



An Exploratory Study of an AI-Supported English Reading Class in Korean High School

Mingi Kim (Yanggu High School)

Received: 13 September 2025
Revised: 23 September 2025
Accepted: 27 September 2025

Kim, Mingi. (2025). An exploratory study of an AI-supported English reading class in Korean high school. *Modern English Education*, 26, 383-402.

Keywords

AI-supported instruction,
personalized learning,
differentiated instruction,
student motivation,
high school English education
AI 활용 수업, 개별화 학습,
수준별 수업, 학습 동기,
고등학교 영어 교육

Mingi Kim

Teacher
Yanggu High School
mgkimedu@gmail.com

Abstract

This year-long exploratory study examines an AI-supported personalized reading approach aimed at improving low student engagement. Conducted with 63 second-grade high school students in a rural area, the researcher, a high school English teacher, utilized AI to generate engaging reading passages at various proficiency levels. Students were allowed to choose texts that best matched their abilities. A mixed-methods design was employed, collecting mock test scores to identify achievement trends and using comparative surveys to evaluate student perceptions compared to the previous grammar-translation method. The findings revealed a positive trend in academic achievement, marked by a decrease in low-level students and an increase in middle and high-level students. Additionally, students reported a significant improvement in their learning experience compared to the prior method, demonstrating much higher levels of satisfaction, motivation, and interest with the new approach. Notably, the autonomy students had in selecting passages successfully preserved their self-esteem, addressing a key challenge in traditional ability grouping. While acknowledging its exploratory nature and context-specific limitations, this research presents a practical model for revitalizing secondary English education and offers a rare long-term perspective on AI integration in an authentic K-12 classroom.

INTRODUCTION

In the South Korean educational context, English is often treated less as a tool for communication and more as a critical subject for university entrance exams. This high-stakes testing environment frequently places high school students in a passive learning situation, where they feel compelled to study English regardless of their actual interest or motivation. As a result, many students experience significant stress and pressure, which can lead to negative attitudes toward English learning (D. Kim & T. Kim, 2018). This persistent issue of low engagement and motivation poses a significant challenge for effective English education in secondary schools.

To address these challenges, this study designed and implemented a novel teaching method over one academic year. This approach is built on several core principles grounded in established motivational theories. First, it utilizes AI-powered

personalization, which in this study refers to the use of AI to generate reading passages tailored to students' individual interests and proficiency levels. Second, to foster student autonomy, learners were empowered to select their own reading passages from the leveled options provided on a single topic. This method enabled differentiated instruction by allowing students to engage with different content within a heterogeneous classroom, avoiding the potential negative effects of ability grouping.

In this pedagogical framework, AI is not envisioned as a replacement for the instructor but is positioned as a powerful tool to augment the teacher's capabilities. The teacher acts as a critical curator, responsible for validating, revising, and adapting the AI-generated materials to ensure their pedagogical suitability and quality. This study, therefore, explores a model where technology serves to enhance, not replace, the professional judgment of the teacher in a real classroom setting.

Positioning AI in this supportive role is particularly relevant given a specific gap in the academic literature. A recent systematic review reveals that most empirical studies on AI in English education are short-term interventions, are concentrated in higher education, or focus on writing rather than reading (Lo et al., 2024). There remains a notable lack of long-term, classroom-based studies investigating the practical implementation and student perceptions of AI-personalized reading within the K-12 context—a gap this study was designed to address.

Accordingly, this year-long, classroom-based exploratory study seeks to address the aforementioned gap in the literature by investigating the practical application of an AI-supported reading approach. To this end, the study is guided by the following research questions: (1) What patterns of change, if any, are observed in students' English academic achievement over the course of a year-long implementation of an AI-supported personalized reading program? (2) How do students perceive the AI-supported personalized reading approach in terms of its impact on their learning experience, engagement, and motivation, compared to the grammar-translation method? By answering these questions, this study aims to provide practical insights for educators and build an empirical foundation for future research on AI-integrated language teaching in K-12 settings.

LITERATURE REVIEW

Leveraging AI for Personalized Language Learning: Affordances & Caveats

The integration of AI, particularly Large Language Models like ChatGPT, offers unprecedented opportunities for realizing personalized and differentiated instruction. A recent systematic review by Tessensohn et al. (2025) confirms that AI-powered tools are increasingly being used to enhance a variety of reading skills, from basic comprehension to critical thinking. A growing body of empirical research further confirms that specific AI applications can effectively generate high-quality, personalized reading materials tailored to student interests and proficiency levels (Xiao et al., 2023). For instance, studies consistently show that AI-generated texts based on learners' personal interests significantly boost affective factors such as reading enjoyment and motivation when compared to traditional textbook materials (J. H. Lee et al., 2023; Yılmaz & Aydın, 2025). This suggests a strong potential for AI to address the affective needs of learners, which is a critical component of engagement.

Beyond these affective benefits, AI's capacity for personalization also yields significant cognitive advantages, particularly when implementing differentiated instruction for learners at varying proficiency levels. Multiple studies have found that personalized interventions yield the most significant gains in reading comprehension and self-efficacy for low and intermediate-proficiency students (Sudin & Swanto, 2024; Sun & Yusof, 2025). This heightened effectiveness is likely due to AI's capability to adjust content difficulty to provide comprehensible input. Supporting this, a study by Çelik et al. (2024) demonstrated that simplifying authentic texts with ChatGPT led to significant improvements in students' reading comprehension and inferencing skills, making challenging materials more accessible for those who might otherwise struggle. However, it is also noted that advanced learners may require more challenging, specially adapted materials to show meaningful improvement, underscoring the need for careful pedagogical design in a differentiated classroom (Sudin & Swanto, 2024).

Despite these promising findings, the application of AI is not without limitations and requires a critical approach from educators. A primary concern is the quality and reliability of AI-generated content, which can contain factual inaccuracies or 'hallucinations', where the AI fabricates non-existent information or references, posing a risk to students' learning and critical judgment (Bektik et al., 2024) and perpetuate cultural biases, often reflecting Western, Educated, Industrial, Rich, and Democratic (WEIRD) cultural values to students (Zheng & Stewart, 2024). These challenges underscore the indispensable role of the teacher, not as a mere implementer of technology, but as a critical curator who evaluates, revises, and adapts AI-generated materials (Snashall, 2023). This teacher-AI collaborative model, where educators refine AI

suggestions to better fit the specific learning context, is emerging as a key factor for successful implementation (Sun & Yusof, 2025; Zhang, 2025).

This body of work, while confirming AI's potential, also reveals critical gaps that the present study aims to address. Many of the aforementioned empirical studies are either short-term interventions (Çelik et al., 2024; Yılmaz & Aydın, 2025) or are concentrated in higher education (Çelik et al., 2024; Yılmaz & Aydın, 2025; Zhang, 2025) and vocational settings (Sun & Yusof, 2025). There remains a notable scarcity of long-term, classroom-based research investigating the practical application of AI within the K-12 secondary education context. This gap is particularly salient in the South Korean high school context, where chronic low student engagement in English classes necessitates novel and sustainable pedagogical approaches. Therefore, this study aims to fill this gap by providing a year-long exploratory case study of an AI-supported English reading class where the teacher, acting as a pedagogical curator, systematically generated, validated, and implemented personalized materials within a real Korean high school setting.

