22]. They argued that practicing prompting and key features as a classroom activity empowered them to become more independent learners who could critically manage their own acquisition process [S7, S14, S33]. This self-directed digital scaffolding also serves as a vital affective buffer; students reported that “*receiving feedback from AI first reduces the teacher apprehension or fear of evaluation,*” providing a psychologically safe environment to refine their professional output privately before it is judged by an instructor [S16].

To summarize, there is a clear discrepancy between standardized results

and perceived ability. While quantitative data showed minimal growth in Listening Comprehension (+0.15), students perceived it as one of the easiest areas to handle. This is because the EMI classroom allows for interactive scaffolding and mediated learning that a standardized test does not. However, the students' call for more discussion suggests that while they are surviving lectures through these scaffolds, they believe they would thrive more in an interactive, dialogic environment.

DISCUSSION

The findings of this study demonstrate that Indonesian pre-service elementary teachers experienced significant overall English proficiency gains over one academic year, progressing from CEFR A2 (Basic User) to B1 (Independent User). However, this development was markedly asymmetrical, with formal grammatical gains (+3.59) and reading (+1.88) significantly outpacing listening comprehension (+0.15). The following discussion situates these patterns of differential growth within established linguistic and pedagogical theories.

The substantial improvement in Structure and Written Expression may

also reflect the influence of the university assessment practices. They claim that they are frequently evaluated through written assignments, academic papers, quizzes, and examinations that require grammatical accuracy and structural awareness. Repeated exposure to such academic tasks may gradually strengthen students' understanding of English sentence structures and written conventions.

The substantial improvement in Structure and Written Expression is deeply connected to the participants' emerging professional identity. As pre-service teachers in an EMI-oriented program, they recognized their future roles as linguistic models for young learners, which created a sense of professional urgency to master grammatical precision. This phenomenon is best explained by Schmidt's (1990) Noticing Theory, where students' professional identity acts as a powerful cognitive filter, forcing them to notice and refine their own linguistic gaps to ensure pedagogical clarity in their future classrooms.

The significant growth observed in high-performing students is a direct result of their learner agency and personal investment. Participants demonstrated

high levels of agency, with top gainers dedicating up to 14 hours per week to independent study beyond the classroom. According to Norton (2013), these students are not merely motivated by a psychological trait; they are investing in English because they believe it will yield "symbolic and material resources" in the future—specifically, high-prestige careers in international or bilingual schools. This investment drives them to seek out additional English exposure and use digital tools for self-directed linguistic refinement

Recent studies indicate that primary teachers are increasingly expected to demonstrate not only a willingness to implement English-Medium Instruction (EMI) but also the adequate English competence necessary for professional responsibility (Ratri et al., 2025). This shift in professional expectations has been linked to teaching excitement, where educators feel challenged and enthusiastic about developing their English skills to subjects beyond language itself, such as science and mathematics. These teachers view their proficiency as a means to increase school credibility and their own global competitiveness (Fanani et al., 2025).

In this study, Reading Comprehension also demonstrated improvement, although the gain was smaller than that of grammatical competence. The increase in reading performance may be related to students' growing exposure to English academic texts throughout their university studies. Students in EMI-oriented institutions frequently interact with English textbooks, journal articles, PowerPoint presentations, and digital learning resources, which may gradually improve vocabulary recognition and reading familiarity. Nevertheless, the relatively moderate increase suggests that reading proficiency development involves more complex cognitive processes beyond vocabulary recognition, including inference-making, identifying main ideas, and strategic reading under time constraints (Hafid et al. (2021). These difficulties in TOEFL reading sections are caused by vocabulary limitations and ineffective reading strategies. Many university learning activities continue to emphasize written assignments, grammar-focused instruction, academic writing, and text-based assessments. As a result, students may receive more opportunities to practice grammatical accuracy and written language production

than communicative listening skills. This finding aligns with Damayanti and Meisani (2025), who argued that teacher education programs in Indonesia often prioritize academic literacy and pedagogical competence through written academic tasks. Similarly, Khoiriyah et al. (2024) found that pre-service teachers involved in CLIL implementation tended to rely heavily on written preparation and structured language use during instruction.

