

A Comparative Case Study of the Effectiveness of an Automated Writing Evaluation System in Improving Grammatical Accuracy Across Myanmar EFL Students of Varying English Proficiency Levels

Nan Ingyin Phyu^a and Tanyaporn Arya ^{b*}

^a English as an International Language Program, Graduate School, Chulalongkorn University, Thailand

^b Chulalongkorn University Language Institute, Chulalongkorn University, Bangkok, Thailand

*Corresponding author: tjarya@gmail.com

Article information	
Abstract	This exploratory case study examines how Grammarly combined with reinforcement strategies impacts grammatical accuracy among three Myanmar EFL students at CEFR levels A2, B1, and B2. Over six weeks, participants completed writing tasks using Grammarly's free version alongside weekly one-on-one meetings and self-reflection with Grammar in Use resources. Progress was assessed through pre- and post-tests and weekly error tracking. Findings indicate positive effects across all levels, with differentiated patterns: the A2 learner showed dramatic improvement relying heavily on automated feedback for remedial support; the B1 learner demonstrated steady progress requiring balanced feedback and reflection for consolidation; the B2 learner achieved refinement using AWE selectively while depending primarily on reinforcement for deeper understanding. Weekly tracking revealed non-linear improvement trajectories, with the greatest gains at lower proficiency levels. While findings cannot be generalized statistically due to the small sample, they provide analytical insights into how learners at different developmental stages process automated feedback. The study offers practical guidance for educators integrating AWE into CEFR-aligned instruction in resource-constrained EFL contexts.
Keywords	Automated Writing Evaluation (AWE), Grammatical Accuracy, CEFR Proficiency Levels, Myanmar EFL Learners, English Writing
APA citation:	Phyu, N. I., & Arya, T. (2026). A comparative case study of the effectiveness of an automated writing evaluation system in improving grammatical accuracy across Myanmar EFL students of varying English proficiency levels. <i>PASAA</i> , 72, 300–329.

1. Introduction

Grammatical accuracy is a fundamental aspect of English language proficiency, influencing the clarity and effectiveness of written communication (Savage, Bitterlin, & Price, 2010). In academic and professional contexts, precise grammar ensures that ideas are conveyed accurately and helps prevent misinterpretation. Despite its importance, many learners face

challenges in achieving grammatical accuracy, particularly in EFL contexts such as in Myanmar, where students often demonstrate stronger receptive skills in listening and reading than productive skills in writing.

Technological tools, such as automated writing evaluation (AWE) systems, have been introduced to support learners in improving their writing accuracy. Grammarly, a widely used AWE tool, provides automated feedback on grammar, vocabulary, and writing mechanics, allowing learners to identify and correct errors independently (Nova, 2018). Prior research suggests that Grammarly can enhance learners' awareness of grammatical errors and support improved writing performance (Fan & Ma, 2022; Miranty & Widiati, 2023). However, limited studies have investigated how such tools combined with reinforcement would affect learners at varying proficiency levels, and little evidence exists in the Myanmar EFL context.

This study attempts to address this gap by examining the effectiveness of Grammarly and reinforcement strategies in improving the grammatical accuracy among three Myanmar EFL students at CEFR levels A2, B1, and B2. Specifically, it explores how automated feedback and reinforcement strategies impact error reduction over a six-week intervention and whether learners at different proficiency levels respond differently to these tools.

This case study aims to address the following questions:

1. What is the effect of Grammarly and reinforcement on the development of grammatical accuracy over time of Myanmar EFL learners at CEFR proficiency levels A2, B1, and B2?
2. To what extent does the effect of Grammarly and reinforcement on grammatical accuracy differ among Myanmar EFL learners at CEFR levels A2, B1, and B2?

To address RQ1, pre-test and post-test writing tasks are analyzed for the frequency and types of correctness errors using Grammarly reports, and OSPT scores are compared to assess changes in grammatical knowledge following the six-week intervention. To address RQ2, weekly error data from the 24 writing tasks are analyzed to compare error reduction trajectories across the three proficiency levels, alongside qualitative observations from weekly meetings to examine differences in learner engagement with automated feedback.

This study adopts an exploratory qualitative case study design to examine how three Myanmar EFL learners at distinct CEFR levels (A2, B1, B2) engage with Grammarly combined with reinforcement over six weeks. This approach enables detailed observation of individual trajectories and differentiated responses—insights obscured in larger quantitative studies (Yin, 2018). While not statistically generalizable, findings provide analytical generalization to feedback processing theory and CEFR-aligned instruction (Stake, 1995).

2. Literature Review

2.1 The Role of Grammar in Language Learning

Grammar provides the structural foundation of language, enabling learners to communicate accurately and coherently (Celce-Murcia & Larsen-Freeman, 1999; Thornbury, 1999). It supports the development of all four language skills: in receptive skills, it aids comprehension by signaling meaning through sentence structure and tense, while in productive skills, it ensures clarity and coherence. Writing demands the highest level of grammatical precision, as errors can obscure meaning and reduce effectiveness.

EFL learners frequently struggle with accuracy in areas such as verb tense, subject-verb agreement, and sentence structure, which often weakens overall communication (Ridha, 2012; Sawalmeh, 2013). Addressing these issues is essential to fostering proficiency and learner confidence.

2.2 Studies on Common Grammatical Errors in EFL Learners' Writing

Research consistently shows that grammatical errors are among the most frequent challenges for EFL learners, directly affecting writing accuracy and overall communication (Ridha, 2012; Singh et al., 2017; Nguyen, 2020; Thi et al., 2023; Inpanich et al., 2023). Studies across diverse EFL contexts highlight recurring difficulties with verb tense, subject-verb agreement, articles, and word choice.

Ridha (2012) identified tense, articles, and prepositions as dominant error types among Iraqi undergraduates. Similar findings were observed by Singh et al. (2017) with Malaysian diploma students and Nguyen (2020) with Vietnamese learners, who both noted verb tense and subject-verb agreement as persistent issues. More recent studies indicate that these challenges persist in Southeast Asian contexts. In Myanmar, Thi et al. (2023) found that EFL learners frequently produced language-related errors involving sentence structure, verb forms, verb tense, subject-verb agreement, articles, conjunctions, and prepositions. Likewise, Inpanich et al. (2023) reported that grammatical errors remained prevalent among Thai EFL learners across different essay genres.

Across these contexts, verb-related errors (tense, form, and agreement) appear most widespread, followed by articles and word choice. Such recurring issues highlight the need for targeted pedagogical strategies. Automated Writing Evaluation (AWE) tools can offer immediate corrective feedback on these error types, supporting learners' grammatical development across proficiency levels. Therefore, researchers have implemented various instructional approaches to support learners' grammatical development across proficiency levels.

2.3 Studies on Enhancing Grammatical Accuracy in Writing via Instructional Methods

Ahangari and Babapour (2015) investigated the impact of self-correction and peer-correction on 80 female EFL learners aged 15 to 19 at the Iran Language Institute. Participants engaged in self-correction, where teachers highlighted errors for individual revision, or peer-correction, providing feedback on classmates' work over seven weeks. Results indicated that self-correction significantly improved writing accuracy across low-intermediate and high-intermediate learners, promoting learner autonomy and reducing error-related anxiety. Peer-correction, however, did not show statistically significant improvements, highlighting the differential effectiveness of feedback types in supporting grammatical development.

Similarly, Solfiyatzahro et al. (2019) implemented classroom action research to enhance grammatical accuracy among tenth-grade Indonesian students producing descriptive texts. Teacher's Written Corrective Feedback (WCF) was provided through direct corrections on drafts, oral explanations of grammatical rules, and revision sessions. This multi-step approach allowed learners to actively engage with feedback, resulting in improved writing performance and positive learner perceptions of the process. These findings demonstrate that

structured feedback, in traditional instructional settings, effectively supports the development of grammatical accuracy in writing.

Particularly in contexts where learners may have limited access to personalized feedback in classrooms, such as in Myanmar's EFL settings, Automated Writing Evaluation tools have the potential to address learners' grammatical challenges by providing immediate, consistent, and detailed feedback. Real-time corrections reduce reliance on teachers, enhance learner autonomy, and allow students to engage interactively with grammatical rules. By complementing traditional instructional strategies, AWE tools offer a dynamic approach to fostering grammatical accuracy.

2.4 Studies with Automated Writing Evaluation (AWE)

Automated Writing Evaluation (AWE) systems, such as QuillBot and e-rater, provide varying features from basic grammar checks to advanced style suggestions, enabling targeted writing improvement (Dembsey, 2017). Among available tools, Grammarly has gained widespread adoption in academic settings due to its accessibility and comprehensive feedback capabilities.

2.4.1 Grammarly as an AWE tool

Grammarly offers multiple plans: the free version provides grammar, spelling, and punctuation checks, while advanced versions include style suggestions and plagiarism detection (Grammarly, 2020). The free version categorizes errors as correctness (grammatical accuracy: subject–verb agreement, verb tense, articles, sentence structure) and clarity (readability: wordiness, complex phrasing). It employs color-coded feedback; red for grammar, blue for clarity for immediate user-friendly correction (Dembsey, 2017). The free version detects approximately 150 error types, covering common EFL challenges suitable for learners across CEFR levels (Abu Guba et al., 2024).