Theoretical Framework: Motivation and Interest in Reading Instruction

A persistent challenge in secondary English education is fostering intrinsic motivation, the drive to engage in a task for its inherent enjoyment rather than for external rewards (Ryan & Deci, 2000). Foundational theories like Self-Determination Theory (SDT) and Interest Theory provide a robust framework for understanding how to cultivate this motivation. SDT posits that intrinsic motivation flourishes when three basic psychological needs are met: autonomy (a sense of choice), competence (a feeling of effectiveness), and relatedness (a sense of belonging) (Ryan & Deci, 2000). The pedagogical design of the present study was a direct attempt to support these needs. Autonomy was addressed by empowering students to select their own reading passages; competence was supported by providing AI-generated texts tailored to their proficiency level; and relatedness was fostered through collaborative activities where students worked with peers who selected the same text.

Complementing SDT, Interest Theory suggests that situational interest, sparked by personally relevant and engaging content, is a critical first step toward developing a more enduring individual interest in a subject (Hidi & Renninger, 2006). The use of AI to generate texts based on topics students found interesting (e.g., humorous stories, current events) was therefore a deliberate strategy to trigger this situational interest. When students are given texts that are either too difficult (hindering competence) or irrelevant to their lives (failing to spark interest), their motivation naturally declines.

In an effort to address this, previous studies in the Korean context explored using materials like English newspapers to connect learning to real-world topics (e.g., H. Kim et al., 2010; Jung & Jang, 2017). These studies reported positive outcomes, such as increased interest and reading comprehension. However, a critical limitation of this approach was the difficulty of systematically catering to the diverse proficiency levels within a single classroom. A single newspaper article, while interesting, could simultaneously be too challenging for some students and not challenging enough for others. This highlights a gap that the current study aims to address: leveraging technology to provide materials that are personalized not only for interest but also for proficiency, thereby supporting both situational interest and the need for competence.

In synthesis, grounded in Self-Determination Theory and Interest Theory, the present study was designed to overcome the key limitation of prior authentic material approaches, such as using newspapers, which failed to cater to diverse proficiency levels. The intervention implemented these theories by using AI to generate materials that supported competence (tailored to proficiency) and fostered autonomy (empowering student choice) and relatedness (facilitating peer collaboration), with the pedagogical design triggering situational interest (personalized to topics).

Theoretical Framework: Differentiated Instruction and Scaffolding

Beyond motivation, effective instruction in a mixed-level classroom requires differentiation—adapting content, process, or products to meet the diverse needs of learners (Tomlinson, 2014). This principle is deeply rooted in theories like Vygotsky's (1978) concept of the Zone of Proximal Development (ZPD), which is the space between what a learner can do independently and what they can achieve with guidance. Similarly, Krashen's (1982) Input Hypothesis posits that language acquisition occurs most effectively when learners receive 'comprehensible input' that is slightly beyond their current proficiency level ($i+1$). Both theories underscore the importance of providing appropriately challenging tasks.

However, despite its strong theoretical grounding, the practical implementation of differentiated instruction has historically been a significant challenge for educators. Preparing tailored materials for diverse proficiency levels demands immense time and effort, often creating an unsustainable workload for teachers (J. H. Lee et al., 2023; Yılmaz & Aydın, 2025). The recent advent of Generative AI, such as ChatGPT, offers a transformative solution to this long-standing problem. These AI tools empower teachers to generate or adapt high-quality, level-appropriate texts with simple commands, thereby

making the systematic implementation of differentiated instruction a practical and achievable goal in the classroom (Yılmaz & Aydın, 2025; Zhang, 2025).

Empirical studies on traditional differentiated instruction, which often involve grouping students into separate classes, have yielded mixed results and highlighted several critical challenges. While some studies reported effectiveness across all proficiency levels (Hong & K. Kim, 2014), others found benefits were limited to high-achieving students (E. Kim & Rha, 2015) or, conversely, only to low-achieving students (Kang & Cho, 2012; J. Lee & Jang, 2014). A key issue is the method of grouping; when students are assigned to lower-level classes based on test scores, they may suffer from decreased self-esteem and a disrupted learning environment, which negates potential academic gains (E. Kim & Rha, 2015). In contrast, allowing students to choose their class level based on willingness has been shown to protect self-esteem and improve scores (J. Lee & Jang, 2014). These findings suggest that while differentiation is valuable, physical ability grouping can introduce negative affective and logistical consequences.

The teaching method in this study sought to apply these principles by combining AI-driven content generation with student-led choice. The instructor first used AI to generate texts at varying difficulty levels on the same topic. Then, crucially, students were empowered to autonomously select the passage they felt best matched their own ability level. This approach enabled the differentiation of content within a single, inclusive classroom, providing each student with a text that approximated their individual ZPD or $i+1$ level. By replacing externally imposed ability grouping with a model based on student autonomy, our approach aimed to reap the academic benefits of differentiation while directly mitigating the potential harm to self-esteem identified in prior research on traditional level-based classes in the Korean context.

METHODOLOGY

Research Context and Participants

This study was conducted at a public high school located in a rural area of Gangwon State, South Korea. The context is generally characterized by a homogeneous student population and limited access to the extensive private tutoring or extracurricular English programs commonly found in urban centers. This specific context should be considered when interpreting the findings, as it may influence their generalizability.

The participants were 63 second-grade students whom the researcher taught over a full academic year. Prior to the intervention, when the participants were in their first grade, English instruction was delivered primarily through the grammar-translation method. This approach, which focused on the teacher translating textbook passages and explaining grammatical rules, is often associated with low student engagement in the Korean high school context (Park et al., 2014). The present study was designed to address this long-standing issue by introducing a novel pedagogical approach aimed at improving student engagement and motivation.

Ethical considerations were paramount throughout the study. Before the intervention began, the purpose and procedures of the research were explained to all students. Informed consent was obtained, and students were assured that their participation would not influence their grades in any way. All data collected, including test scores and survey responses, were anonymized to protect student privacy.

Instructional Design and Procedures

Preliminary Trials and Rationale for the New Method

Initially, reading passages from mock tests, chosen for their potential interest, were used as class materials. However, this approach failed to engage the majority of students; only a quarter of the class, primarily high-level students, actively participated. A second attempt involved using AI to revise standard textbook passages to different difficulty levels. This also proved ineffective, as the core topics of the textbook remained uninteresting to the students. The persistent low engagement, even with these adjustments, necessitated the introduction of the new teaching method, which was designed to systematically address both student interest and proficiency level simultaneously.

Materials Preparation and Validation

The class materials consisted of two primary types: contemporary articles from The Wall Street Journal (WSJ) and original

humorous narratives centered on student experiences. These texts were adapted into low, middle, and high-level reading passages to accommodate varying student proficiency levels, utilizing AI tools.

