ISSN: 2288-7709
JEMM website: https://accesson.kr/jemm
doi: http://dx.doi.org/10.20482/jemm.2025.13.3.75

# Data Analytics in Language Education: Assessing the Effectiveness of an AI Learning Platform*

**Ryan HATCHER[1], Myeong Hee SHIN[2]**

**Received: June 17, 2025. Revised: June 22, 2025. Accepted: June 30, 2025.**

## Abstract

**Purpose:** This study investigates the effectiveness of an artificial intelligence (AI)-based language learning platform in enhancing English proficiency among undergraduate students. **Research design, data and methodology:** A quasi-experimental design was employed to compare the learning outcomes of students using the AI platform with those receiving traditional instruction. Quantitative data, including exam scores and platform-generated analytics, were analyzed to assess progress over a single academic semester. **Results:** Students who used the AI-powered platform consistently outperformed those in the control group, demonstrating significantly higher test scores and more stable progress throughout the semester. The findings suggest that integrating adaptive AI learning technologies into tertiary English education can lead to measurable improvements in student performance. The study underscores the pedagogical potential of data-driven instructional approaches and supports the incorporation of AI tools to enhance learner engagement and academic outcomes.

**Keywords :** AI in education, language learning, MagniLearn, Power BI, learner performance, data-driven

**JEL Classification Code**: I21, I23, C88, O33

## 1. Introduction

However, despite the growing popularity of AI-assisted tools, empirical research remains limited regarding their long-term effectiveness in higher education settings. While previous studies have noted improvements in engagement and learner outcomes, many have relied on short-term trials or lacked detailed analysis of learner interaction patterns. This leaves unanswered questions about the sustainability and scalability of AI integration in formal university curricula.

In particular, there is a need for research that goes beyond simple pre- and post-test comparisons to include actual learner usage data—such as time spent on exercises, error correction rates, and task completion behaviors—to understand how AI-supported systems influence learning outcomes. Furthermore, few studies have explored how different learner profiles interact with AI-based tools and what pedagogical strategies best support such platforms.

This study seeks to address these gaps by investigating the effectiveness of an AI-powered adaptive learning platform in improving English language proficiency among Korean university students. By analyzing learner interaction data across multiple proficiency levels and course types, the study examines how different usage patterns correlate with learning outcomes such as vocabulary acquisition, grammar accuracy, and self-reported learning confidence.

*This work was supported by 2022 Hannam University Research Fund.

1 First Author. Assistant Professor, General English, Hannam University, Republic of Korea. Email: rjhatcher@hnu.kr
2 Corresponding Author. Associate Professor, General English, Hannam University, Republic of Korea. Email: scindy@hnu.kr

Through a mixed-methods approach combining quantitative data (e.g., CEFR scores, TOEIC performance, platform usage logs) and qualitative feedback (e.g., learner reflections, interviews), this research offers a comprehensive view of how AI-assisted instruction compares to traditional English teaching in terms of both efficacy and learner engagement.

## 2. Literature Review

### 2.1. AI Integration in Language Education

The application of AI in language education has shifted from theoretical promise to pedagogical practice over the past decade, driven by improvements in algorithmic precision and real-time data processing. AI-enhanced platforms now provide customized instruction based on learners' real-time performance, promoting individualized pacing and scaffolded content delivery (Zawacki-Richter et al., 2019).

Multiple studies have documented the benefits of AI in improving language skills, particularly in grammar, vocabulary, and pronunciation (Li & Hegelheimer, 2013; Xie et al., 2019). Wu et al. (2021) conducted a systematic review of AI-based platforms and found that such systems support both adaptive sequencing and increased learner autonomy. By simulating aspects of one-on-one tutoring, these systems have the potential to narrow proficiency gaps and support differentiated instruction (Colpaert, 2020).

Despite this promise, many existing studies have focused on short-term outcomes or specific skills, lacking longitudinal evidence of impact across instructional contexts. Moreover, research comparing AI-supported instruction with traditional classroom instruction in authentic settings remains limited (Holmes et al., 2022). This study builds on calls to expand empirical evidence through the use of platform analytics and mixed-method approaches.

### 2.2. NLP-Based Assessment Tools

Natural Language Processing (NLP) techniques have become increasingly central to the development of automated assessment systems in language education. Among these, semantic similarity measures such as Word Mover's Distance (WMD) have gained prominence for their ability to assess content-level meaning beyond surface lexical matches. Originally introduced by Kusner et al. (2015), WMD calculates the minimal cumulative distance required to transform one text into another by comparing word embeddings. Tashu and Horváth (2018) demonstrated that WMD outperformed traditional approaches such as TF-IDF and bag-of-words models in evaluating student essays, particularly by capturing semantic relationships in learner writing.

However, WMD presents challenges, especially in terms of computational efficiency for long-form or batch essay processing. Sato et al. (2022) proposed optimization techniques, including normalization and pruning, to improve scalability while preserving interpretive depth. More recent developments incorporate syntactic cues—such as syntax-aware WMD—and neural attention models (e.g., Fused Gromov-Wasserstein distance), further refining alignment accuracy at the sentence level and enhancing diagnostic feedback in formative assessment contexts.