2.4.2 Research on AWE Effectiveness in Grammar Development

Studies demonstrate AWE tools enhance grammar and mechanics while fostering autonomy, though effectiveness varies across proficiency levels and requires pedagogical support.

Research on Automated Writing Evaluation (AWE) tools generally indicates that they can improve learners' grammatical accuracy and writing mechanics by providing immediate and individualized feedback. Studies have reported improvements in grammar, spelling, punctuation, and other language-related features among learners using Grammarly, suggesting that automated feedback can facilitate error recognition and revision (Qassemzadeh & Soleimani, 2016; Ghufon & Rosyida, 2018; Dizon & Gayed, 2021). In addition, Grammarly has been found to support learner autonomy by allowing students to independently identify and correct errors during the writing process (Nova, 2018; Qassemzadeh & Soleimani, 2016).

However, the effectiveness of AWE tools is not uniform across all learners and writing dimensions. Research suggests that learners with stronger linguistic foundations may benefit more from automated feedback, while lower-proficiency learners often demonstrate greater improvement in grammatical accuracy than in broader aspects of writing such as fluency and content development (Ghufon & Rosyida, 2018; Dizon & Gayed, 2021). Furthermore, although Grammarly demonstrates relatively high error-detection accuracy, it may

occasionally provide inaccurate suggestions and is generally less effective in addressing higher-order concerns such as coherence, organization, and context-dependent language use (Dodigovic & Tovmasyan, 2021; Ghuftron & Rosyida, 2018; Nova, 2018).

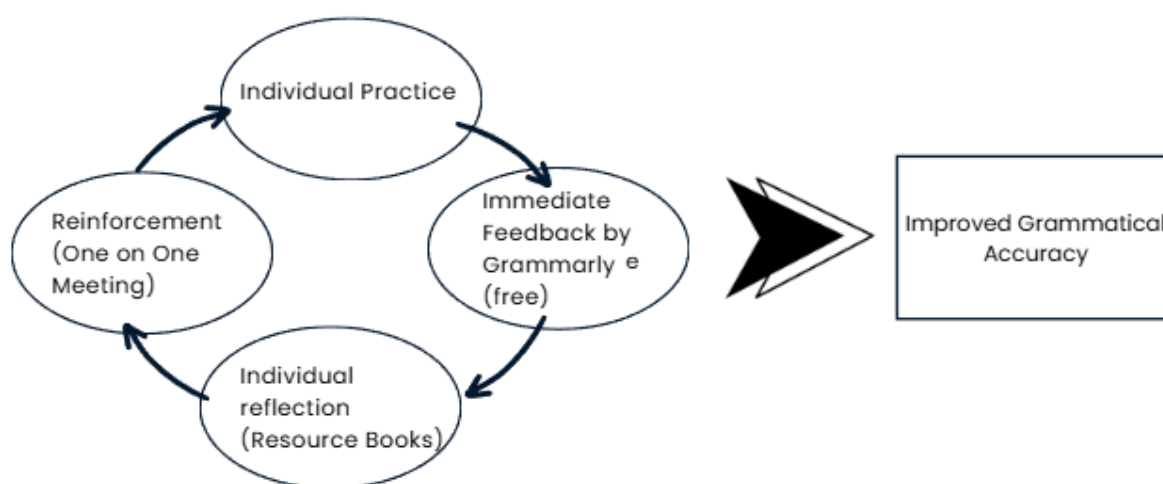
Overall, the literature suggests that Grammarly is most effective when used as a complement to, rather than a replacement for, teacher support. While automated feedback can enhance grammatical development and encourage autonomous learning, pedagogical guidance remains necessary to address complex writing issues and help learners interpret feedback appropriately.

While research demonstrates Grammarly's potential, significant gaps remain. Myanmar, classified as a Least Developed Country with substantial infrastructure constraints, remains unexplored in AWE literature despite its growing EFL learner population. Most existing studies focus on university settings in East Asia; evidence from small-scale tutorial contexts in South and Southeast Asia is virtually absent. Furthermore, few studies examine the free version of Grammarly, despite it being the only accessible option in resource-constrained environments where internet connectivity and electricity access are limited and where learners face both L1 transfer challenges, such as non-alphabetic script, no article system, different punctuation conventions, and severely limited access to personalized feedback. This study addresses this gap by integrating Grammarly with weekly meetings and self-reflection resources across A2, B1, and B2 learners, building on prior research emphasizing combined automated-human support (Ghuftron & Rosyida, 2018; Nova, 2018).

The Learner-Centered Grammar Enhancement Model introduced in this study integrates Grammarly with reinforcement through weekly one-on-one meetings and Grammar in Use self-reflection, drawing on Cognitive Learning Theory (Piaget, 1952) emphasizing active engagement and Sociocultural Theory (Vygotsky, 1978) highlighting scaffolding. This framework aims to improve accuracy while fostering autonomy (Figure 1).

Figure 1

Learner-Centered Grammar Enhancement Model



3. Methodology

3.1 Participants

This case study involved three Myanmar students at different CEFR proficiency levels (A2, B1, and B2) to examine Grammarly's impact on grammatical accuracy in writing. Three participants were deliberately selected to represent one learner at each target CEFR level (A2, B1, B2), consistent with the comparative case study design which prioritizes depth of analysis over breadth (Yin, 2018).

Their CEFR levels were determined based on their most recently completed Cambridge English qualification and the examination they were preparing for at the time of the study. Student A (15 years old) had successfully completed the Cambridge English Flyers examination and was preparing for the A2 Key for Schools (KET for Schools). Student B (16 years old) had passed the KET for Schools with Grade A and was preparing for the B1 Preliminary for Schools (PET for Schools). Student C (17 years old) had passed PET for Schools with Grade A and was preparing for the B2 First for Schools (FCE for Schools). These examinations are aligned with the Common European Framework of Reference for Languages (CEFR), corresponding to the A2, B1, and B2 levels respectively (Council of Europe, 2001; Cambridge, 2023). Participants were therefore categorized as A2, B1, and B2 level learners.

Although the extremely small sample size and absence of a comparison group limit the generalizability of the findings, the inclusion of learners at three distinct proficiency levels enabled in-depth longitudinal tracking of how students at different developmental stages responded to automated feedback (Yin, 2018).

Participants were selected through convenience sampling from the researcher's private tutorial classes. Their prior experience with Cambridge English qualifications provided sufficient language proficiency and familiarity with grammar instruction to engage meaningfully with the intervention. Participation was voluntary, with informed consent obtained from both participants and their parents or legal guardians. The study was scheduled during summer vacation to avoid disrupting regular school activities.

The researcher acted as both tutor and study facilitator, providing one-on-one guidance while ensuring that participation did not interfere with normal learning. This setup allowed the study to assess Grammarly's effectiveness across distinct proficiency levels while maintaining ethical standards of transparency, confidentiality, and voluntary participation. Ethical approval for this study was obtained through the standard review process of the researchers' institution.

3.2 Instruments

This study employed four primary instruments: Grammarly (free version), pre- and post-tests, Cambridge writing tasks, and level-appropriate Grammar in Use textbooks. Grammarly served as the automated writing evaluation tool for all writing submissions during the six-week intervention. It provided feedback in two domains: correctness (i.e., errors in the usage of verb tense, subject–verb agreement, articles, prepositions, pronouns, punctuation, and spelling) and clarity (i.e., conciseness and sentence flow). As the study focused on grammatical accuracy, only correctness errors were analyzed.

These were further classified into two subcategories based on their impact on meaning. Major errors refer to correctness errors that affect the accuracy of meaning, including faulty subject-verb agreement, confused words, incorrect verb forms, incorrect noun number, determiner use, wrong or missing prepositions, incorrect phrasing, pronoun use, and

conjunction use. Minor errors refer to correctness errors that do not significantly affect meaning, including improper formatting, misspelled words, incorrect punctuation, comma misuse within clauses, closing punctuation, misplaced words or phrases, and misuse of modifiers. This distinction is broadly consistent with the global-local error framework discussed in second language writing research (Ellis, 2009; Ferris, 2002)

Pre- and post-tests each comprised two components: a picture-based story writing task modeled after Cambridge English exams (KET, PET, FCE for Schools), which required participants to write narratives based on a visual prompt, and the Oxford Solutions Placement Test (OSPT), a multiple-choice grammar assessment measuring knowledge of tense, agreement, and sentence structure. The grammar test was administered via Google Forms. Both tests were reviewed by three experts to ensure level-appropriate difficulty.

Weekly practice writing tasks were selected from CEFR-aligned Cambridge Trainer books (Cambridge University Press, 2019, 2020) and were separate from testing materials. Participants also used the Grammar in Use series (Murphy, 2015, 2019) corresponding to their proficiency levels to complete targeted grammar exercises.

3.3 Data Collection

The study began with the pre-test, including a picture-based story writing task and a grammar test (OSPT), to establish a baseline of participants' grammatical accuracy. Immediately following the pre-test, participants attended an orientation session. They received personalized instruction on using Grammarly, including how to submit writing tasks, interpret automated feedback on errors such as tense, articles, and subject-verb agreement, and revise their work. Participants were shown a demonstration of Grammarly's key features during the orientation session.