For the instructional materials derived from WSJ articles, the instructor first selected recent articles based on student interest and educational relevance. For instance, in response to student requests for aerospace-related topics, an article discussing ‘SpaceX’ was chosen. Additionally, an article on ‘Disney characters created with deepfake technology’ was selected to address a significant social issue, while a piece on ‘fentanyl overdose’ was used as a preventative educational tool regarding substance abuse. After an article was selected, complementary YouTube videos were chosen to increase student engagement and provide essential background knowledge. For example, when teaching about fentanyl overdoses, a video illustrating the severe struggles of those affected by the drug was presented to provide context and deepen students’ understanding.

In parallel, humorous stories reflecting events in school life were created using AI tools from basic outlines developed by the teacher. This approach allowed for the generation of relatable and engaging content. For instance, inspired by an incident where a student’s loud scream from a leg cramp amused the class, a story was created in which the student transformed into a ‘mouse.’ This narrative was based on the fact that the Korean words for ‘cramp’ and ‘mouse’ are both ‘jwi.’ In the creation of these stories, students with lower academic achievement were often cast as the main characters. This strategy was employed to enhance their motivation, as high-achieving students typically demonstrated consistent engagement with the course material.

Following the selection of articles from The Wall Street Journal and the creation of humorous stories, a systematic process was employed to adapt each source text for low, middle, and high-level readers. This was accomplished by providing specific commands to an AI tool, enabling the generation of three differentiated passages from a single subject. For low-level texts, the command was, ‘Revise this text to be suitable for kindergarten and elementary school English learners.’ For the middle level, the prompt was modified to ‘middle school English learners.’ For the high level, one of two methods was used: either the text was adjusted with the prompt ‘undergraduate English learner,’ or the original news articles were utilized with difficult vocabulary and expressions manually replaced by simpler alternatives.

Following the initial AI generation, the instructor further refined the texts to ensure they were pedagogically appropriate. This critical step involved manually adjusting factors such as sentence length, vocabulary complexity, and overall conceptual density to better align with the students’ observed proficiency. This process acknowledges that the AI prompts served as practical heuristics rather than precise diagnostic tools for determining a student’s optimal learning level.

Ultimately, the final judgment of a text’s appropriateness, and the necessary instructional support to help students comprehend it, remained the critical responsibility of the human teacher. Therefore, AI was utilized as a starting point for differentiation, not as an autonomous arbiter of student proficiency. This teacher-led validation was also essential to review the AI-generated content for factual accuracy and coherence.

The First Period of the English Class

At the beginning of the lesson, students either watched YouTube videos when the class materials were the latest articles, or listened to an overview of the storyline when the class materials were humorous stories. Afterwards, they voluntarily selected reading passages at low, middle, or high-levels and engaged in reading activities with classmates who had chosen the same level, working together to comprehend the text. The researcher circulated throughout the classroom, answering students’ questions and facilitating their understanding. Once students had grasped the contents of the text, they answered comprehension questions generated by AI tools. Additionally, high-level students were sometimes required to write their answers in English to enhance their writing skills. A sample of the differentiated reading passages and corresponding comprehension questions is provided in Appendix B.

Occasionally, the researcher prepared foundational level reading passages, designed to match the English proficiency of a first-semester, first-year middle school student, for students performing below basic academic achievement and provided detailed explanations during class due to their difficulty in comprehending low-level reading passages. In some cases, high-level students selected middle-level passages to assist their peers, while low and middle-level students challenged themselves by choosing more advanced texts.

The Second Period of the English Class

In the second period, grammar instruction was delivered briefly and repeatedly. Instead of translating all English sentences in the texts, the researcher selected a few sentences that contained key grammar points for explanation. The focus was on

the use of to-infinitives (as nouns, adjectives, and adverbs), the -ing form (gerunds, present participles, and participle phrases), and the distinction between the relative pronoun ‘that’ and the noun clause ‘that’. These elements were chosen because, despite their similar forms, their diverse functions often confuse students.

During the 50-minute session, the first 15 minutes were dedicated to explaining grammar points in a low-level text, the next 15 minutes to those in a middle-level text, and the final 15 minutes to high-level grammar points. Students were required to pay attention for their chosen passage and gently reminded to remain focused for the full 15 minutes. Afterward, they had the option to listen to grammar explanations for other passages or enjoy free time.

At the end of the second period, the researcher distributed all three passages, along with their translations, to all students. These passages were included in the midterm and final exams, and grammar-related questions were excluded because grammar was not the central focus of the English classes. Appendix A lists the reading passage topics covered in each English class.

Data Collection and Analysis

This study employed a mixed-methods approach to explore the implementation of the new teaching method. Given the exploratory nature of the research, the focus was on observing patterns and perceptions rather than establishing causal links.

Academic Achievement Data

To observe trends in academic achievement, a one-group pre-post design was used, collecting scores from mock tests administered in March, June, and September of the intervention year. It is critical to note that due to the nature of the study—a classroom-based intervention implemented with all second-grade students taught by the researcher—a concurrent control group was not feasible. Therefore, this study does not claim to measure the causal effect of the intervention.

Instead of a control group, historical data from two sources were used as contextual reference points: 1) the 2023 test scores of the same participants (when they were first-grade students), and 2) the 2023 test scores of the previous year's second-grade cohort. This comparison was not intended to serve as a rigorous control but to provide a broader context for interpreting the trends observed during the intervention year, acknowledging that confounding variables such as student maturation or variations in test difficulty could not be controlled.

Student performance on the mock tests was categorized into three proficiency tiers based on the national rating system: high-level (ratings 2-4), middle-level (ratings 5-6), and low-level (ratings 7-9). Changes in the proportion of students in each tier were analyzed descriptively.

The grading system for the mock tests, managed by the education office, is an absolute grading system. Scores above 90 are classified as rating 1, and scores above 80 are rating 2 (Korean Institute for Curriculum and Evaluation, 2016). Table 1 shows the grading system for the mock tests. All participants were categorized into three tiers: rating 2-4 (high-level), 5-6 (middle-level), and 7-9 (low-level).

TABLE 1

The Grading System for the Mock Test

| Rating | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------|--------|-------|-------|-------|-------|-------|-------|-------|------|
| Scores | 100~90 | 89~80 | 79~70 | 69~60 | 59~50 | 49~40 | 39~30 | 29~20 | 19~0 |

Source: Korean Institute for Curriculum and Evaluation (2016).

Assessing Student Perceptions: A Comparative Approach

To evaluate student perceptions in a comparative manner, this study collected both survey data and observational records. A researcher-developed survey, informed by a review of prior studies on differentiated instruction and student motivation (e.g., Ahn & E.-S. Park, 2014; K. Kim, 2011; Lim, 2014; Shim & Lim, 2017), was administered at two different time points. Items on the survey were rated on a five-point Likert scale. The questionnaire was intentionally kept brief, comprising only the most essential items. This was a deliberate methodological choice to counteract the risk of respondent fatigue, a common issue with lengthy surveys where participants may provide rushed or insincere answers. The full list of survey items for the grammar-translation method and the new teaching method can be found in Appendix C-1 and Appendix C-2, respectively.