Beyond WMD, several other NLP-based scoring methods have gained traction. Metrics such as BERT Score, ROUGE, BLEU, and GPTScore leverage contextualized embeddings and transformer models to evaluate coherence, fluency, and lexical diversity. BERT Score, for example, matches predicted tokens to reference answers using semantic vectors, enabling more nuanced analysis. GPT-based models (e.g., GPT-3, GPT-4) have shown strong agreement with human raters in standardized assessments like IELTS and TOEFL, although questions remain about transparency, fairness, and explain ability. These models are now being examined not only for their scoring accuracy but also for their diagnostic capacities—identifying grammar issues, coherence gaps, or lexical repetition.

In practice, large-scale language assessments have already adopted these tools. ETS (TOEFL) and the Duolingo English Test (DET) employ transformer-based NLP algorithms to assess both written and spoken responses. DET, for instance, delivers AI-generated results in under five minutes, supported by deep learning models that analyze semantic and structural features. However, because many of these commercial systems are proprietary, their inner workings remain opaque, highlighting the need for independent academic validation.

Despite these advancements, relatively few studies have directly compared NLP-based scoring models across diverse learner populations or within classroom-based writing contexts where genre, register, and proficiency vary widely. This study thus contributes to the field by examining the implementation of the MagniLearn platform through the dual lens of scoring accuracy and pedagogical interpretability, situating its analysis within the broader landscape of AI-mediated formative assessment.

### 2.3. Automated Speech Recognition (ASR) in Language Learning

Automated Speech Recognition (ASR) technology has gained considerable attention as a transformative tool for improving oral language skills, particularly in pronunciation and speaking fluency. ASR enables learners to receive instant, objective feedback on their speech, often highlighting errors in stress, intonation, and articulation in real-time (McCrocklin, 2019).

Commercial applications such as ELSA Speak, Google Read Along, and Microsoft's Azure-based ASR have made these tools increasingly accessible in both formal and informal learning settings. These platforms use speech models trained on large corpora to detect deviations from native-like pronunciation and offer corrective guidance, allowing learners to repeat and refine their output autonomously.

Recent studies indicate that ASR-based systems enhance pronunciation accuracy and oral fluency, especially for intermediate-level learners (Trinh et al., 2022; Lee, 2020). However, their effectiveness varies depending on task type and learner profile. For example, while controlled speech tasks (e.g., sentence repetition or reading aloud) yield high recognition accuracy, spontaneous speech tends to result in higher error rates, especially among learners with strong accents or limited prosodic control (Wang & Chen, 2021).

Moreover, ASR tools provide a nonjudgmental, low-anxiety practice space, which can help learners build confidence before participating in interactive classroom speaking activities. Nevertheless, the pedagogical impact of ASR depends heavily on how instructors integrate it into instruction, including the design of tasks, the alignment with curricular goals, and the scaffolding of feedback interpretation.

In sum, while ASR offers promising support for speaking development, further empirical studies are needed to assess its impact across diverse learning contexts and its interaction with other instructional modalities.

## 2.4. Learner Analytics and Behavioral Indicators

The emergence of learner analytics has revolutionized how educators and researchers evaluate student engagement and performance in digital learning environments. AI-enhanced platforms now generate detailed logs of learner behavior, including time-on-task, error patterns, repetition frequency, completion rates, and navigation pathways (Ifenthaler & Yau, 2020; Kovanović et al., 2017).

Such behavioral indicators offer a multidimensional view of learning that goes beyond test scores, revealing how learners interact with content and how motivation, persistence, and self-regulation play out over time. For example, learners who revisit challenging exercises multiple times or who gradually reduce error rates may demonstrate a deeper engagement with the learning process, even if their initial performance is modest (Pardo et al., 2019).

In recent years, studies have begun linking these behavioral indicators to affective and cognitive variables, such as self-efficacy, perceived usefulness, and emotional engagement (Schneider et al., 2022). Theories such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) have been used to explain how learners' perceptions of ease of use, usefulness, and feedback immediacy influence their continued use of AI platforms (Teo, 2011; Venkatesh et al., 2003).

However, there is a growing consensus that raw usage data must be contextualized within pedagogical design. Simply logging completion rates or login frequency is insufficient unless tied to learning goals, task difficulty, and learner characteristics. This study addresses this gap by analyzing platform log data alongside CEFR outcomes and learner interviews, enabling a more grounded interpretation of what behavioral engagement actually means in terms of learning efficacy.

## 2.5. Synthesis and Research Gap

The reviewed literature collectively underscores the growing promise of AI technologies in supporting language acquisition, yet several critical gaps remain that this study seeks to address.

First, many studies emphasize system capabilities without sufficiently investigating the learner experience or instructional integration. AI systems such as ASR and NLP tools are often evaluated in isolation, without examining how learners interact with them in sustained ways over the course of a full semester or academic year (Holmes et al., 2022).

Second, few empirical studies triangulate platform usage data with performance metrics and affective variables, such as self-efficacy or learner satisfaction. As AI tools increasingly influence pedagogical decisions, it becomes vital to understand not only what students achieve but how they achieve it.

Third, most comparative studies are based in Western or test-preparation-focused contexts, leaving gaps in our understanding of how these tools perform in regular academic courses across diverse cultural settings. In East Asian contexts, where classroom practices, teacher authority, and learner expectations differ, AI adoption patterns and learner responses may also diverge (Zhang & Hyland, 2021).

This study contributes to filling these gaps by conducting a data-rich, mixed-methods investigation of an AI-assisted English learning platform used across multiple course types in a Korean university. By analyzing learner behavior, academic performance, and learner reflections, the study aims to provide a holistic view of how AI tools are adopted, experienced, and evaluated in formal instruction.