During the orientation, participants were also introduced to graded Grammar in Use books (Murphy, 2015; 2019) corresponding to their CEFR proficiency level. The researcher explained the structure of the books, demonstrated how exercises were organized by grammar topic, and guided learners on completing exercises relevant to errors highlighted by Grammarly. Participants were instructed to track recurring mistakes, reflect on their errors, and revisit topics requiring further practice to consolidate grammatical understanding.

During the six-week intervention, participants completed four writing tasks per week (24 total) drawn from CEFR-aligned Cambridge Trainer materials (Cambridge University Press, 2019, 2020). Specifically, tasks were modeled after the official writing test formats of the Cambridge English examinations corresponding to each participant's proficiency level: KET for Schools (Student A), PET for Schools (Student B), and FCE for Schools (Student C). Writing tasks were sourced from A2 Key for Schools Trainer 1 and 2 (Cambridge University Press, 2019), B1 Preliminary for Schools Trainer 1 (Cambridge University Press, 2019) and Trainer 2 (Cambridge University Press, 2020), and B2 First for Schools Trainer 1 (Cambridge University Press, 2019) and Trainer 2 (Cambridge University Press, 2020). Prompts varied in genre and topic to reduce practice effects. Time allocations reflected the official exam writing time for each level, as summarized in Table 1 below. Full sample writing tasks for each proficiency level are provided in Appendices C, D, and E.

Table 1*Summary of Weekly Writing Tasks by Proficiency Level*

Student	Task Type	Time Limit
Student A (A2)	Short email/note; picture-based story	20
Student B (B1)	Guided email; article or story	45
Student C (B2)	Essay; article, email, report, or review	80

After each submission, participants reviewed automated feedback, revised their texts, and completed related grammar exercises. Weekly 30-minute individual meetings with the researcher provided clarification and additional support. At the end of the intervention, participants completed a post-test parallel to the pre-test. All activities were conducted online, and confidentiality and data security were maintained.

3.4 Data Analysis

This exploratory case study (Stake, 1995; Yin, 2018) examined the effectiveness of automated writing evaluation across three CEFR proficiency levels. Given the small sample ($N = 3$), the study aimed for analytical rather than statistical generalization, focusing on individual learning trajectories.

To address the first research question, which examined the impact of Grammarly and reinforcement on grammatical accuracy over time, pre-test and post-test results were compared. Data were obtained from two sources: (1) the picture-based writing task and (2) the Oxford Solutions Placement Test (OSPT). For the writing task, grammatical errors identified through Grammarly reports were counted and compared between the pre-test and post-test. For the OSPT, participants' scores were compared to examine changes in grammatical knowledge following the six-week intervention.

To address the second research question, which examined how the effect of Grammarly and reinforcement differed across proficiency levels, weekly error data from the 24 writing tasks were analyzed. Grammarly feedback reports were used to identify and categorize errors into major errors, which affect the accuracy of meaning, and minor errors, which do not significantly affect meaning. Error frequencies and trajectories were tracked and compared across the A2, B1, and B2 learners to identify differences in the pattern and rate of grammatical improvement at each proficiency level.

In addition, a qualitative descriptive analysis explored learners' engagement with feedback, autonomy, and support needs, based on researcher observations from weekly meetings. While formal observation protocols were not employed, patterns in learner behavior and engagement were noted throughout the intervention period, highlighting differences in responsiveness to automated versus human guidance across proficiency levels. Data from multiple sources were triangulated to enhance credibility.

4. Results/Findings

4.1 Pre-test and Post-test Results

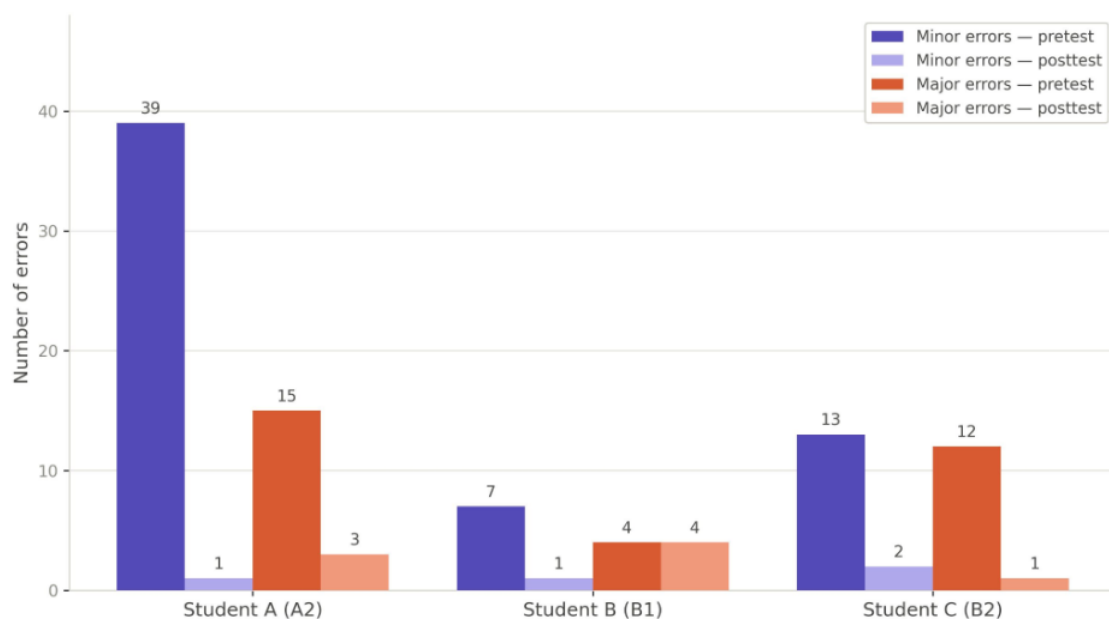
To investigate the effect of Grammarly and reinforcement strategies on grammatical accuracy, both the pre-test and post-test results were analyzed. Each test consisted of two sessions: a story-writing task and a multiple-choice grammar test. In the first session, correctness errors were classified into two subcategories as described in Section 3.2.

4.2 Writing Test Results

Figure 2 summarizes the number of correctness-related errors identified in the writing tasks.

Figure 2

Major and minor errors in pretest and posttest



In the pre-test, Student A recorded 39 minor errors and 15 major errors, Student B recorded 7 minor errors and 4 major errors, and Student C recorded 13 minor errors and 12 major errors. In the post-test, errors decreased across both categories for all three students. Student A reduced minor errors to 1 and major errors to 3, Student B reduced minor errors to 1 while major errors remained at 4, and Student C reduced minor errors to 2 and major errors to 1. These results suggest that while both error categories declined following the intervention, minor errors showed a more consistent reduction across all three proficiency levels.

However, comparisons across proficiency levels should be interpreted cautiously. As shown in Table 1, writing tasks differed across proficiency levels in terms of task type and time allocation, which may have influenced raw error counts. Given that higher-proficiency learners were expected to produce longer texts within their allocated time, raw error totals are not directly comparable across levels. Future studies may benefit from using normalized measures, such as error rates per 100 words, to facilitate more precise cross-level comparisons.

4.3 Grammar Test Results

Table 2 presents the OSPT scores.

Table 2

Test Scores in Second Session: Pre-test vs. Post-test

Student	Pre-test Score	Post-test Score
Student A (CEFR A2)	30	38
Student B (CEFR B1)	38	42
Student C (CEFR B2)	42	47

All participants showed modest score increases. Student A improved by eight points, Student B by four points, and Student C by five points. These gains are consistent with the reductions observed in the writing tasks and suggest parallel development in explicit grammatical knowledge.

4.4 Weekly Error Trends

To examine development over time, weekly Grammarly correctness reports were compiled across the 24 writing tasks. Table 3 summarizes the total number of errors per week.

Table 3

Weekly Grammatical Correctness Errors Across Six Weeks

Student	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Student A (A2)	69	47	35	25	35	24
Student B (B1)	38	35	31	14	17	11
Student C (B2)	60	58	49	49	43	40

All participants exhibited overall downward trends in error frequency, although the rate of change differed by proficiency level. Student A showed the steepest reduction, particularly during the early weeks, followed by some fluctuation. Student B demonstrated steady, gradual improvement. Student C displayed smaller, incremental reductions across the intervention. These patterns suggest differing developmental trajectories across proficiency levels, with larger observable changes among lower-level learners and more incremental refinement among higher-level learners.