First, a baseline survey was administered in early March to establish students' perceptions of the grammar-translation

method they had experienced in the previous academic year. This survey assessed core constructs including overall satisfaction, perceived learning effects, motivation, and self-reported classroom disengagement (i.e., experience of sleeping in class).

Second, a post-intervention survey was administered at the end of the academic year to evaluate the new AI-supported teaching method. This survey included the same core constructs as the baseline survey to allow for direct comparison. It also incorporated four additional items specific to the intervention, assessing perceptions of autonomy in differentiated instruction, interest level, views on minimized grammar instruction, and impact on self-esteem. The inclusion of the self-esteem item was specifically informed by prior research highlighting its potential decrease in traditionally differentiated settings. Both surveys concluded with an open-ended question for students to provide overall qualitative feedback. Open-ended question responses were analyzed using inductive content analysis. The researcher read through all comments multiple times to identify key patterns and themes emerging directly from the data itself.

Furthermore, to complement the survey data, the researcher maintained observational records during the intervention year. This involved tracking the number of students observed sleeping during the AI-supported English classes. This observational data provided an objective counterpoint to the baseline survey item that specifically asked students to self-report ‘whether they had the experience of sleeping in class’ during the previous year’s grammar-translation method.

The limitations of the survey instrument are acknowledged. The survey was not formally pilot-tested, nor were its psychometric properties (e.g., reliability, construct validity) statistically validated. Furthermore, it is acknowledged that certain items, such as the question regarding students’ experience of sleeping in the previous year’s class, are subject to recall bias and could be perceived as leading. Separately, observational records of classroom engagement were maintained to contrast with the survey data. While unable to correct the survey’s inherent limitations, these records allowed for a direct comparison between students’ recalled past and their observed present. Additionally, to mitigate potential social desirability bias, the surveys were conducted anonymously after final exams were completed, and students were explicitly assured that their responses were confidential and would not affect their grades. Therefore, the survey data should be interpreted as exploratory insights into student perceptions.

Measures to Enhance Reliability and Validity

Several methodological measures were deliberately integrated into the research design to enhance the credibility and trustworthiness of the findings.

First, to strengthen the validity of the interpretations, the study employed methodological triangulation. Rather than relying on a single data source, the research collected and cross-referenced multiple forms of data. Quantitative data from mock test scores and comparative surveys were analyzed alongside qualitative data from the researcher’s direct observational records of student engagement and students’ open-ended written feedback. By synthesizing these different data sources, the study was able to develop a more holistic understanding of the intervention’s impact.

Second, the study’s prolonged engagement over a full academic year serves to increase the reliability of the findings. Unlike short-term interventions that may be influenced by a temporary ‘novelty effect’, this year-long implementation allowed for the observation of more stable patterns of change in both academic achievement and student perceptions. This extended timeframe ensures that the positive student feedback reflects a sustained response to the new pedagogical model rather than a fleeting reaction to a new activity.

Third, the research was conducted in an authentic setting—a real public high school classroom in a rural area—thereby strengthening its ecological validity. The findings are not derived from a controlled, artificial laboratory environment but from the complex, dynamic context of everyday school life. This enhances the practical relevance of the study, suggesting that the proposed teacher-AI collaborative model is viable for implementation in similar real-world educational settings.

RESULTS: OBSERVED TRENDS AND STUDENT PERCEPTIONS

Observed Trends in Academic Achievement

This section presents the data collected during the study. In line with the exploratory nature of this research, the findings are reported as observed trends and perceptions. As noted in the methodology section, inferential statistical analyses were not conducted due to the study’s design limitations; therefore, the data are presented descriptively.

Changes in Proportions of Proficiency Groups

An analysis of mock test scores reveals shifts in the distribution of proficiency groups among the participants over the academic year. As shown in Table 2, the proportion of students categorized as low-level decreased from 62% in March to 40% in September. Concurrently, the proportion of middle-level students increased from 21% to 37%, and the proportion of high-level students rose from 17% to 23%.

This pattern of change contrasts with the trends observed in the historical comparison data from 2023. For the same participants in their first-grade year, the proportion of low-level students actually increased over the year, from 56% to 64%. Similarly, for the previous cohort of second-grade students (third graders in 2024), the proportion of low-level students also saw a slight increase from 56% to 59% during their second-grade year. While direct causal claims are limited due to the study's design, the observed trend of an increasing proportion of high and middle-level students and a corresponding decrease in low-level students for the participant group in 2024 is distinct from the patterns seen in the comparison data.

TABLE 2
The Changes of Ratio in Three Groups

| Group | Level | March | June | September |
|---|--------|-------|------|-----------|
| Participant Group (Intervention Year: 2024) | High | 17% | 21% | 23% |
| | Middle | 21% | 27% | 37% |
| | Low | 62% | 52% | 40% |
| Comparison Group 1 (Participants in 2023) | High | 16% | 16% | 11% |
| | Middle | 28% | 26% | 24% |
| | Low | 56% | 58% | 64% |
| Comparison Group 2 (Previous Cohort in 2023) | High | 8% | - | 11% |
| | Middle | 38% | - | 30% |
| | Low | 56% | - | 59% |

Analysis of Changes within the High-Level Student Group

A closer examination of the students within the high-level category (ratings 2-4) also indicates specific shifts, as detailed in Table 3. The percentage of these high-level students achieving the highest rating (rating 2) increased, rising from 9% in March to 21% in September. Correspondingly, the proportions of students with rating 3 and rating 4 showed some fluctuation.

The data for June shows a temporary surge in rating 4 (85%) and a disappearance of rating 3. This anomaly might be attributed to students' unfamiliarity with the specific format of the mock test questions. To address this, a one-week instructional period was dedicated to familiarizing participants with appropriate test-taking strategies before the September mock test. Following this, the distribution in September showed a more balanced pattern, with an increase in the rating 2 group compared to March.

TABLE 3
The Changes of Ratio in High-Level Students

| Rating | March | June | September |
|--------|-------|------|-----------|
| 2 | 9% | 15% | 21% |
| 3 | 36% | 0% | 29% |
| 4 | 55% | 85% | 50% |

Comparative Analysis of Student Perceptions

This section presents a comparative analysis of student perceptions, drawing from survey data collected at the beginning and end of the academic year. A total of 63 students participated in both surveys. Responses were collected on a five-point

Likert scale and converted to a 100-point scale for analysis, where ‘strongly agree’ was assigned 100, ‘agree’ 80, ‘neutral’ 60, ‘disagree’ 40, and ‘strongly disagree’ 20. This conversion enabled the calculation of a mean score (MS) for each item. The findings reveal a stark contrast between student perceptions of the grammar-translation method and the new AI-supported method.

Comparative Survey Findings

A significant disparity was observed across all core constructs. Regarding classroom disengagement, 63% of students self-reported having the experience of sleeping in the previous year's grammar-translation class. In contrast, observational data from the AI-supported classes in 2024 indicated that instances of disengagement were minimal, with fewer than two students per class observed sleeping, typically due to severe fatigue.