### 2.6. AI Tools and Learner Self-Efficacy in Language Learning

Learner self-efficacy—defined as a student's belief in their ability to succeed in specific learning tasks—plays a critical role in language acquisition, particularly in technology-mediated environments (Bandura, 1997). As AI-powered tools increasingly support instruction, understanding how they impact self-efficacy has become a central concern in both applied linguistics and educational technology research (Huang et al., 2019; Zhang & Hyland, 2021).

AI tools can support learner confidence by providing immediate, individualized feedback, enabling low-stakes practice, and visualizing progress through analytics dashboards. These features have the potential to foster a sense of competence, control, and progress, all of which are positively correlated with higher self-efficacy (Tschannen-Moran & Hoy, 2001).

For instance, NLP-based platforms that allow repeated revision cycles, such as Grammarly or Write & Improve, have been shown to reduce writing anxiety and improve learners' confidence in self-editing (Wilson & Roscoe, 2020). Likewise, ASR tools that simulate pronunciation coaching can give learners a private, judgment-free environment to rehearse speaking skills, thereby reducing anxiety associated with oral production in public settings (McCrocklin, 2019).

Moreover, adaptive learning platforms that scaffold tasks based on proficiency data can promote a growth mindset by showing learners that improvement is attainable through effort and strategic practice. These experiences align with mastery experiences—one of the strongest sources of self-efficacy according to Bandura (1997).

However, not all learners interpret automated feedback positively. Studies have found that when feedback is vague, overly frequent, or linguistically complex, it can cause confusion or discouragement, particularly among lower-proficiency learners (Ranalli et al., 2022). In such cases, AI tools may inadvertently reinforce feelings of inadequacy or failure.

In addition, students' trust in the AI system, their understanding of its limitations, and their beliefs about the role of technology in education significantly mediate its impact on self-efficacy (Ifenthaler & Yau, 2020). Learners who perceive AI tools as accurate, helpful, and fair are more likely to experience increases in confidence and motivation (Huang et al., 2019).

Cultural attitudes toward error correction and technology also play a role. In collectivist and exam-oriented contexts such as South Korea, learners may place more trust in human instructors or may hesitate to rely entirely on automated systems (Park & Son, 2020). Therefore, self-efficacy gains from AI use must be understood in context, with attention to learners' prior experiences, cultural expectations, and digital literacy levels.

This study contributes to this discussion by examining how Korean university students experience AI-based instruction in relation to their perceived language competence. Through survey data and reflective interviews, we explore how learners interpret the usefulness and accuracy of feedback, how they respond emotionally to performance tracking, and whether these interactions influence their motivation to persist and improve.

Findings from this research aim to deepen our understanding of how AI tools not only affect learning outcomes but also shape learner identity, confidence, and long-term attitudes toward language learning.

## 3. Research Methods and Materials

### 3.1. Research Design

This study employed a mixed-methods research design to investigate the pedagogical effectiveness of the AI-powered MagniLearn platform in enhancing English language proficiency and learner engagement among undergraduate students enrolled in university-level English courses. The design integrated quantitative and qualitative data sources to provide a multidimensional evaluation of learning outcomes, usage behavior, and learner perceptions.

A quasi-experimental model was implemented, comprising a treatment group (n = 12) and a control group (n = 8). The treatment group used the MagniLearn platform in conjunction with regular classroom instruction, while the control group received conventional instruction without AI-enhanced support. Both groups were enrolled in similar English courses (general English and English for Specific Purposes), taught under the same institutional curriculum framework to ensure instructional consistency.

The intervention lasted one academic semester (15 weeks), during which students in the treatment group engaged with the MagniLearn platform for approximately 30–45 minutes per week. The platform employed natural language processing (NLP) and adaptive algorithms to deliver individualized tasks in grammar, vocabulary, listening, and pronunciation. Based on learners' responses and error patterns, MagniLearn adjusted task difficulty, sequencing, and feedback in real time, thereby constructing personalized learning trajectories.

Data collection focused on both outcome-based and process-based indicators. Academic outcomes included CEFR level gains, MagniLearn internal scores, and results from standardized tests such as TOEIC, TOEFL, and IELTS where available. Process-oriented data captured behavioral metrics such as time spent on the platform, number and type of completed exercises, success rates by grammatical category, and error correction frequency.

Data were processed and analyzed using Power BI, a business intelligence platform that enabled automated Extract, Transform, Load (ETL) procedures and dynamic data visualization. Visual analytics, including clustered bar charts, stacked graphs, and doughnut charts, were employed to compare performance across groups, track learner progress over time, and identify engagement trends. The platform's automation capabilities also supported real-time monitoring and reduced the risk of human error in data processing.

To complement the quantitative findings, qualitative data were collected from learner reflections and platform-generated feedback logs. These data were analyzed thematically to provide insight into students' perceptions of the platform's effectiveness, usability, and emotional impact on language learning. The integration of these data sources enabled the triangulation of results and supported a more comprehensive interpretation of how AI-mediated instruction influences learner development in higher education settings.

## 3.2. Research Context and Participants

This study was conducted within the framework of English language instruction at A University, a private institution located in Daejeon, South Korea. The research was embedded in a required English curriculum spanning both general English and English for Specific Purposes (ESP) courses.

Participants consisted of a total of 140 undergraduate students enrolled in English courses during the spring semester. For the core quasi-experimental comparison, 20 students were selected—12 assigned to the treatment group using the MagniLearn platform, and 8 to the control group receiving traditional instruction without AI support. These groups were selected to ensure equivalent class sizes, course types, and instructional time.