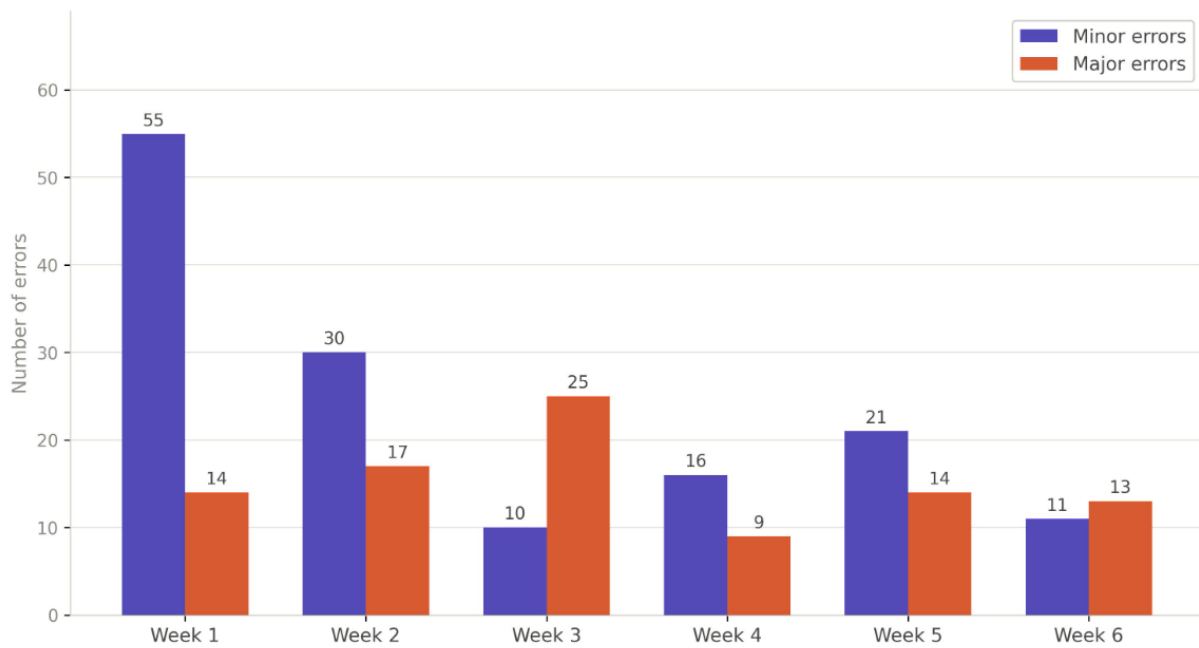
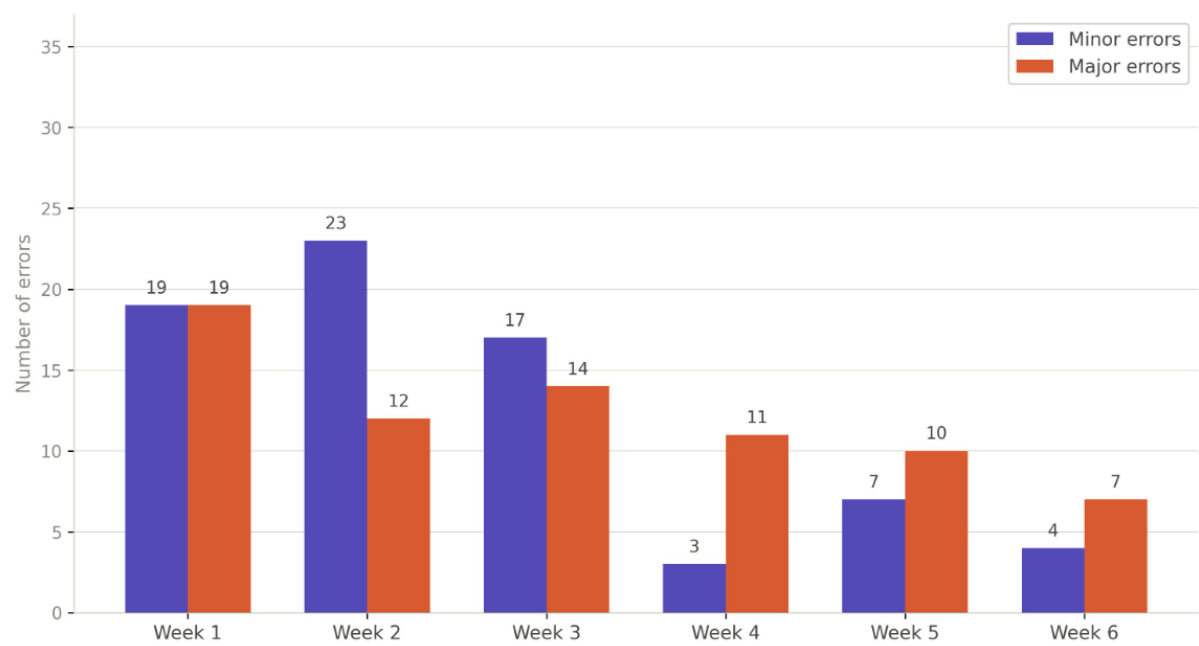
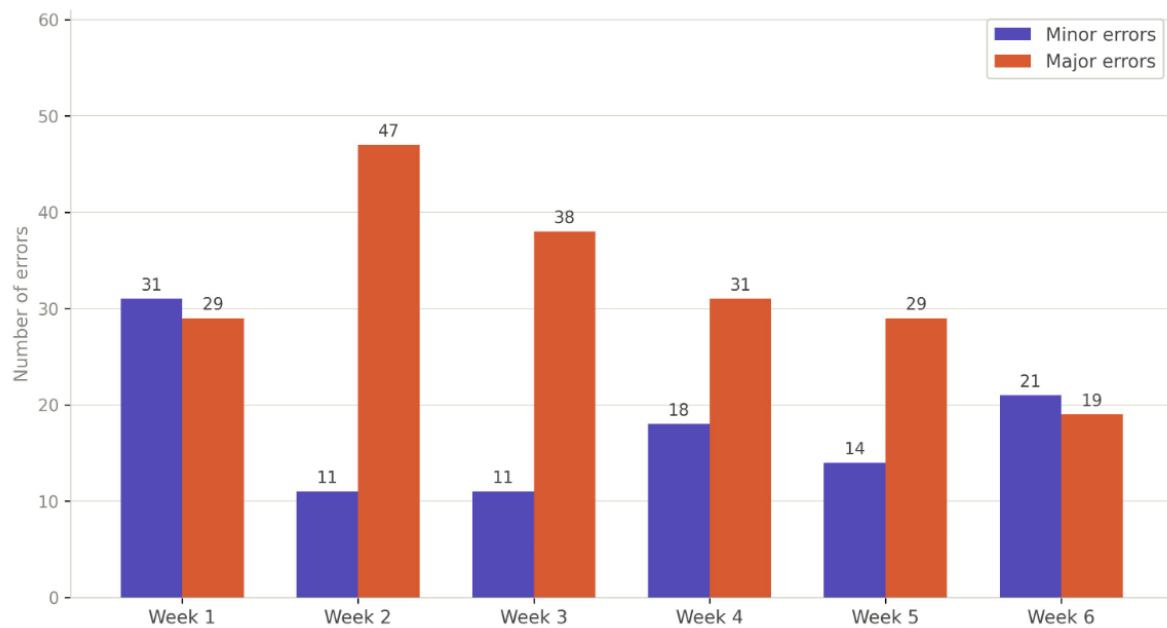
Figure 3*Major and minor errors of Student A made across six weeks***Figure 4***Major and minor errors of Student B made across six weeks*

Figure 5*Major and minor errors of Student C made across six weeks*

Figures 3, 4, and 5 further illustrate the weekly changes in major and minor errors for each participant. Overall, all three participants demonstrated a downward trend in error frequency, although the rate of improvement varied by proficiency level. Minor errors, including formatting, punctuation, and spelling, generally decreased more rapidly and consistently than major errors. In contrast, major errors, such as verb form, determiner, preposition, and pronoun errors, showed more gradual reductions. Student C exhibited the slowest overall improvement, with major errors remaining comparatively higher throughout the intervention period.

Taken together, quantitative data from the weekly error reports and pre- and post-test results indicate consistent improvement in grammatical accuracy across all three cases during the six-week period. The magnitude and pattern of change varied by proficiency level: the A2 learner showed larger absolute reductions in errors, the B1 learner demonstrated steady progress, and the B2 learner exhibited smaller but consistent refinements.

Observational notes from weekly meetings suggest that participants engaged with feedback and supplementary materials in different ways. A2 learners appeared to focus on immediate error correction, whereas B2 learners tended to use feedback for refinement and self-monitoring. These differences may reflect varying instructional needs across proficiency levels.

Given the exploratory case study design, small sample size, and absence of a control group, these findings should be interpreted as descriptive rather than causal. Improvements observed during the intervention cannot be attributed solely to Grammarly or reinforcement strategies; other factors, such as practice effects may also have contributed. Nonetheless, the convergence of multiple data sources provides evidence of positive developmental trends during the study period.

5. Discussion

This discussion addresses the two research questions in turn. Section 5.1 examines how Grammarly and reinforcement affected grammatical accuracy over time (RQ1), Section 5.2 explores how effects differed across proficiency levels (RQ2), and Section 5.3 presents implications including the effectiveness of the Learner-Centered Grammar Enhancement Model, persistent grammatical challenges for Myanmar learners, and recommendations for CEFR-aligned instruction.

The results showed that all three learners reduced their grammatical errors during the six-week intervention. However, the amount and pattern of improvement differed across proficiency levels. These differences suggest that learners at different stages may use automated feedback in different ways.

5.1 Effect of Grammarly and Reinforcement on Grammatical Accuracy Over Time

The findings indicate that the combined use of Grammarly's free version and reinforcement strategies contributed positively to grammatical accuracy among all three participants. In the writing tasks, total errors decreased across all participants following the six-week intervention: Student A from 54 to 4, Student B from 11 to 5, and Student C from 25 to 3. Reductions were observed across both major and minor error categories, suggesting that learners became more accurate in their written grammar over time.

OSPT scores further support this trend. All three participants showed improved performance in the post-test, with Student A gaining eight points (30 to 38), Student B four points (38 to 42), and Student C five points (42 to 47). The alignment between writing-based error reduction and improved grammar test performance suggests that learners were not only correcting errors mechanically but also developing deeper grammatical knowledge.