This pattern was reflected in the survey data. For overall satisfaction, the mean score for the AI-supported method (MS = 91.8) was substantially higher than for the previous method (MS = 52.6). While 93.5% of students expressed satisfaction with the new method, only 22.4% reported satisfaction with the previous year's instruction. Similarly, for perceived learning effects, the mean score for the new method was 83.3, compared to 46.1 for the previous method. A majority of students (75.4%) felt their English ability improved under the new approach, whereas only 12.1% felt the same about the grammar-translation method. The disparity was also evident in motivation, with the AI-supported class receiving a mean score of 80.7, far exceeding the 42.6 recorded for the previous method.

Perceptions of the AI-Supported Method's Key Features

The post-intervention survey included items assessing the specific components of the new teaching method, which were rated overwhelmingly positively. The principle of student autonomy was exceptionally well-received, with the statement, ‘I think the teaching method in which students choose passages that match their ability is a good approach,’ earning a mean score of 93.1. The engaging nature of the content was also highly valued, as the statement, “I found this year's English class interesting,” received a mean score of 92.1. The pedagogical shift away from intensive grammar instruction was supported by a large majority of students (80.3%), resulting in a mean score of 86.9. Crucially, the method appeared successful in mitigating the potential negative affective consequences of differentiation. For the statement, “I think the teaching method of allowing students to choose passages that match their ability did not lower my self-esteem,” responses were positive across all proficiency levels, including a substantial majority (80%) of low-level students.

Qualitative Feedback from Open-Ended Responses

The qualitative data from the open-ended survey questions contextualized and reinforced the quantitative findings. Comments regarding the baseline survey on the grammar-translation method were predominantly negative, citing the tedious nature of grammar memorization and the perceived ineffectiveness of repetitive translation exercises. A high-level student noted that this approach put their “journey to learn English in jeopardy”, while others admitted they “often chose to fall asleep rather than participating.”

In stark contrast, feedback on the post-intervention survey was overwhelmingly positive. Several key themes emerged. First, students praised the autonomy and engaging content, with one low-level student writing, “Being able to choose a passage that fit my level was the best part,” and another commenting, “The topics were interesting, so I actually wanted to read them, which is a first for me in English class.” Second, students consistently made favorable comparisons to the previous year's instruction. A high-level student remarked, “Last year's class was just about memorizing grammar for tests, and it was boring. This year, I felt like I was actually using English to learn about interesting things.” Third, several comments provided direct evidence for the unintended scaffolding effect of the exam preparation process. A middle-level student noted, “I chose the middle-level passage in class, but I had to study the high-level one for the midterm and final exam. Because the story was fun, I enjoyed studying the harder version too, and I felt like my reading skills actually improved.” Finally, some students offered constructive criticism, particularly regarding the minimized grammar instruction. One middle-level student explained, “Since the high-level passage was also on the exam, I wanted to understand its grammar points more deeply. But the explanation was sometimes too brief to fully understand for students who did not choose a high-level passage.” This comment highlights a valid tension between the new method's approach and students' pragmatic needs for comprehensive test preparation.

TABLE 4
Students' Responses to the Grammar-Translation Method

| Statement | Yes(%) | | | | | |
|--|--------|-------|-------|-------|--------|------|
| I had the experience of sleeping in last year's English class. | 63% | | | | | |
| Statement | SA (%) | A (%) | N (%) | D (%) | SD (%) | MS |
| I was generally satisfied with last year's English class. | 8.6 | 13.8 | 36.2 | 15.5 | 25.9 | 52.6 |
| My English ability was improved through last year's English class. | 5.2 | 6.9 | 34.5 | 20.7 | 32.8 | 46.1 |
| Last year's English class motivated me to study English. | 3.4 | 5.2 | 36.2 | 12.1 | 43.1 | 42.6 |

Note: Based on responses from N = 63 students. SA = Strongly Agree (100), A = Agree (80), N = Neutral (60), D = Disagree (40), SD = Strongly Disagree (20), MS = Mean Score. Mean score calculated based on the assigned values.

TABLE 5
Students' Responses to the New Teaching Method

| Statement | SA (%) | A (%) | N (%) | D (%) | SD (%) | MS |
|---|--------|---------------------------|-------|-------|--------|------|
| I was generally satisfied with this year's English class. | 65.6 | 27.9 | 6.6 | - | - | 91.8 |
| | | for high-level students | | | | 95.7 |
| | | for middle-level students | | | | 92.7 |
| My English ability was improved through this year's English class. | | for low-level students | | | | 88.8 |
| | 45.9 | 29.5 | 19.7 | 4.9 | - | 83.3 |
| | | for high-level students | | | | 92.9 |
| This year's English class motivated me to study English. | | for middle-level students | | | | 88.1 |
| | | for low-level students | | | | 73.6 |
| | 41 | 26.2 | 27.9 | 4.9 | - | 80.7 |
| I think the teaching method in which students choose passages that match their ability is a good approach. | | for high-level students | | | | 92.9 |
| | | for middle-level students | | | | 83.6 |
| | | for low-level students | | | | 71.2 |
| I found this year's English class interesting. | 75.4 | 16.4 | 6.6 | 1.6 | - | 93.1 |
| | | for high-level students | | | | 97.1 |
| | | for middle-level students | | | | 92.7 |
| I liked that this year's English class minimized grammar points, focusing only on what was necessary for reading. | | for low-level students | | | | 91.2 |
| | 68.9 | 23 | 8.2 | - | - | 92.1 |
| | | for high-level students | | | | 94.3 |
| | | for middle-level students | | | | 93.6 |
| | | for low-level students | | | | 89.6 |
| | 54.1 | 26.2 | 19.7 | - | - | 86.9 |
| | | for high-level students | | | | 92.9 |
| | | for middle-level students | | | | 90.1 |
| | | for low-level students | | | | 80.8 |

| Statement | SA (%) | A (%) | N (%) | D (%) | SD (%) | MS |
|--|--------|-------|-------|-------|--------|------|
| | | | | | | 90.3 |
| | | | | | | 97.3 |
| I think the teaching method of allowing students to choose passages that match their ability did not lower my self-esteem. | 86.7 | 13.3 | - | - | - | 94.5 |
| | 77.2 | 18.1 | 4 | - | - | 82.4 |
| | 32 | 48 | 20 | - | - | |
| | | | | | | |

Note: Based on responses from N = 63 students. SA = Strongly Agree (100), A = Agree (80), N = Neutral (60), D = Disagree (40), SD = Strongly Disagree (20), MS = Mean Score. Mean score calculated based on the assigned values.

DISCUSSION

Interpretation of Key Findings Through Theoretical Frameworks

The findings of this year-long exploratory study are twofold: 1) a positive descriptive trend in students' academic achievement, and 2) a dramatic and overwhelmingly positive shift in student perceptions of their English classes when compared to the grammar-translation method. While the study's design limitations preclude causal claims, these findings can be interpreted through the extensive theoretical and empirical frameworks established in the literature review, offering significant insights into the potential of AI-supported, student-centered reading instruction.