In addition to the primary participants, a supplementary dataset comprising 197 students who engaged with MagniLearn during the same semester was used for broader learning analytics and engagement pattern analysis. These additional participants were not part of the controlled comparison but contributed valuable data for identifying usage trends and performance variation.

Students were drawn from a variety of academic departments and participated in different sections of the English curriculum. For each student, the following data were collected: academic department, class type (General English or ESP), total instructional hours, and English proficiency indicators including CEFR levels and scores from TOEIC, TOEFL, and IELTS (when available). This detailed background information allowed for contextualized interpretation of performance differentials and informed subgroup analysis in the results section.

## 3.3. Instructional Content and Tools

The core instructional intervention in this study was the MagniLearn platform, an AI-driven language learning system that leverages natural language processing (NLP) and adaptive algorithms to personalize learning sequences. The system is designed to deliver micro-skill-level practice in grammar, vocabulary, listening comprehension, and pronunciation by dynamically adjusting task difficulty and content sequencing based on individual learner responses and ongoing performance trends.

Students in the treatment group accessed MagniLearn for a minimum of 30 to 45 minutes per week as part of their coursework. Tasks were tailored in real time, offering learners repeated exposure to challenging structures, targeted feedback, and progression scaffolding. The learning content spanned a wide range of CEFR-aligned skills, including verb tense usage, preposition accuracy, collocation, and synonym discrimination.

Instructors were granted administrative access to a teacher dashboard powered by Microsoft Power BI, allowing them to monitor individual and group-level performance. Real-time insights included completion rates, score trajectories, error frequency by linguistic feature, and success rates on key grammar and vocabulary patterns. This dashboard supported formative intervention and longitudinal tracking of learning gains.

Metrics collected through the platform included: 1) Number of exercises completed per student. 2) Success rates per linguistic rule. 3) Time spent per task and session. 4) Grammar and vocabulary categories targeted. 5) Types of errors (e.g., lexical, syntactic, phonological).

These data served as the foundation for both performance-based comparisons and behavioral engagement analyses, which are reported in Section 4.

## 3.4. Data Collection

Data for this study were gathered over the course of a 15-week academic semester and derived from both institutional records and the MagniLearn platform. Three distinct datasets were collected to support multi-layered analysis of learner performance, engagement, and instructional context. All data were anonymized and coded using randomized identifiers in accordance with A University's institutional review board (IRB) protocols.

Dataset 1: Student Proficiency Profiles

This dataset included pre- and post-intervention English proficiency indicators for each student. Variables consisted of: 1) CEFR levels, 2) Internal MagniLearn assessment scores, 3) Standardized English proficiency test scores (TOEIC, TOEFL, IELTS, when available)

These metrics served as the primary indicators of linguistic development over the intervention period and were used to evaluate learning outcomes across both treatment and control groups.

Dataset 2: Group and Class Assignment

This dataset documented the contextual and demographic information required to distinguish experimental conditions. Variables included: 1) Group status (treatment or control), 2) Course type (General English or English for Specific Purposes), 3) Instructor identification code, 4) Academic department affiliation. This information allowed for the stratification of participants and ensured comparability across groups in terms of instructional context and curricular exposure.

Dataset 3: Interaction and Engagement Logs

Exported directly from the MagniLearn platform, this dataset provided fine-grained, real-time behavioral data capturing learners' interaction with the AI system. Key variables included: 1) Total time spent on the platform, 2) Number of platform access sessions, 3) Task completion rates by activity type, 4) Success and failure rates by linguistic rule (e.g., past tense, modal verbs, and prepositions), 5) Frequency of repeated errors per learner. These data enabled the analysis of individual and group-level engagement patterns, rule-specific learning trajectories, and behavior-performance correlations. They were essential for triangulating learning outcomes with usage intensity and identifying points of instructional effectiveness or difficulty.

Together, these three datasets supported both comparative outcome analysis and exploratory behavioral modeling. Their integration in the data analysis process allowed for multi-dimensional interpretation of how AI-mediated instruction influenced learner proficiency, engagement, and instructional differentiation.

### 3.5. Data Analysis

Data were processed and analyzed using Power BI, a business intelligence platform that supports Extract, Transform, Load (ETL) operations and dynamic visualization. The ETL process involved cleaning raw data by removing duplicates, handling null values, adjusting data types, and splitting complex fields into analyzable components. This transformation enabled efficient visualization and analysis without extensive manual intervention. Key visualizations included bar charts, doughnut charts, stacked bar charts, and clustered column graphs. These were used to compare performance between groups, analyze progress over time, and identify areas of difficulty or improvement.

Power BI's automation capabilities minimized human error and expedited the processing of large datasets. By integrating academic indicators with learner interaction metrics, the analysis provided multi-dimensional insights into the platform's impact on learner performance and engagement.

## 4. Results and Discussion

### 4.1. Key Findings

This section reports key findings from the Power BI analysis, focusing on learner performance across course groups, overall proficiency levels, and patterns of learner engagement. From the eight original dashboards generated during analysis, only the most representative visualizations are included here to support clarity and alignment with the study's objectives.

Figure 1 displays the average MagniLearn performance scores across six course groups. Among these, Communicative English 64 recorded the highest average platform score at 21.83%, followed by Communicative English 26 with 19.14%. The group with the lowest average was Communicative English 04, at 11.56%.