The role of reinforcement strategies should also be considered. Participants were not only exposed to automated feedback through Grammarly but also engaged in weekly one-on-one meetings and used grammar reference materials for self-study. Drawing on Vygotsky's (1978) Sociocultural Theory, the one-on-one meetings served as a form of scaffolding, helping learners bridge the gap between what they could achieve independently and what they could understand with teacher support. Therefore, the improvements observed cannot be attributed solely to Grammarly but rather to the integrated learning process facilitated by both technological and pedagogical support.

5.2 Differences in Effectiveness Across Proficiency Levels

The A2 learner showed the most substantial improvement in terms of overall error reduction. Total errors decreased from 69 in Week 1 to 24 in Week 6, with only 4 errors recorded in the post-test writing task. Major errors, including incorrect verb forms and subject-verb disagreement, showed the most dramatic reduction, as these frequent, rule-based errors are more easily identified and corrected through automated feedback. This corroborates findings from Dizon and Gayed (2021), who reported that Grammarly significantly improved grammatical accuracy among Japanese A1-A2 college students, particularly for rule-based errors. However, a temporary increase in errors was observed in Weeks 4 and 5 before stabilizing in Week 6, suggesting that error recognition alone was insufficient and that repeated practice was necessary for consolidation.

The B1 learner exhibited a more gradual and consistent pattern of improvement. Total errors decreased steadily from 38 in Week 1 to 11 in Week 6, with 5 errors identified in the post-test. Unlike the A2 learner, Student B demonstrated a wider range of error types across both major and minor categories, suggesting that B1 learners are transitioning from basic grammatical control to more flexible language use, requiring both immediate feedback and continued practice for noticeable improvement. This gradual pattern is consistent with Qassemzadeh and Soleimani's (2016) finding that Grammarly's benefits for intermediate learners emerged more strongly in long-term retention than immediate performance.

The B2 learner presented the slowest overall improvement, with total errors declining from 60 in Week 1 to 40 in Week 6, and 3 errors recorded in the post-test. Several major errors persisted throughout the intervention, including confused words, determiner use, and incorrect phrasing. This suggests a ceiling effect where AWE tools' rule-based algorithms effectively address A2 and B1 errors but cannot fully support the nuanced, context-sensitive accuracy required at B2 level, consistent with Nova's (2018) findings and Dodigovic and Tovmasyan's (2021) report of decreased accuracy for context-dependent errors at higher proficiency levels.

Overall, these differing trajectories suggest that lower-proficiency learners may experience more rapid initial gains from automated feedback, while higher-proficiency learners show more gradual refinement. Such patterns are consistent with skill acquisition theory, which posits that language learning requires repeated practice to move from declarative to procedural knowledge (DeKeyser, 2007). In line with this, Abdi Tabari et al.'s (2025) meta-analysis found that task repetition with spaced intervals improved grammatical accuracy particularly when combined with feedback, which may explain the consolidation patterns observed across levels. However, these interpretations should be viewed cautiously. With only three participants and no comparison group, the findings describe observed trends rather than causal effects.

Beyond error reduction patterns, observations from weekly one-on-one meetings revealed differentiated engagement with each intervention component across proficiency levels. While formal observation protocols were not employed, the researcher documented learner progress through weekly records of self-reflection activities and performance notes from one-on-one meetings. These records, combined with triangulation across multiple data sources, support the credibility of the patterns described (Yin, 2018; Stake, 1995). Consistent patterns emerged that illuminate how learners at different developmental stages interacted with automated feedback, self-reflection resources, and teacher guidance.

In terms of Grammarly feedback usage, Student A (A2) demonstrated limited ability to interpret automated corrections independently. Even after orientation, this learner required substantial teacher support to understand why suggested corrections were appropriate. For instance, when Grammarly flagged verb tense errors, Student A (A2) could identify that an error existed but could not articulate the underlying grammatical principle without explicit explanation. This pattern persisted throughout the intervention, with the learner consistently seeking confirmation before accepting automated suggestions. Grammarly functioned primarily as an error-detection tool, with comprehension dependent on subsequent teacher mediation. This dependence on teacher mediation for comprehension echoes concerns raised in previous AWE research. Dodigovic and Tovmasyan (2021) noted that Grammarly occasionally suggested incorrect corrections, highlighting that lower-proficiency learners lack the metalinguistic knowledge to evaluate feedback critically. Without teacher verification, A2 learners may accept erroneous suggestions or fail to understand why corrections are

appropriate, potentially limiting learning gains. This finding challenges implementations of AWE as standalone tools and supports Ghufon and Rosyida's (2018) argument that teacher input remains essential, particularly for lower-proficiency learners.

Regarding self-reflection with grammar resources, Student A required structured guidance for effective engagement with the Grammar in Use materials. Initially, when asked which grammar areas needed attention based on Grammarly feedback, this learner could not identify patterns without direct teacher intervention. The researcher had to explicitly point to recurring error types and direct Student A to specific Grammar in Use units. Even by Week 5, self-directed reflection remained limited, with the learner relying on teacher identification of priority areas for practice.

With respect to teacher-mediated reinforcement for Student A, meetings focused primarily on explicit grammar instruction: explaining why corrections were necessary, providing additional examples, and ensuring comprehension through guided questioning. Approximately two-thirds of session time involved direct instruction, with the teacher functioning as the primary source of grammatical knowledge.

Student B (B1) showed moderate independence in using Grammarly. This learner could interpret most corrections without extensive support but benefited from teacher prompting to identify recurring error patterns. When presented with multiple instances of article or preposition errors, Student B initially did not recognize these as systematic issues until guided to do so. Once patterns were highlighted, however, the learner could reflect on them more independently. This suggests that B1 learners possess emerging analytical capacity that requires scaffolding to activate fully. This need for scaffolding parallels findings from traditional corrective feedback research. Ahangari and Babapour (2015) found that self-correction with teacher guidance significantly improved writing accuracy among intermediate Iranian EFL learners, whereas peer correction without structured support did not. The present study extends these findings to AWE contexts, suggesting that Grammarly can serve the error-highlighting function previously performed by teachers, but B1 learners still require human scaffolding to move from error recognition to pattern identification and autonomous application.

Over the intervention period, Student B developed emerging reflective capacity. By mid-intervention, this learner began recognizing broad error categories (e.g., "I have problems with articles") when prompted, though unprompted reflection remained surface-level. Student B showed moderate independence in selecting relevant Grammar in Use exercises, typically consulting the table of contents and then seeking teacher confirmation that chosen units were appropriate. This pattern suggests B1 learners can engage in reflection but require initial scaffolding and periodic guidance.

In terms of teacher-mediated reinforcement, Student B required balanced support combining explanation and guided discovery. Rather than full explanations, meetings often employed prompting questions that helped the learner reach understanding independently. Session time was distributed more evenly between reviewing Grammarly feedback, facilitating pattern recognition, and planning targeted practice. The teacher's role shifted from instructor to facilitator, scaffolding rather than directly providing knowledge.

Student C (B2) demonstrated high autonomy from the outset. This learner not only interpreted Grammarly suggestions independently but occasionally questioned them when contextually inappropriate. During one meeting, Student C explained why a particular comma

suggestion did not align with the intended emphasis, demonstrating metalinguistic awareness beyond error correction. This critical evaluation capacity was similarly documented by Nova (2018), whose Indonesian postgraduate participants independently recognized when Grammarly's suggestions were contextually inappropriate and made deliberate decisions to reject them. This ability to evaluate automated feedback critically distinguishes B2 learners from A2 and B1, who in the present study accepted Grammarly suggestions with little questioning. The development of this evaluative stance may explain why previous research (Ghufron & Rosyida, 2018) found that students with stronger linguistic foundations benefited more from Grammarly. They possessed the metalinguistic resources to use it strategically rather than prescriptively. For this learner, Grammarly served as a proofreading tool rather than primary instruction, with independent critical evaluation of automated feedback.

Student C, exhibiting strong independent reflection throughout, was able to analyze error patterns without prompting, categorize mistakes by type, and independently select targeted practice activities. Rather than requiring direction, Student C brought specific questions about nuanced grammar points to meetings, indicating that self-reflection at this level serves to deepen existing knowledge rather than establish foundational understanding.

Student C used meetings primarily for metalinguistic discussion rather than grammar instruction. Sessions involved exploring why certain grammatical choices were more appropriate than others, discussing nuanced rules, and addressing questions that extended beyond basic correctness. The teacher functioned more as a discussion partner than instructor, with meetings serving to refine advanced knowledge rather than establish it.

These differentiated patterns confirm that intervention effectiveness depends not solely on tool quality but on the interaction between automated feedback, learner developmental readiness, and appropriate pedagogical support. What constitutes 'support' varies substantially: A2 learners require explicit instruction, B1 learners need guided discovery, and B2 learners benefit from facilitated reflection and discussion.

This differentiation challenges assumptions underlying many AWE implementations that automated feedback functions uniformly across proficiency levels. The present findings suggest that pedagogical effectiveness depends critically on proficiency-mediated interaction with feedback rather than feedback quality alone. For A2 learners, AWE without human mediation may reinforce surface correction without developing grammatical understanding (Ferris, 2010), while B2 learners can leverage the same feedback for autonomous learning. This proficiency gradient has implications for large-scale AWE deployment in heterogeneous classrooms, where standardized automated feedback may inadvertently advantage B2 learners while providing insufficient scaffolding for A2 learners.