AI-Personalized Learning: Validating and Extending Prior Research

This study's findings strongly align with and extend the existing literature on AI in language education. The exceptionally high levels of student satisfaction (MS = 91.8), interest (MS = 92.1), and motivation (MS = 80.7) provide robust, long-term evidence for what previous short-term studies have suggested: AI-generated texts based on learners' personal interests can significantly boost affective factors like reading enjoyment and motivation (J. H. Lee et al., 2023; Yılmaz & Aydın, 2025). Furthermore, the observed decrease in the proportion of low-level students and the corresponding increase in middle and high-level students support the findings of Sudin and Swanto (2024) and Çelik et al. (2024), who argued that AI-driven personalization yields the most significant cognitive gains for low to intermediate learners by providing comprehensible input.

Crucially, this study offers a practical, year-long case study of the 'teacher-AI collaborative model' advocated in the literature (Sun & Yusof, 2025; Zhang, 2025). The instructor's role as a 'critical curator' — selecting topics, adapting AI-generated texts, and providing pedagogical support—was central to the intervention's success. This confirms that AI may be most effective not as a replacement for the teacher, but as a powerful tool that augments the teacher's capacity to implement pedagogically sound, personalized instruction.

A Successful Operationalization of Motivational Theories

The dramatic improvement in student perceptions can be directly interpreted through the lenses of Self-Determination Theory (SDT) and Interest Theory. The intervention was explicitly designed to support the three basic psychological needs central to SDT (Ryan & Deci, 2000): autonomy was fostered by allowing students to choose their own texts; competence was supported by providing level-appropriate materials that made successful reading achievable; and relatedness was encouraged through collaborative work with peers. The fulfillment of these needs, in stark contrast to the passive nature of the grammar-translation method, likely contributed to the significant increase in students' intrinsic motivation.

Similarly, the use of engaging and personally relevant topics (e.g., humorous stories about classmates, contemporary news) appears to have successfully sparked the initial situational interest necessary to draw students into the learning process (Hidi & Renninger, 2006). This approach addresses a key limitation of previous authentic material interventions, such as using

newspapers, which could be interesting but failed to cater to diverse proficiency levels (H. Kim et al., 2010; Jung & Jang, 2017). By leveraging AI, this study successfully personalized for both interest and proficiency simultaneously, thereby supporting both situational interest and the need for competence.

A New Model for Differentiation: Overcoming Traditional Pitfalls

Perhaps the most significant contribution of this study is the demonstration of a viable model for differentiation that overcomes the well-documented pitfalls of traditional methods. As highlighted in the literature review, the practical implementation of differentiation has long been hindered by the ‘unsustainable workload’ it places on teachers (J. H. Lee et al., 2023; Yılmaz & Aydın, 2025). This study serves as a direct case for how AI can solve this problem, empowering the teacher to consistently produce high-quality, multi-level materials.

More importantly, the study's model of student-driven differentiation within a single classroom addresses the critical issue of negative affective consequences. Prior research on physical ability grouping in the Korean context showed that assigning students to lower-level classes can harm their self-esteem (E. Kim & Rha, 2015). The finding that 80% of low-level students in this study affirmed their self-esteem was not negatively affected is particularly noteworthy. This suggests that by replacing teacher-assigned grouping with student autonomy—a method shown to be effective in prior research (J. Lee & Jang, 2014)—it is possible to reap the academic benefits of differentiation while mitigating its potential for social-emotional harm.

Unintended Scaffolding through Exam Preparation

Beyond the direct motivational aspects of the in-class activities, the observed positive trend in academic achievement may be partially explained by a synergy with an unintended factor: the structure of out-of-class exam preparation. This interpretation is directly supported by the qualitative data presented in the section on qualitative feedback, where students explicitly commented on this phenomenon. The requirement to study all three proficiency-level passages for exams created what students perceived as a form of self-paced scaffolding. As one middle-level student reported, the engaging nature of the content motivated them to tackle the more difficult high-level passage, leading to a perceived improvement in their reading comprehension. While this finding is based on qualitative self-reports, it offers a plausible, student-grounded explanation for the observed academic trends that warrants further, more systematic investigation.

Pedagogical Implications

While exploratory, the findings of this study offer three key pedagogical implications for educators seeking to integrate AI into secondary English reading instruction.

First, while the intervention showed a positive overall trend in academic achievement, it also highlighted that AI-driven personalization is not a panacea, particularly for the lowest-proficiency learners. This group consistently reported the lowest scores in perceived improvement and motivation, suggesting that engaging, topic-personalized texts may be insufficient for students who lack foundational language skills. This implies that for AI-enhanced pedagogy to be truly effective for all learners, it must go beyond content personalization to include more fundamental, targeted skill-building support. Future AI tools could, for example, integrate adaptive exercises that target specific grammatical weaknesses or vocabulary gaps identified in a student's reading process.

Second, this study highlights a clear and urgent need for a more sophisticated diagnostic tool. The limitations of adjusting content difficulty based solely on a teacher's informal observations became apparent during the intervention. For instance, the teacher significantly increased passage difficulty after observing that students seemed to be improving based on their in-class performance and written answers. However, this sharp adjustment resulted in many students expressing difficulty with the new materials. Therefore, an educational platform that can accurately assess students' current proficiency levels and then use that data to generate appropriately matched reading passages is required to effectively personalize learning.

Third, the study reveals a practical tension between the new method's pedagogical focus and the pragmatic demands of assessment. Student feedback indicates a desire for more in-depth grammar explanations for high-level passages. This highlights a significant challenge in preparing for midterm and final exams. This implies that instructors should consider providing supplementary and more thorough grammar instruction. This is especially important for the most advanced texts when an exam includes all proficiency levels. This could be achieved through optional video resources, written guides, or structured peer-tutoring activities, ensuring all students have adequate support for test preparation.

CONCLUSION

This year-long exploratory case study investigated the implementation of an AI-supported, personalized reading approach in a Korean high school English class, aiming to address persistent issues of low student motivation and engagement associated with traditional grammar-translation methods. The study sought to understand the resulting patterns in academic achievement and student perceptions. The findings revealed two principal outcomes. First, there was a positive descriptive trend in academic achievement, with a decrease in low-proficiency students and an increase in middle and high-proficiency students. Second, students' perceptions of their English learning experience showed a profound and overwhelmingly positive shift. Students reported dramatically higher levels of satisfaction, interest, and motivation compared to the previous year, attributing this change to the autonomy of choosing level-appropriate texts and the engaging, personalized content.

This study makes several contributions to the literature on AI in language education. It addresses a notable gap by providing a rare, long-term, classroom-based investigation within a K-12 context, extending a field of research largely dominated by short-term interventions or studies in higher education. Furthermore, it presents a practical and effective 'teacher-AI collaborative model,' where the teacher acts as a critical curator, leveraging AI to overcome the long-standing challenge of implementing differentiated instruction without an unsustainable workload. The approach successfully operationalized key motivational principles from Self-Determination Theory and Interest Theory, demonstrating a viable method for differentiation that mitigates the potential for harm to student self-esteem often associated with traditional ability grouping.