These discrepancies suggest variation in learner engagement or instructional implementation across classes, despite being aligned under a shared institutional syllabus. Higher scores may indicate more consistent platform use, better integration into classroom activities, or stronger learner motivation. Conversely, the lower scores observed in some groups may reflect reduced time-on-task, passive usage, or limited instructor involvement in AI-facilitated learning.

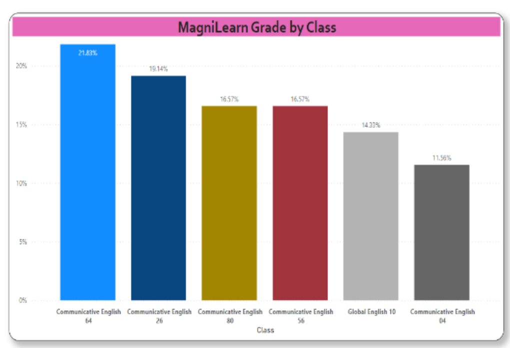


**Figure 1:** MagniLearn grade with respect to Class

Figure 2 illustrates the CEFR-level distribution of students based on MagniLearn's internal placement assessments. A clear majority of learners were positioned at the B1 level (53.5%), followed by B2 (29.3%), with smaller groups at A2 (14.0%), C1–C2 (2.0%), and A1 (1.1%).

These proportions suggest that the platform was most frequently accessed or retained by learners within the intermediate proficiency range, where adaptive feedback and scaffolded instruction may be optimally effective. This distribution aligns with prior research in intelligent CALL systems, which has noted that intermediate-level learners often benefit more from personalized, data-driven learning environments due to their sufficient baseline skills and receptiveness to corrective feedback (e.g., Vasylets et al., 2021; Bai & Wang, 2023).

In contrast, learners at the C1–C2 level may have required more advanced content than what the platform provided, while those at A1 may have struggled to fully engage without external support or foundational instruction. The small percentage at both ends of the spectrum highlights a potential need to further tailor platform features to accommodate both high- and low-proficiency learners more effectively.

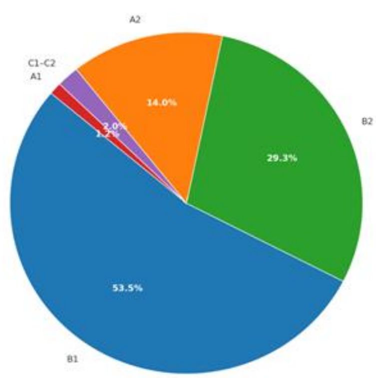


**Figure 2:** CEFR-Level Distribution of Learners

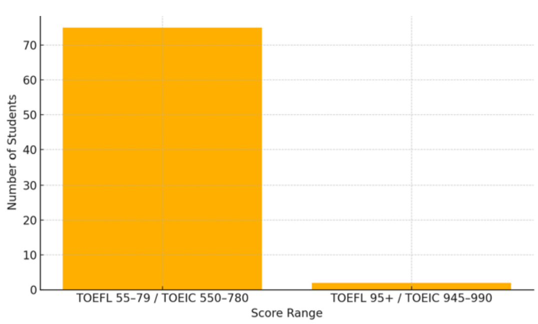


**Figure 3:** TOEFL and TOEIC Correlation

Figure 3 presents the distribution of learners based on standardized English test scores, specifically TOEFL and TOEIC. A substantial majority of students (n = 75) were situated within the mid-range proficiency band, defined as TOEFL 55–79 or TOEIC 550–780. In contrast, only two students (n = 2) achieved scores in the advanced range (TOEFL 95+ / TOEIC 945–990), accounting for just 2.6% of the sample.

This distribution pattern reinforces the CEFR-level findings reported in Figure 2, where the vast majority of learners were clustered at the B1 and B2 levels. The convergence between standardized test scores and CEFR placement data suggests that the sampled population represents a predominantly intermediate proficiency cohort. This is relevant for evaluating the effectiveness of the MagniLearn platform, as prior studies have shown that intermediate-level learners are most responsive to adaptive learning systems due to their readiness for individualized feedback and pattern-based instruction (Huang et al., 2019; Vasylets et al., 2021).

Furthermore, the low number of students in the highest proficiency range may reflect the institutional context, where English instruction is typically required of learners who have not yet met advanced-level benchmarks. This underscores the importance of aligning AI-based instruction with the actual proficiency profiles of the target learner population.

Figure 4 presents a dual-axis visualization showing the number of exercises completed (left axis) and grammar/vocabulary items practiced (right axis) by the top three students using the MagniLearn platform. Hyunjang completed the highest number of exercises (n = 3,850) and reviewed over 1,300 distinct grammar and vocabulary points, significantly exceeding the other two students. Park Sangseo and Ji Suhyun completed approximately 2,700 and 2,600 exercises respectively, each with under 1,000 unique items practiced.

This steep contrast suggests a possible association between high engagement levels—measured through both task volume and linguistic focus—and accelerated learning outcomes. While causality cannot be inferred from this small sample, the trend is consistent with adaptive learning literature indicating that repetitive exposure and sustained practice can enhance retention and performance gains, particularly in grammar-focused tasks (e.g., Teo, 2011; Bai & Wang, 2023).

These findings also highlight individual variation in learner behavior, reinforcing the need for adaptive platforms to accommodate different pacing and usage intensity. Future analyses should investigate whether high-volume users demonstrate proportionately greater gains in CEFR levels or platform-derived proficiency scores.