A central paradox in this study is the scaffolding discrepancy: while quantitative scores for Listening Comprehension remained nearly stagnant, students perceived their classroom listening as highly successful. This phenomenon is an operationalization of Vygotsky's (1978) Zone of Proximal Development (ZPD). The TOEFL ITP practise set measures independent performance (what a student can do alone), where students failed to show growth. Conversely, the EMI classroom provides assisted performance (what a student can do with support). Lecturers utilized simplification strategies, such as step-by-step explanations and clear language, which functioned as a temporary support structure. The BUDDY program, where learners

negotiated meaning and received immediate feedback through translanguaging, acts as a social scaffolding allowing them to function at a higher conceptual level than their independent linguistic test results suggest.

In the context of English-Medium Instruction (EMI), learner agency and self-directed digital scaffolding serve as critical determinants of a pre-service teacher's ability to navigate the asymmetrical development of language subskills. In this study, students exercised high levels of agency by actively seeking English exposure beyond the classroom—with top-performing students dedicating up to 14 hours per week to independent study—to bridge the gap between their current CEFR A2/B1 proficiency and the B2 level typically required for academic success.

Studies have shown that providing individualized diagnostic feedback through AI platforms effectively facilitates learners' writing performance across multiple linguistic levels (Alharbi, 2023; Marzuki et al., 2023). This intervention has been linked to a significant increase in students' self-efficacy and confidence in academic writing, as learners feel more secure in

their ability to express complex ideas. A key factor in this confidence gain is the ability of AI tools to act as a partner for a self-directed scaffolding, allowing students to simplify complex academic jargon into more digestible concepts and ensure that conceptual clarity is not compromised by linguistic complexity. By leveraging digital resources for self-directed scaffolding, pre-service teachers can effectively function within their Zone of Proximal Development (ZPD), performing at higher conceptual levels than their independent standardized test scores might otherwise indicate.

The findings regarding digital mediation reveal that participants utilized AI-powered tools not as mere crutches, but as a sophisticated form of self-scaffolding that enabled them to bridge the gap between their current proficiency and the demands of the EMI environment. AI tools are pedagogically embedded to support learners in taking control of their own acquisition processes (Busso & Sanchez, 2024). Specifically, the participants' gratitude for classroom

activities focused on prompting and key features indicates a successful transition toward AI literacy—the critical ability to interact with technology effectively and ethically.

Rather than succumbing to thoughtless reliance, these pre-service teachers leveraged chatbots and platforms like NotebookLM for strategic triangulation, using them to simplify complex academic jargon and ensure that conceptual clarity was not sacrificed to linguistic complexity. This self-directed digital scaffolding also served as a vital affective buffer; by receiving immediate and personalized feedback from AI first, students reduced their teacher apprehension, creating a psychologically safe space to refine their professional output before human. Ultimately, the teacher's role in this study shifted from a traditional instructor to an essential innovator and trainer, providing the structured AI literacy training necessary for students to become independent learners.

CONCLUSIONS

This study investigated the differential growth of English subskills among 34 Indonesian pre-service

elementary teachers during their first year in an English-Medium Instruction (EMI) environment. Utilizing an explanatory

sequential mixed-methods design, the research revealed that while participants' overall English scores improved from 416.12 to 437.56 (CEFR A2 to B1), development across language domains was markedly asymmetrical.

Structure and Written Expression demonstrated the greatest growth (+3.59 points), driven by the students' emerging professional identity to serve as accurate linguistic models and targeted institutional scaffolding in academic writing courses. Conversely, Listening Comprehension showed the least quantitative improvement (+0.15 points), despite students' perceptions that navigating lectures was manageable. This scaffolding discrepancy indicates that while students function effectively in the classroom through interactional teacher support, the BUDDY program, and digital/AI mediation, their independent auditory processing skills remain stagnant. Furthermore, students identified one-way, lecture-heavy instructional formats as a significant barrier to real-time listening development.

The findings of this study are subject to several limitations. First, the study involved a relatively small sample of 34 participants from a single university, which may limit the generalizability of the results to other teacher education contexts

in Indonesia. Also, the research relied primarily on TOEFL performance and self-reported questionnaire data. It did not include direct, longitudinal classroom observations of language use in real-time.