These patterns suggest that the necessity of teacher-led reinforcement likely varies by proficiency level. For A2 learners, teacher involvement appears essential to provide explanations that automated feedback alone cannot offer. For B1 learners, teacher support remains valuable but might be reduced or delivered through alternative structured activities. For B2 learners, greater independence seems achievable. However, these interpretations should be viewed cautiously. Without experimental comparison of different component combinations, it remains unclear whether similar improvements would occur with only two components. The observed effectiveness may have resulted from the synergistic interaction of all three elements rather than any single component.

5.3. Implications

5.3.1. Effectiveness of the Learner-Centered Grammar Enhancement Model

The three-component model: Grammarly feedback, learner reflection, and teacher-led reinforcement, revealed differences in how learners at each proficiency level engaged with automated feedback during the intervention. A key question is whether learners would achieve similar improvements using only Grammarly and self-reflection without teacher-mediated reinforcement sessions.

Student A (A2 level) showed limited ability to use Grammarly independently. During the 30-minute reinforcement sessions, this learner frequently needed teacher explanations to understand why corrections were appropriate. For instance, when Grammarly flagged verb tense errors, Student A could accept the suggested change but struggled to explain the underlying grammar rule without support. Self-reflection at this level appeared to focus mainly on surface corrections rather than pattern recognition. The teacher spent considerable time providing explicit grammar explanations and asking guided questions. Based on these observations, improvement with only Grammarly and self-reflection seems unlikely for A2 learners. Without teacher mediation to explain grammatical concepts, the learner would likely have continued making surface-level corrections without developing deeper understanding of the concepts. Research confirms that explicit metalinguistic explanation is particularly beneficial for lower-proficiency learners who lack the linguistic foundation to induce grammatical rules from examples alone (Ellis, 2006).

Student B (B1 level) demonstrated moderate independence. This learner could interpret most Grammarly suggestions without extensive explanation but benefited from teacher prompting to identify recurring patterns. For example, when presented with multiple article errors, Student B needed guiding questions such as "What do these mistakes have in common?" to recognize the pattern. Once prompted, however, this learner could articulate explanations and attempt self-correction. The teacher's role shifted from direct instruction to facilitation, encouraging error categorization and providing targeted exercises. Whether B1 learners would improve with only two components is uncertain. While Student B showed emerging reflective capacity, the teacher's role in sustaining motivation and providing timely clarification likely contributed to the steady progress observed.

The importance of teacher presence for maintaining motivation is consistent with broader findings on technology-enhanced language learning. Koltovskaia (2020) found in a multiple case study that students' engagement with Grammarly's automated written corrective feedback varied significantly based on their motivation and self-regulation skills, with some students abandoning the tool without teacher encouragement. This suggests that while Grammarly provides feedback mechanically, the social and motivational elements of teacher interaction remain critical for sustained engagement, particularly for B1 learners who have not yet developed the autonomous learning strategies characteristic of B2 level students. Carefully designed reflection prompts or structured peer activities might partially substitute for teacher involvement, though this would require further investigation.

Student C (B2 level) showed the greatest autonomy. This learner independently analyzed Grammarly feedback and sometimes questioned suggestions when they seemed contextually inappropriate. During reinforcement sessions, Student C required minimal instruction and used the time mainly for metalinguistic discussion. The teacher functioned more as a discussion partner than an instructor. For B2 learners, improvement with only Grammarly

and self-reflection appears more plausible, as this learner possessed both the linguistic knowledge and metacognitive skills to use automated feedback productively. Metacognitive awareness, including the ability to plan, monitor, and evaluate learning processes, develops alongside linguistic proficiency and enables the learners to engage in strategic self-regulation (Wenden, 1998). However, teacher involvement may still enhance engagement depth and provide opportunities for linguistic inquiry that might not occur independently.

It seems that teacher-mediated reinforcement remains a necessary component since feedback provided during interaction helps learners notice gaps between their production and the target form (Ellis, 2009). Moreover, previous research (Hattie, 2009) has found that teacher feedback, clarity, and support have strong effects on student engagement and achievement.

5.3.2. Persistent Grammatical Challenges for Myanmar Students

Despite overall error reductions across all three learners, certain grammatical areas showed limited improvement or persisted throughout the intervention. Analysis of pre-test and post-test writing tasks, along with weekly error reports, reveals specific challenges within major and minor error categories that automated feedback and reinforcement appeared insufficient to fully address.

Major errors showed varying patterns across proficiency levels. While Student A (A2) showed dramatic reduction in major errors (15 to 3), and Student C (B2) also improved substantially (12 to 1), Student B (B1) showed no change in this category (4 errors in both pre- and post-test). For the A2 learner, major errors tended to involve fundamental rule violations such as incorrect verb forms and subject-verb disagreement: errors that are frequent, systematic, and therefore more responsive to automated feedback. For the B1 learner, however, major errors may have occurred less frequently but involved more complex structures where rule application is conditional or context-dependent. Observational notes indicated that learners sometimes struggled to generalize rules across contexts, suggesting that error recognition did not always translate to autonomous rule application.

Among major errors, certain types proved particularly resistant. Article errors appeared particularly resistant to correction. In Burmese, which is the mother tongue of Myanmar people, does not employ a grammatical article system, requiring learners to acquire an entirely new category without L1 reference points. While Grammarly consistently flagged article omissions and misuse, learners often struggled to understand why specific articles were required in particular contexts. Similarly, preposition errors persisted, especially in temporal and spatial expressions. These findings align with research on cross-linguistic influence, which suggests that grammatical features absent in the first language pose persistent acquisition challenges that extend beyond simple error correction (Jarvis & Pavlenko, 2008).

Minor errors also showed notable persistence despite being among the most consistently identified issues by Grammarly. While all three students reduced minor errors (Student A from 39 to 1, Student B from 7 to 1, and Student C from 13 to 2), weekly error reports revealed that punctuation issues recurred throughout the intervention. This persistence may be attributed to fundamental differences between Burmese and English writing systems. Burmese does not use punctuation in the same way as English; sentence boundaries and pauses are indicated differently, and marks such as commas, semicolons, and colons are employed less systematically.

Learners showed particular difficulty with comma placement in complex sentences, apostrophe use in possessives and contractions, and appropriate deployment of semicolons and colons. Even when Grammarly highlighted these errors, learners often could not articulate the underlying punctuation conventions. Research on punctuation acquisition suggests that these conventions, being arbitrary rather than semantically motivated, require explicit instruction and extensive practice (Hall, 2009).

These persistent challenges indicate that error types rooted in fundamental L1-L2 structural differences require frontloaded explicit instruction before learners can productively engage with automated feedback. Future interventions could benefit from providing targeted exercises and preliminary instruction on articles, prepositions, and punctuation conventions to help Myanmar learners overcome these language transfer obstacles. The findings suggest that while AWE tools effectively support high-frequency error correction, they may be less effective for addressing deeply embedded transfer errors without accompanying explicit instruction and extended practice opportunities.

5.3.3 Implications for CEFR-Aligned Instruction

For educators implementing CEFR frameworks, AWE tools should be deployed with proficiency-appropriate functions. At A2, maximize AWE exposure for high-frequency errors (verb tense, subject-verb agreement, punctuation) paired with explicit grammar instruction that explains why corrections are needed. At B1, balance automated feedback with structured reflection activities—error logs tracking Grammarly-identified patterns, combined with targeted exercises, help learners transition from error recognition to independent rule application. At B2, position AWE as a proofreading and noticing tool rather than primary instruction; benefit most from self-directed error analysis and metalinguistic reflection.

The persistence of punctuation errors despite automated feedback suggests CEFR descriptors may underestimate mechanical accuracy challenges for learners from non-alphabetic writing systems. Current A2-B2 grammatical accuracy descriptors focus primarily on morphosyntactic accuracy but treat punctuation as mechanical convention. For Myanmar learners, punctuation represents a substantive grammatical challenge requiring explicit pre-teaching before AWE feedback becomes interpretable. CEFR-aligned curricula in such contexts should frontload punctuation instruction earlier than typical European EFL progressions.

For test developers and curriculum designers, AWE-generated error profiles could supplement traditional CEFR placement tests, providing diagnostic information about error types and frequency aligned with grammatical accuracy descriptors. However, error quantity alone does not predict proficiency. Student C initially produced more errors than Student B but was higher-proficiency, suggesting error type and severity must also be considered in placement decisions.

5.3.4 Practical Classroom Applications and Scalability

Teachers should differentiate AWE use by proficiency level: A2 learner benefits from frequent AWE exposure on every writing task with teacher follow-up ensuring corrections translate to rule understanding; B1 learner should combine AWE with reflective activities, error journals, peer discussions of common mistakes, targeted grammar exercises to consolidate

emerging knowledge; B2 learner may use AWE selectively for final drafts, focusing more on self-directed error analysis than automated correction.

Pre-teaching L1-influenced error types before AWE introduction prevents frustration from feedback on unfamiliar conventions. Teachers working with learners from non-alphabetic L1s should frontload punctuation conventions (comma usage, apostrophe rules, sentence-final punctuation) through direct instruction and controlled practice before introducing automated tools.