Despite these promising findings, the study's limitations must be acknowledged. First, as a single-case study conducted in a public high school in a rural area, the findings possess limited generalizability and may not be applicable to different educational contexts, such as urban schools with more diverse student populations. Second, the methodological design was exploratory and descriptive, not experimental. The absence of a concurrent control group means that no causal claims can be made regarding the intervention's impact on observed academic achievement, as confounding variables such as student maturation or variations in test difficulty were not controlled. The reliance on a researcher-developed, non-validated survey, which included some potentially leading questions, and the dual role of the researcher as the teacher also introduce potential for bias. Furthermore, despite the study being conducted for a year, the possibility that the 'novelty effect' of the new method contributed to the positive perceptual changes in the early stages, in addition to the effect of the teaching method itself, cannot be ruled out.

Based on these findings and limitations, several avenues for future research are recommended. Future studies should employ quasi-experimental or experimental designs with control groups to establish a clearer causal link between this pedagogical model and academic outcomes. Replicating this study in a variety of educational settings would also be crucial to assess the generalizability of its findings. Further research could also systematically investigate the 'unintended scaffolding' effect observed during exam preparation to understand how it can be intentionally integrated into pedagogical design.

In conclusion, this study provides compelling evidence that a teacher-led, AI-supported pedagogical approach can transform the secondary English reading classroom from a passive, test-oriented environment into an active, engaging, and motivating learning space. While acknowledging its exploratory nature, this research offers valuable practical insights and a promising model for educators seeking to harness technology to foster genuine student interest and support diverse learning needs.

References

- Ahn, Ae-Rang, & Park, Eun-Soo. (2014). Effects of level-based English class on low-level middle school students' affective domain. *Journal of Learner-Centered Curriculum and Instruction*, 14(9), 47-64.
- Bektik, D., Ullmann, T. D., Edwards, C., Herodotou, C., & Whitelock, D. (2024). AI-powered curricula: Unpacking the potential and progress of generative technologies in education. *Ubiquity Proceedings*, 4(1), Article 38.
- Çelik, F., Yangın Ersanlı, C., & Arslanbay, G. (2024). Does AI simplification of authentic blog texts improve reading comprehension, inferencing, and anxiety? A one-shot intervention in Turkish EFL context. *International Review of Research in Open and Distributed Learning*, 25(3), 287-303.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111-127.
- Hong, Eun-Hwa, & Kim, Kyung Ja. (2014). High school students' changes in English achievement, interest, and self-confidence in level-differentiated English classes. *Studies in Foreign Language Education*, 28(2), 233-254.
- Jung, Ji-Sun, & Jang, Jong-Duk. (2017). A study on the efficacy of English newspapers to improve students' reading comprehension.

Studies in British and American Language and Literature, 126, 215-237.

- Kang, Bo-ram, & Cho, Yunkyoung. (2012). Effects of level-based English instructions in middle school on students' academic achievements. *Modern English Education*, 13(2), 101-120.
- Kim, Damee, & Kim, Tae-Young. (2018). The structural relationship between Korean high school students' resilience, English learning motivation, English class attitude, and English learning achievement. *Modern English Education*, 19 (1), 120-136.
- Kim, Eun Ju, & Rha, Kyeong-Hee. (2015). The effect of level-differentiated English classes on Korean middle school students' English achievement. *Journal of the English Language and Literature*, 57(1), 131-156.
- Kim, Hye-Ju, Im, Byung-Bin, Park, Jimin, & Mun, Chang-Sik. (2010). Using English newspapers in high school English reading classes. *English Language & Literature Teaching*, 16(3), 183-206.
- Kim, Kyung-Hoon. (2011). Improving English reading competence for middle school students through newspapers in education. *Journal of the Korea Contents Association*, 11(4), 477-484.
- Korean Institute for Curriculum and Evaluation. (2016). *2018 Hangnyeondo Suneung Yeongeo Jeoldaepyeongga Hakseubannaee*.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Lee, Ji-Eun, & Jang, Jong-Duk. (2014). Efficiency of level-based class on high school students' academic achievement. *The Journal of Linguistics Science*, 71, 197-216.
- Lee, Jang Ho, Shin, Dongkwang, & Noh, Wonjun. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, 54(2), 508-516.
- Lim, Jeongwan. (2014). A study of high school students' and teachers' perceptions towards level-differentiated English classes. *Journal of the Korea English Education Society*, 13(1), 111-133.
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11(1), Article 50.
- Park, Ha-eun, Hong, Chanlan, Choi, Hana, Hwi, Ho, & Larsen, C. G. (2014). Affective filter in second language learning in South Korea. *International Journal of Foreign Studies*, 7(2), 43-58.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78.
- Shim, Jihyun, & Lim, Hyun-Woo. (2017). Exploring students' and teachers' perceptions of the effects of English ability grouping teaching. *Secondary Education Research*, 65(4), 883-912.
- Snashall, P. (2023). Empowering ESL Teachers with generative AI. *Proceedings of the 10th National and International Conference on "Research to Serve Society"*, Huachiew Chalermprakiet University, 243-252.
- Sudin, V. D. K. A., & Swanto, S. (2024). Enhancing ESL learners reading comprehension and motivation via AI-generated personalised reading texts. *International Journal of E-Learning Practices*, 7(1), 1-11.
- Sun, X., & Yusof, S. M. (2025). AI-supported differentiated instruction for Chinese vocational English reading: Enhancing proficiency and motivation. *International Journal of High Speed Electronics and Systems*, 2540649.
- Tessensohn, T. C., Yunus, M. M., & Ismail, H. H. (2025). Using AI-powered tools in enhancing reading skills in the ESL classroom: A systematic review (2020-2024). *International Journal of Academic Research in Progressive Education and Development*, 14(2), 57-70.
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard University Press.
- Xiao, C., Xu, S. X., Zhang, K., Wang, Y., & Xia, L. (2023). Evaluating reading comprehension exercises generated by LLMs: A showcase of ChatGPT in education applications. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics, 610-625.
- Yılmaz, Ö. K., & Aydın, S. (2025). The impact of the use of artificial intelligence-generated materials on reading motivation among EFL Learners. *Reading Research Quarterly*, 60(3), e70016.
- Zhang, L. (2025). Exploring differentiated instruction in college English teaching in the context of ChatGPT. *Pacific International Journal*, 8(2), 154-158.
- Zheng, Y. D., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Computers and Education: Artificial Intelligence*, 6, Article 100223.