Based on the results, the following recommendations are proposed. First, future research should utilize longitudinal designs and direct classroom observations to investigate whether structured training in AI-powered self-scaffolding can permanently bridge the scaffolding discrepancy observed between pre-service teachers' assisted classroom performance and their independent linguistic growth across different language subskills. Second, institutions should move toward interactive classroom discussions to provide the meaningful output necessary for listening and speaking growth. Third, teacher education programs should incorporate specialized vocabulary glossaries and systematic linguistic scaffolding across all pedagogical content subjects to help pre-service teachers overcome the vocabulary bottleneck. Finally, further research is needed to examine the long-term impact of self-directed digital and AI mediation on subskill development and whether these tools can permanently bridge the gap

between independent and assisted performance.

ACKNOWLEDGMENTS

The author would like to express sincere gratitude to President University for providing financial support for this research. The author also extends appreciation to the Elementary Teacher

Education Study Program students from the 2025 cohort who participated in this study and willingly contributed their time and experiences throughout the data collection process

REFERENCES

- Al Hakim, M. A. (2021). Revealing the potential impacts of English-medium instruction (EMI) in Indonesian higher education context. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 9(2), 461–472.
- Alharbi, Wael. AI in the Foreign Language Classroom: A Pedagogical Overview of Automated Writing Assistance Tools, *Education Research International*, 2023, 4253331, 15 pages, 2023. <https://doi.org/10.1155/2023/4253331>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, Dale., Yaguchi, Michiko., Murray, Lewis., & Oyabu, Kana. (2022). The relationship between English proficiency and student experiences in EMI courses. In *外国語教育フォーラム: 金沢大学外国語教育論集= Forum of Language Instructors* (Vol. 16, pp. 17-35). 金沢大学国際基幹教育院外国語教育系.
- Busso, A., & Sanchez, B. (2024). Advancing communicative competence in the digital age: A case for AI tools in Japanese EFL education. *Technology in Language Teaching & Learning*, 6(3), 1211-1211.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Sage Publications.
- Damayanti, I. L., & Meisani, D. R. (2025). Teacher education for primary English teachers. In *The Routledge International Handbook of English Language Education in Indonesia* (pp. 358–375). Routledge.
- German, E., Pujiastuti, A., & Estillomo, R.B. (2025). English as a Medium of Instruction (EMI) in Indonesian elementary schools: Features, problems, possibilities, and prospects. *EduLite: Journal of English Education, Literature and Culture*.

- Fanani, R. A., Syahria, N., & Nurjati, N. (2025). Exploring A Private Islamic Primary School Teachers'perspectives On English As Medium Of Instruction (Emi): A Phenomenological Case Study. *Jurnal Pendidikan Indonesia*, 6(4).
- Fişne, F. N., Sata, M., & Karakaya, I. (2022). Standard setting in academic writing assessment through objective standard setting method. *International Journal of Assessment Tools in Education*, 9(1), 80-97.
- Hafid, N., Purnawarman, P., & Sukyadi, D. (2021). Investigating constraints and strategies faced by the non-English students in answering reading comprehension TOEFL test. *English Review: Journal of English Education*, 9(2), 247–256.
- Khoiriyah, K., Idris, M. F. B., Rahma, A. P., & Ulhaq, A. H. D. (2024). Teaching science through English: Pre-service teachers' implementation of CLIL approach in primary school. In *Strengthening Professional and Spiritual Education Through 21st Century Skill Empowerment in a Pandemic and Post-Pandemic Era* (pp. 89–96). Routledge.
- Lestari, D. E. (2020). Behind the improvement of students' English proficiency in an EMI university. *Jurnal SMART*, 6(1), 10–23.
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1), 36–76. <https://doi.org/10.1017/S0261444817000350>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469.
- Rachmajanti, S., Anugerahwati, M., & Unsiah, F. (2023). A profile of primary school students' literacy through EMI in CLIL Contexts. *JEES (Journal of English Educators Society)*, 8(2). <https://doi.org/10.21070/jees.v8i2.1802>
- Ratri, D. P., Gozali, A., Inayati, D., & Razali, K. A. B. (2025). From willingness to competence: Investigating primary teachers' readiness for EMI implementation. *PANYONARA: Journal of English Education*, 7(2), 247–266.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Sukmawati, I. D., & Pujiani, T. (2024). English as medium of instruction (EMI) at primary schools in Indonesia: Challenges and opportunities. *The Journal of English Teaching for Young and Adult Learners*, 3(1), 14–20.