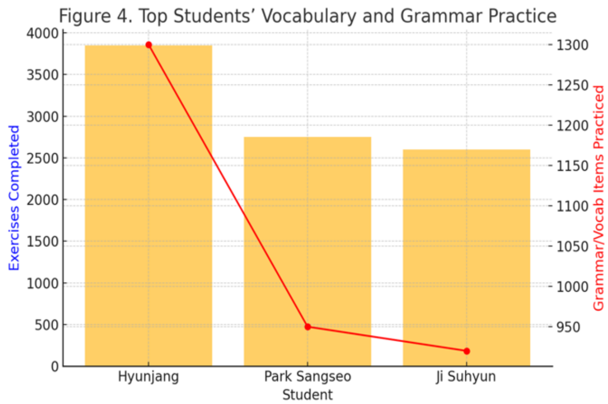


**Figure 4:** Top Students' Vocabulary and Grammar Practice

## 4.2. Student Performance Highlights

Figure 5 presents a comparative summary of top-performing students' engagement with vocabulary and grammar exercises, measured by the number of exercises completed (blue bars) and the number of words and grammar rules practiced (green bars). Notably, Hyunjang completed nearly 4,000 exercises and practiced approximately 1,500 vocabulary and grammar items, followed by Park Sangseo and Ji Suhyun with slightly lower—but still substantial—engagement.

While this figure does not reflect time in minutes, the volume of activity implies a consistently high level of learner persistence. Prior research suggests that such sustained exposure to personalized input—particularly when scaffolded through adaptive systems—can result in measurable gains in linguistic accuracy and confidence (Bai & Wang, 2023; Vasylets et al., 2021).

Figure 6 displays lesson completion rates across selected "Communicative English" classes. Students enrolled in Class 64 and Class 56 reached a 100% completion rate for assigned lessons, marking the highest engagement among peer groups. A particularly notable case is student Kim Gun Ho, who was concurrently enrolled in both classes and achieved full task completion in both. This case provides micro-level support for the platform's ability to maintain motivation and task continuity across contexts.

These findings reinforce the argument that frequent, consistent, and diversified practice—especially when combined with feedback and learner autonomy—supports successful integration of AI-enhanced instruction in language classrooms.

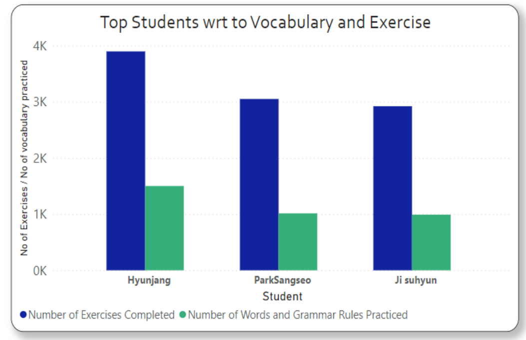


**Figure 5:** Time Spent on Vocabulary and Exercises

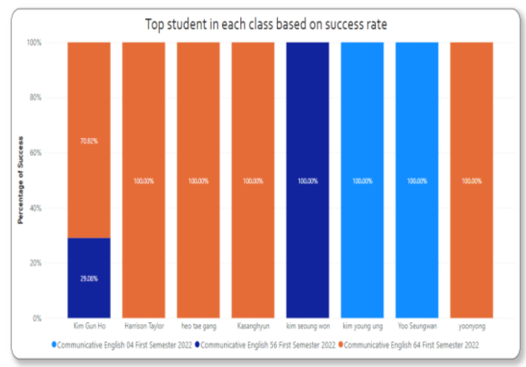


**Figure 6:** Lesson Success Rates by Course

Figure 7 illustrates the distribution of active learners by class. In this study, active learners were operationally defined as students who completed at least one full lesson on the MagniLearn platform. The class "Communicative English 64" recorded the highest proportion of active learners (22.14%), followed by "Communicative English 56" (18.57%) and "Communicative English 26" (17.86%). In contrast, "Communicative English 04" had the lowest proportion of active students (11.43%).

These differences may be attributed to a range of pedagogical or contextual factors, including variations in instructor facilitation style, learner autonomy, or integration of the platform into class routines. While further qualitative data would be required to confirm causality, the observed disparity suggests that instructor engagement and classroom culture may significantly influence platform uptake.

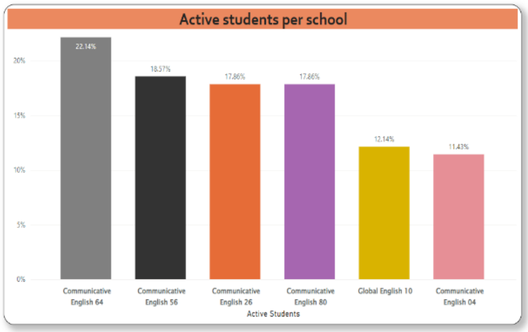


**Figure 7:** Active Student Distribution by Class

Figure 8 presents the top five students ranked by overall MagniLearn grade performance. Nguyen Thi Hong Ngoc achieved the highest score at 27.53%, significantly outperforming peers such as Gu Yerin (18.78%) and Lee Soohyun (18.50%).

This result highlights a potential link between consistent platform engagement and academic achievement, particularly for learners demonstrating strong digital literacy and self-regulated study habits. While further investigation is required to isolate causal factors, the exceptional performance of this learner suggests that the MagniLearn platform may be especially beneficial when paired with high motivation and frequent interaction.