While this study used intensive one-on-one meetings, more scalable alternatives include peer review sessions where learners discuss Grammarly feedback collaboratively, digital error logs tracking recurring patterns for teacher-led mini-lessons, asynchronous video feedback addressing common patterns across multiple students, or biweekly group workshops analyzing anonymized reports. Even in large classes (30+ students), brief check-ins combined with peer accountability can maintain reinforcement benefits without overwhelming teacher workload. The key is ensuring learners actively process feedback rather than passively accepting corrections.

Timeline expectations should reflect proficiency-based differences: rapid gains at A2 (2-3 weeks), steady progress at B1 (4-6 weeks), incremental refinement at B2 (6+ weeks). Teachers should avoid assumptions of uniform rapid improvement. Curriculum planners should allocate minimum 8–10-week units for grammar-focused AWE interventions at B1 and B2 levels to allow sufficient consolidation of time.

From a pedagogical perspective, the findings highlight AWE's potential to support CEFR-aligned instruction in resource-constrained EFL contexts by extending feedback opportunities beyond classroom time. For teachers and curriculum developers, integrating Grammarly into classroom practice may help address limited access to personalized feedback, while reinforcement activities ensure learners engage meaningfully with automated corrections rather than relying solely on them. Overall, AWE with structured reinforcement can serve as a flexible, learner-centered approach to grammar development, with its role understood as differentiated: remedial for A2 learners, consolidating for B1 learners, and refining for B2 learners.

6. Limitations and Recommendations

This study faced some contextual and methodological limitations. First, the research was conducted with a small sample size of only three learners at different CEFR levels. While the exploratory case-study design enabled detailed insights into individual progress and highlighted the differentiated impact of Grammarly and reinforcement, the findings cannot be generalized to broader EFL populations. Future studies should include larger and more diverse samples to better capture variability in how learners respond to automated feedback and reinforcement strategies.

Second, the study was conducted entirely online, which made it vulnerable to infrastructural challenges common in Myanmar as a Least Developed Country. During the data collection period, scheduled electricity outages meant power was not available continuously, and internet connectivity remained unstable throughout. These constraints required flexible scheduling and demonstrated the resilience necessary for implementing educational technology in resource-limited contexts. Although the participants demonstrated notable discipline and resilience by following the given timeline despite these disruptions, such challenges highlight

the fragility of online-only approaches in contexts with limited infrastructure. The successful completion of the intervention despite these infrastructure limitations actually strengthens the study's practical value, demonstrating that AWE integration is feasible even in challenging technological environments. If future research is also conducted in countries facing similar challenges, such as Myanmar, in-person or blended designs may be more suitable to minimize technical barriers and ensure smoother data collection.

Moreover, the study was conducted on a small scale with intensive researcher involvement, including one-on-one meetings and individualized reinforcement. While this approach was feasible and effective for a case study with three participants, it raises concerns about scalability. In larger-scale studies, individualized support would be time-consuming and difficult to sustain. Alternative reinforcement methods such as peer feedback, structured online platforms, or regular digital check-ins, may provide more scalable ways of maintaining personalized support.

Finally, while the participants in this study remained highly motivated and disciplined, it is uncertain whether the same level of commitment could be maintained in larger groups over longer periods. Future research should explore strategies to sustain engagement at scale, such as automated reminders, group accountability mechanisms, or blended learning approaches that combine automated feedback with structured instructor oversight.

By acknowledging these limitations, this case study indicates the importance of considering not only the pedagogical effectiveness of AWE tools like Grammarly, but also the practical realities of implementation in Myanmar's EFL context, where sample size, infrastructural challenges, and scalability all influence outcomes.

7. Conclusion

This exploratory case study provides the first empirical examination of automated writing evaluation in Myanmar's EFL context. By investigating how Grammarly's free version, combined with teacher-mediated reinforcement, impacts grammatical accuracy among three learners at A2, B1, and B2 levels, this study addresses a significant gap in AWE research from under-represented, low-resource contexts. All learners showed improvement in both writing tasks and grammar test scores after the six-week intervention. However, the amount and type of improvement differed by proficiency level.

For A2 learners, automated feedback helps correct frequent basic errors. For the B1 learner, feedback and practice worked together to strengthen developing grammar knowledge. For the B2 learner, Grammarly mainly supported proofreading and helped refine smaller, more subtle errors. These results suggest that automated feedback may serve different purposes depending on the learners' language level.

The findings also show that automated feedback alone may be insufficient. Reinforcement activities, such as reflection, grammar exercises, and guided support, helped learners understand their mistakes and apply corrections more consistently. Improvement was gradual and sometimes uneven, indicating that grammar development requires time and repeated practice. In addition, punctuation errors remained challenging for all Myanmar learners, suggesting that some problems related to first-language influence may need explicit instruction.

The study's contributions extend beyond Myanmar's borders. It demonstrates that accessible, free AWE tools can be pedagogically effective when embedded within teacher-

mediated instruction; a finding with direct relevance for resource-constrained EFL contexts globally, particularly in Least Developed Countries facing infrastructure limitations. Despite challenges such as scheduled electricity outages and unstable internet connectivity during data collection, the intervention proved feasible and effective, confirming that AWE integration does not require optimal technological infrastructure to yield meaningful learning outcomes. The fine-grained analysis across CEFR levels reveals that educational technology does not function uniformly; rather, its role and effectiveness are mediated by learners' proficiency stages. These insights inform technology integration decisions in similar small-scale, tutorial-based teaching environments throughout South and Southeast Asia where personalized instruction remains feasible but access to premium educational technology and stable infrastructure is limited.

Although the study involved only three participants, it provides useful insights into how AWE tools can be used in resource-limited EFL contexts. When combined with structured support, tools like Grammarly can extend feedback opportunities and encourage learner independence. Future studies with larger groups and more scalable designs are needed to confirm these results.

8. About the Author

Nan Ingyin Phyu is currently a MA student of the English as an International Language Program at Chulalongkorn University in Bangkok, Thailand. She obtained a Bachelor of Arts in Oriental Studies from University of Yangon, and a Diploma in English from Yangon University of Foreign Language. Her research interests are in language assessment and evaluation in EFL.

Tanyaporn Arya, Ph.D. is a faculty member of Chulalongkorn University Language Institute, where she teaches communication and language assessment courses. Her research interests are in language assessment and intercultural communication.

9. Acknowledgement

Our gratitude goes to all the participants for their dedication and perseverance. Moreover, we are grateful to everyone who contributed to this research, directly or indirectly, for their encouragement and support.

10. Declaration of AI Use

The authors declare that AI tools, specifically Claude Sonnet 4.6 and ChatGPT 5.5, were used for grammatical and spelling proofreading, reference formatting, and minor language refinement.

11. References

- Abdi Tabari, M., Zhuang, J., & Farahanynia, M. (2025). Task repetition and L2 written performance: A meta-analysis. *Journal of Second Language Writing*. <https://doi.org/10.1016/j.jslw.2025.101255>
- Abu Guba, M. N. A., Awad, A., & Abu Quba, A. (2024). Grammarly in Teaching Writing to EFL learners at low levels: How useful is it? *World Journal of English Language*, 14(3), 1. <https://doi.org/10.5430/wjel.v14n3p1>
- Ahangari, S., & Babapour, M. (2015). The Effect of Self-Correction and Peer- Correction on EFL Learners' Writing. *Modern Journal of Language Teaching Methods*, 5(2), 465–473.
- Al-Alami, S. E. (2024). EFL learners' attitudes towards utilizing ChatGPT for acquiring writing skills in higher education: A case study of computing students. *Journal of Language Teaching and Research*, 15(4), 1029–1038. <https://doi.org/10.17507/jltr.1504.01>
- Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.
- Cambridge University Press. (2019). A2 key for Schools trainer 1.
- Cambridge University Press. (2019). A2 key for Schools trainer 2.
- Cambridge University Press. (2019). B1 preliminary for Schools trainer 1.
- Cambridge University Press. (2020). B1 preliminary for Schools trainer 2.
- Cambridge University Press. (2019). B2 first for Schools trainer 1.
- Cambridge University Press. (2020). B2 first for Schools trainer 2.
- Cambridge English. (2023). The Cambridge English Scale explained. <https://www.cambridgeenglish.org/images/210434-converting-practice-test-scores-to-cambridge-english-scale-scores.pdf>
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. Heinle & Heinle.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97-113). Routledge.
- DeKeyser, R. M. (2018). Task repetition for language learning: A perspective from skill acquisition theory. In M. Bygate (Ed.), *Learning language through task repetition*.
- Dembsey, J. M. (2017). Closing the Grammarly Gaps: A Study of Claims and Feedback from an Online Grammar Program. *Deleted Journal*, 36(1). <https://doi.org/10.7771/2832-9414.1815>
- Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *JALT CALL Journal*, 17(2), 74–92.
- Dodigovic, M., & Tovmasyan, A. (2021). Automated writing evaluation: The accuracy of Grammarly's feedback on form. *International Journal of TESOL Studies*, 3(2), 71–88.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83-107.
- Ellis, R. (2009). Corrective Feedback and teacher development. *L2 Journal: An Open Access Refereed Journal for World Language Educators*, 1(1).