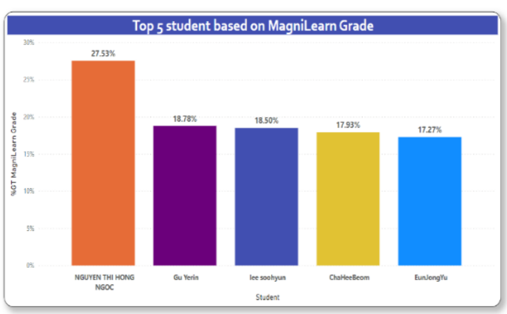


**Figure 8:** Top 5 student based on MagniLearn Grade

### 4.3. Interpretation and Implications

The analysis of Figures 1 through 8 suggests that AI-assisted language instruction via the MagniLearn platform positively influenced learner engagement and performance, particularly among intermediate-level students. A substantial majority of participants were concentrated in the B1 (53.55%) and B2 (29.30%) CEFR bands (Figure 2), with only a small proportion advancing beyond this range. This pattern was mirrored in standardized assessment data (Figure 3), where most learners scored within the TOEFL 55–79 and TOEIC 550–780 brackets. These results indicate that the platform effectively supports core language development but may have limited impact on propelling learners to advanced proficiency without supplementary instructional input.

High-performing individuals such as Hyunjang and Nguyen Thi Hong Ngoc (Figures 4 and 8) demonstrate a clear association between sustained engagement, task completion volume, and learning outcomes. These findings align with previous research emphasizing the importance of consistent, personalized input in driving language acquisition. Furthermore, class-level data reveal that certain cohorts—most notably "Communicative English 64"—achieved complete task completion rates (Figure 6), suggesting a strong instructional match and learning climate. In contrast, lower participation and success rates in other groups (Figures 6 and 7) may reflect gaps in instructional integration, learner support, or motivational scaffolding.

The pedagogical implications are twofold. First, MagniLearn's adaptive capacity to address grammar and vocabulary micro-skills in real time appears to reinforce learner retention and accuracy. Second, the platform's limitations at higher proficiency levels highlight the need for continued human facilitation. AI tools alone cannot replace the nuanced feedback, encouragement, and curriculum alignment that teachers provide—especially when aiming for C1-level or beyond.

In conclusion, this study underscores the potential of AI-enhanced platforms as valuable supplements to tertiary-level language instruction. Institutions should explore blended learning models that integrate adaptive technologies with structured classroom practices. Future research should further investigate the long-term effects of learner-AI interaction, motivational design, and differentiated scaffolding strategies across varied proficiency profiles and educational contexts.

## 5. Conclusions

This study explored the impact of an AI-powered adaptive learning platform, MagniLearn, on undergraduate students' English language development over a semester-long course. A total of 140 students participated across various Communicative English classes, engaging with the platform to varying degrees. By triangulating MagniLearn performance data with CEFR classifications and standardized test scores (TOEFL, TOEIC), the study identified patterns of learner engagement and proficiency growth that underscore the growing potential of AI-assisted instruction in language education.

Findings revealed a positive correlation between sustained, high-intensity use of the platform and measurable improvements in language performance. Students who completed more exercises and devoted greater time to targeted vocabulary and grammar practice consistently achieved higher proficiency scores. Notable cases—such as Hyunjang and Nguyen Thi Hong Ngoc—demonstrated significant gains, suggesting that MagniLearn's micro-skill scaffolding and adaptive feedback mechanisms are particularly effective for learners at the intermediate (B1–B2) level, where progress often stalls.

However, progression to advanced proficiency (C1 and above) remained limited. Only a small minority reached high TOEFL/TOEIC equivalency scores, indicating that while AI platforms like MagniLearn are effective for reinforcing foundational skills, they may require integration with instructor-led strategies and advanced content to support more complex linguistic development.

This study also acknowledges several limitations. The absence of a consistently applied control group and the modest sample size, particularly for between-group comparisons, limit the generalizability of the results. Moreover, the lack of qualitative data—such as learner perceptions, motivational profiles, or usability feedback—restricts insight into how students experience and respond to AI-based learning environments.

Future research should adopt a mixed-methods approach, engaging larger and more diverse learner populations across institutions. Incorporating interviews, reflective journals, and classroom observations would offer richer understanding of how AI tools influence learner autonomy, motivation, and long-term engagement.

In sum, this study affirms the pedagogical value of adaptive AI platforms in second language instruction. When integrated thoughtfully with human facilitation and reflective learning design, such technologies hold significant promise for enhancing learner outcomes—particularly at scale and in blended learning contexts.

## 5.1. Limitations and Future Research

While this study provides valuable insights into the application of the MagniLearn platform in a university-level English learning context, several limitations must be acknowledged.

First, the relatively small sample size in the experimental component—specifically within the control group (n = 8) and treatment group (n = 12)—constrains the statistical power of comparative analysis. Although broader usage data from 140 students were analyzed descriptively, the lack of robust experimental control limits the ability to isolate the effects of the platform from confounding variables such as instructor style, learner background, or extracurricular exposure to English.

Second, the study's timeframe was confined to a single academic semester. While short-term improvements in vocabulary acquisition, lesson completion, and CEFR-level progression were observed, the durability of these gains over time remains unclear. Longitudinal use across multiple semesters may yield different insights into learner motivation, autonomy, and potential plateau effects, which were beyond the present study's scope.