- Fan, N., & Ma, Y. (2022). The Effects of Automated Writing Evaluation (AWE) feedback on Students' English writing quality: A Systematic literature review. *Language Teaching Research Quarterly*, 28, 53–73. <https://doi.org/10.32038/ltrq.2022.28.03>
- Ferris, D. (2002). *Treatment of error in second language student writing*. University of Michigan Press.
- Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32(2), 181-201.
- Fitria, T.N., 2021. Artificial intelligence (AI) in education: using AI tools for teaching and learning process. Proceeding Seminar Nasional & Call for Paper STIE AAS 4 (1), 134–147. <https://prosiding.stie-aas.ac.id/index.php/prosenas/article/view/106/>
- Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a Foreign Language (EFL) writing. *Lingua Cultura*, 12(4), 395–403.
- Grammarly. (2020). The Grammarly Editor: New look, more suggestions, better writing. <https://www.grammarly.com/blog/product/better-writing-with-grammarly/>
- Grammarly. (2024a). AI-powered paraphrasing tool. <https://www.grammarly.com/paraphrasing-tool>
- Grammarly. (2024b). Releases. <https://www.grammarly.com/releases>
- Hall, N. (2009). Punctuation and grammar. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 271-283). SAGE Publications.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Inpanich, P., Chaya, P., Wongsittikan, N., & D. Evans, M. (2023). A Comparison of Grammatical Errors Made by Thai EFL Undergraduate Students in Different Types of Essays. *Manusat Paritat: Journal of Humanities*, 45(2), 165–191.
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Miranty, D., Widiati, U., Cahyono, B. Y., & Tengku Sharif, T. I. S. (2023). Automated writing evaluation tools for Indonesian undergraduate English as a foreign language students' writing. *International Journal of Evaluation and Research in Education (IJERE)*, 12(3), 1705–1715. <https://doi.org/10.11591/ijere.v12i3.24958>
- Murphy, R. (2015). *Essential grammar in use* (4th ed.). Cambridge University Press.
- Murphy, R. (2019). *English grammar in use* (5th ed.). Cambridge University Press.
- Ng, V., Ling, Y. L., & Kho, M. G. W. (2020). *Error analysis of the written English essays of diploma students in Kuching, Sarawak*. Paper presented at the National Research Innovation Conference (NRICon 2020), Kuching, Sarawak, Malaysia.
- Nguyen, T. H. (2020). Common grammatical errors in English writing: A case study with second-year students of information technology at HAUI. *Can Tho University Journal of Science*, 12(1), 37–44. <https://doi.org/10.22144/ctu.jen.2020.005>
- Nova, M. (2018). Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience. *Premise: Journal of English Education and Applied Linguistics*, 7(1), 80–96.

- Nurmayanti, N., & Suryadi. (2023). The effectiveness of using Quillbot in improving writing for students of the English education study program. *Jurnal Teknologi Pendidikan: Jurnal Penelitian dan Pengembangan Pembelajaran*, 8(1), 32–40. <https://e-journal.undikma.ac.id/index.php/jtp/index>
- Petchprasert, A. (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 1–16. <https://doi.org/10.1186/s40862-020-00107-w>
- Piaget, J. (1952). *The origins of intelligence in children*. W. W. Norton.
- Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies*, 6(9), 1884–1894. <https://doi.org/10.17507/tpls.0609.23>
- Ridha, N. (2012). The effect of EFL learners' mother tongue on their writings in English: An error analysis study. *Journal of the College of Arts, University of Basrah*, 60, 22–45.
- Sawalmeh, M.H.M. (2013). Error analysis of written English essays: The case of students of the Preparatory Year Program in Saudi Arabia. *English for Specific Purposes World*, 14(40), 1-17
- Savage, K. L., Bitterlin, G., & Price, D. (2010). *Grammar matters: Teaching Grammar in Adult ESL Programs*. Cambridge University Press.
- Singh, C. K. S., Singh, A. K. J., Razak, N. Q. A., & Ravinthar, T. (2017). Grammar errors made by ESL tertiary students in writing. *English Language Teaching*, 10(5), 16–27.
- Singh, C. K. S., Singh, A. K. J., Razak, N. Q. A., & Ravinthar, T. (2017). Grammar errors made by ESL tertiary students in writing. *English Language Teaching*, 10(5), 16–27.
- Solfiyatuzzahro, S., Santihastuti, A., & Erfan, E. (2019). Grammatical Accuracy using Teacher's Written Corrective Feedback. *Eralingua Jurnal Pendidikan Bahasa Asing Dan Sastra*, 3(2). <https://doi.org/10.26858/eralingua.v3i2.9051>
- Stake, R. E. (1995). *The art of case study research*. Sage Publications.
- Thi, N. K., Vo, D. V., & Nikolov, M. (2023). Investigating syntactic complexity and language-related error patterns in EFL students' writing: Corpus-based and epistemic network analyses. *Language Learning in Higher Education*, 13(1), 127–151. <https://doi.org/10.1515/cereles-2023-2009>
- Thornbury, S. (1999). *How to teach grammar*. Pearson Education.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wenden, A. (1998). Metacognitive knowledge and language learning. *Applied Linguistics*, 19(4), 515-537.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.
- Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

11. Appendix

Appendix A. Sample: Pre-Test and Post Test First Session (Writing Task)

Instructions: You will be given three pictures that together tell a story. Your task is to write a story based on these pictures. Focus on conveying your ideas clearly and ensuring your writing is grammatically accurate and coherent. There is no specific word count requirement; however, please continue writing until the time is over. Be creative in your writing and make sure to write to the best of your ability.

The allotted time for this task is 30 minutes. Please use this time to produce as much written material as you can.



Appendix B. Pre-Test and Post Test Second Session (Grammar Test)

<https://forms.gle/7L5zwj31EufTXHhF9>

Appendix C. Sample: Weekly Writing Tasks for A2 level student

You want to invite your friend Helen to the cinema next week.

Write an email to Helen.

In your email:

- Invite Helen to the cinema.
- Ask Helen what time and day suits her.
- Explain where the cinema is.

Write **25 words** or more.

Look at the pictures and write a story



Appendix D. Sample: Weekly Writing Tasks for B1 level student

Read this email from your English-speaking friend Sarah and *the notes* you have made.
Write your email to Sarah using all *the notes*.

 New Message

 From: Sarah

 Subject: Birthday party!

Hi,

Can you come to my birthday party downtown on Saturday evening at 7 pm? 🍕 **Yes, I can.**

We're trying to decide if it's better to have dinner at an Italian bistro or to go to a Chinese restaurant called the Golden Duck. What do you think is the best idea? What do you prefer? 🍕 **Say which I prefer.**

Also, we're going skiing in the mountains on Sunday morning. Would you like to join us? 🍕 **Yes, I would.**

Let me know if you have any questions. 🍕 **Ask Sarah...**

See you soon,

Sarah

Articles wanted!**TRAVEL**

Which is your favourite place to visit? Would you recommend others to go there? Why?

Write an article answering these questions, and we will publish it in the library magazine!

Appendix E. Sample: Weekly Writing Tasks for B2 level student



Some people believe that modern technology is creating a less sociable world. Do you agree?
Notes Write about:
1. Online communication 2. Impact on education 3. (your own idea)

You see this announcement in your school's student newsletter.

Movie Reviews Wanted

Have you watched a film recently that was particularly thought-provoking or inspiring?

Write us a review of the movie, describing the plot and what aspects made it thought-provoking or inspiring. Discuss the performances, direction, and whether the film succeeded in conveying its message.

Tell us if you think this film is worth watching and why.

The most compelling reviews will be published in our next student newsletter.

Write your **review**.

Appendix F. Sample: Correctness Errors Identified by Grammarly in Student Writing

Writing Issues

54	Correctness	
2	Misspelled words	
11	Incorrect punctuation	
19	Improper formatting	
11	Incorrect verb forms	
2	Pronoun use	
1	Comma misuse within clauses	
2	Incorrect phrasing	
2	Wrong or missing prepositions	
1	Determiner use (a/an/the/this, etc.)	
1	Conjunction use	
1	Faulty subject-verb agreement	
1	Incorrect noun number	

Appendix G. Sample: Student Writing Annotated for Correctness Errors by Grammarly

One morning, Emily, who is a 14 years old student woke up because her mom was waking her. She was annoyed because she slept late at night and she can't get up. She brushed her teeth, showered and dressed up slowly. When she was about to go downstairs to have breakfast, she looked at her clock and realized she was late for school. She quickly rushed down to the dining table and ate her breakfast really quick. Then, she ran to her school as fast as she can but that day was really strange because she wasn't tired at all. She thought it was a very special day for her. Well, it was. When she got to her school, she thought she was the only one who is late to school and was afraid to get scolded by the teacher. But, when she arrived to her school she saw a new girl who is also late to school. She knew that she is a new girl because she has never seen her before. She said Hi to her and tell new girl was very friendly that she replied with a nice Hi. They became friends and went to the classroom together. Strangely, their English teacher hadn't come yet so it's their free day. They were very happy until the principal came into their classroom. He said that they have a running race. Most of the students were disappointed but Emily and the new girl was very excited. Then, they went to the sports field and started racing. Three girls on top three winners and on the top three girl, Emily was in it. She was very nervous but she was not tired. She knew that the next two girls are tired so she plans to run very fast. Well, as we all were thinking she won! She was very happy and she knew it was a very special day.