Third, the study relied primarily on quantitative data—MagniLearn usage statistics, CEFR classifications, and standardized test scores (TOEFL/TOEIC)—without triangulation from qualitative sources. Learners' subjective experiences, perceptions of AI-generated feedback, cognitive load, and emotional engagement were not explored. These affective and metacognitive dimensions are critical to understanding how students interact with AI-powered learning environments.

Future research should adopt a mixed-methods approach that combines analytics with qualitative instruments such as interviews, learner journals, or focus groups to capture user experiences in greater depth. Longitudinal studies conducted across diverse institutional and cultural contexts would also help assess the scalability and long-term efficacy of AI-enhanced instruction. Furthermore, comparative studies examining different adaptive platforms—or contrasting AI-supported versus instructor-only learning models—could provide a more nuanced understanding of the pedagogical affordances and limitations of current AI tools in language education.

## 5.2. Final Reflection

In a rapidly evolving educational landscape, flipped learning entails more than a reordering of instructional delivery. It demands intentional and pedagogically sound design—content that is modular, adaptive, and learner-centered. The creation of flipped learning materials is thus not merely a technical process but an instructional commitment. It reflects an ethical responsibility to foster learning environments that are not only efficient but also empowering and inclusive.

Instructional design, in this context, becomes an ethical practice: a deliberate act of ensuring that educational content is accessible, relevant, and meaningful. At its core, the design of learning experiences should embody the very principles we seek to instill in our students—autonomy, sustained engagement, and deep understanding.

This study offers one such model of integration, rooted in learner performance data, refined through iterative practice, and attuned to the demands of 21st-century education. As adaptive technologies like MagniLearn continue to evolve, so too must our pedagogical frameworks, ensuring that innovation in education remains anchored in empathy, equity, and meaningful learning.

## References

Bai, B., & Wang, J. (2023). The application of AI-powered writing feedback on learners' engagement and self-efficacy: Evidence from EFL writing classrooms. *System, 114*, 102996. https://doi.org/10.1016/j.system.2023.102996

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Colpaert, J. (2020). Editorial position paper: How virtual is your research? *Computer Assisted Language Learning, 33*(7), 653–664. https://doi.org/10.1080/09588221.2020.1744663

Huang, F., Teo, T., & Scherer, R. (2019). Investigating the antecedents of university students' perceived ease of use and perceived usefulness of a learning analytics dashboard. *Interactive Learning Environments, 29*(2), 262–277. https://doi.org/10.1080/10494820.2019.1587462

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education. *Education and Information Technologies, 25*(2), 971–990. https://doi.org/10.1007/s10639-019-10020-8

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2017). What public media reveals about learning analytics: A systematic review of media articles. *British Journal of Educational Technology, 48*(3), 693–705. https://doi.org/10.1111/bjet.12438

Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. Q. (2015). From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (37, 957–966). JMLR.org. https://doi.org/10.5555/3045118.3045221

Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology, 17*(3), 135–156.

McCrocklin, S. A. (2019). ASR-based dictation practice for pronunciation improvement. *Language Learning & Technology, 23*(1), 64–84.

Park, M., & Son, J.-B. (2020). Technology integration in English language teaching: A qualitative study of EFL teachers' perceptions. *The Asia-Pacific Education Researcher, 29*(3), 217–226. https://doi.org/10.1007/s40299-019-00496-5

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2022). Automated written corrective feedback: How learners respond. *System, 109*, 102893. https://doi.org/10.1016/j.system.2022.102893

Sato, R., Yamada, M., & Kashima, H. (2022). Re-evaluating word mover's distance. *Machine Learning, 111*(2), 3692–3720. https://doi.org/10.1007/s10994-022-06296-z

Tashu, Y., & Horváth, I. (2018). Semantic similarity-based automatic assessment of textual answers in e-learning. *Computers & Education, 126*, 1–13. https://doi.org/10.1016/j.compedu.2018.06.012

Teo, T. (2011). Factors influencing teachers' intention to use technology: Model development and test. *Computers & Education, 57*(4), 2432–2440. https://doi.org/10.1016/j.compedu.2011.06.008

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Vasylets, O., Niemeier, S., & Götz, S. (2021). Intelligent CALL and learner autonomy: A systematic review. *ReCALL, 33*(2), 161–177. https://doi.org/10.1017/S0958344020000222

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Wei, C., Wang, B., & Kuo, C.-C. J. (2023). SynWMD: Syntax-aware word mover's distance for sentence similarity evaluation. *Pattern Recognition Letters, 170*, 48–55. https://doi.org/10.1016/j.patrec.2022.05.016

Wilson, K., & Roscoe, R. (2020). Automated writing evaluation and feedback: Multiple case studies of learners revising with Grammarly. *Journal of Educational Computing Research, 58*(2), 327–355. https://doi.org/10.1177/0735633119845694

Woo, J., & Choi, J. (2021). A systematic review on the use of artificial intelligence in English education. *The Journal of Asia TEFL, 18*(2), 649–660. https://doi.org/10.18823/asiatefl.2021.18.2.14.649

Zhai, X., Chu, X., Wang, M., & Wang, J. (2023). ChatGPT in education: A systematic review and future research agenda. *Education and Information Technologies.* https://doi.org/10.1007/s10639-023-11753-9

Zhang, Z., & Hyland, K. (2021). Student engagement with automated feedback: Insights into feedback literacy. *Assessing Writing, 49*, 100524. https://doi.org/10.1016/j.asw.2021.100524