Appendix A

Explanations of each English class

| No. | Time of Use | Topic | Source | Implementation Details |
|-----|-----------------------|---|-----------|--|
| 1 | April (Week 1) | The future outlook for Tesla sales | WSJ | Used drag race video (EV vs. gasoline car) for situational interest; activated background knowledge with sales trend data. |
| 2 | April (Week 1) | Apple's monopoly issue | WSJ | Adjusted difficulty of low-level passage downward based on student feedback. |
| 3 | April (Week 2) | The new Disney CEO's strategy to increase profits | WSJ | Provided additional foundational-level passages for struggling students. |
| 4 | April (Week 2) | Barack Obama's speech at the 2004 DNC convention | TAPP | Created level-appropriate passages based on proficiency identified over 3 sessions; conducted Jigsaw activity. |
| 5 | April (Week 3) | The story of Student A who becomes a billionaire with the help of Elon Musk | TGN | Used original story featuring student volunteers as main characters. |
| 6 | April (Week 3) | How the Russian military is using Starlink | WSJ | Utilized WSJ article reflecting student interest in Starlink. |
| 7 | May (Week 1) | The story of Student B who becomes the greatest father in the world | TGN | Created story based on student ideas following enthusiastic response to Lesson 5. |
| 8 | May (Week 2) | The effectiveness of a double major in college | WSJ | Used relevant WSJ article to address student concerns about choosing a major. |
| 9 | May (Week 3) | Apple's plan to build its own electric vehicle | WSJ | Provided direct, explicit instruction for students who chose foundational-level texts instead of collaborative reading. |
| 10 | May (Week 3) | The large wave of immigration impacting U.S. classrooms | WSJ | Increased passage complexity with longer sentences based on students' perceived proficiency improvement. |
| 11 | May (Week 4) | New EU laws and high chocolate prices | WSJ | Selected article requiring deeper analysis to enhance reading comprehension. |
| 12 | June (Week 2) | Why Apple fell behind in the AI race | WSJ | Based on original articles, replacing difficult vocabulary with simpler alternatives for high-level passages. |
| 13 | June (Week 2) | The story of Student C who helped Steve Jobs establish Apple | TGN | Developed story based on student's funny episode, including peers as characters to boost motivation. |
| 14 | June (Week 3) | Conan O'Brien's 2011 Dartmouth College Commencement Speech | Dartmouth | Used YouTube summary of a speech to inspire academic motivation and provide background knowledge. |
| 15 | June (Week 3) | Why Nike fell behind in innovation | WSJ | Engaged students with a YouTube video comparing popular shoe brands before reading activity. |
| 16 | August (Week 3) | The new CEO of Starbucks | WSJ | Generated interest with stock chart and teacher's humorous anecdote; explained new CEO's strengths. |
| 17 | August (Week 3) | How Microsoft is building an AI empire | WSJ | Significantly raised difficulty of mid/high-level passages due to increased reading speed; students found it challenging. |
| 18 | September (Week 1) | What fentanyl is and how to prevent overdoses | WSJ | Used relevant news articles to educate on the seriousness of the drug problem. |
| 19 | September (Week 2) | How deepfake technology is used to create the illusion of Disney characters smoking | WSJ | Used news articles for deepfake education as requested; engaged students with visual of teacher's superimposed face. |

| No. | Time of Use | Topic | Source | Implementation Details |
|-----|----------------------|---|--------|---|
| 20 | October (Week 1) | Why Boeing's stock price keeps falling | WSJ | Provided background knowledge with competitor comparison video; linked to safety education with incident video. |
| 21 | October (Week 2) | The various difficulties that Volkswagen is experiencing | WSJ | Conducted creative problem-solving activity after watching video explaining a company's struggles. |
| 22 | October (Week 5) | The story of Students D, E, F, and G who built the world's greatest strawberry farm | TGN | Compiled funny school episodes into a single story. |
| 23 | November (Week 1) | The Israel-Hamas war | WSJ | Linked content to school violence prevention after watching a video explaining the Israel-Hamas war. |
| 24 | November (Week 1) | SpaceX's new rocket launch | WSJ | Selected topic based on news coverage; used comparison videos of old vs. new technology to engage students. |
| 25 | November (Week 2) | The story of Student H who triumphed in the Black and White Cooking Contest | TGN | Created a story about a student winning a competition, based on a popular Netflix show. |
| 26 | November (Week 3) | Tesla's Robotaxi event | WSJ | Selected topic based on extensive news coverage; used event highlights video before reading activity. |
| 27 | November (Week 3) | The various difficulties that Samsung is experiencing | WSJ | Addressed some students' investment concerns with a relevant WSJ article; provided brief financial literacy lesson. |
| 28 | November (Week 4) | Why Intel's share price has not recovered | WSJ | Introduced company's struggles to capture interest; used videos on history and reasons for decline. |
| 29 | November (Week 4) | The revival of AMD and its CEO Lisa Su | WSJ | Engaged students by presenting a contrasting success story; used videos on the CEO and company comeback. |
| 30 | December (Week 1) | NEOM City Project: The Line | WSJ | Reflected some students' desired major (architecture) by using WSJ article on a large-scale construction project. |
| 31 | December (Week 1) | The story of Student I who transformed into a rat. | TGN | Created a fun story based on a student's episode, using a pun (play on homonyms in Korean). |

Note: WSJ: The Wall Street Journal, TAPP: The American Presidency Project, TGN: Teacher-Generated Narrative All topics listed above were adapted into three distinct reading passages (low, middle, and high-level) using AI tools, as described in the Methodology section. Students autonomously selected one of the three passages in class.

Appendix B

Sample of reading passages and comprehension questions

Topic: NEOM City Project: The Line

[Low-level reading passage]

Have you heard of a city called The Line? It's a very special city being built in Saudi Arabia, a country far away in the desert. This city is not like other cities—it is long and straight, like a pencil line. That's why it's called The Line. The Line will be 170 kilometers long. But it will only be 200 meters wide, which is not very wide at all. Even though it is long and thin, it will have everything a city needs—homes, schools, parks, and even a place to play.

What makes The Line very special is that it's built to be friendly to nature. There will be no cars or roads. Instead, people will walk or use fast trains that go from one end of the city to the other in just 20 minutes. This way, the air will stay clean, and the animals nearby can live safely.

The buildings in The Line will be very tall, like skyscrapers, and they will have shiny walls made of mirrors. These mirrors will make the city look amazing because it will reflect the sunlight and desert around it.

The people who designed The Line want it to be a place where everyone can live happily, with clean air, green trees, and smart technology. So, The Line is not just a city—it's a dream of the future, where people and nature live together peacefully.

[Low-level comprehension questions]

왜 이 도시의 이름이 “더 라인(The Line)”인가요?

더 라인의 길이와 폭은 각각 얼마나 되나요?

더 라인이 자연친화적인 이유는 무엇인가요?

더 라인의 건물은 어떤 독특한 특징을 가지고 있나요?

더 라인의 설계자들이 이루고자 하는 주요 목표는 무엇인가요?

[Middle-level reading passage]

The giant future city of Neom is causing problems for Saudi Arabia. Costs are rising, plans are delayed, and recently, the world's largest building project replaced its leader of six years.

In the background, Saudi Arabia is facing money problems. Even though the country is known for being rich, it can't pay for all the fancy projects and ideas linked to Vision 2030, a plan by Crown Prince Mohammed bin Salman to move away from depending on oil.

Inside Neom, some people say money has been wasted on unnecessary work. At The Line—a pair of skyscrapers planned to stretch farther than the distance from New York to Philadelphia—builders dug out 60 miles of sand. This was done even though the first phase of the project was only 10 miles and was later reduced to 1.5 miles.

Neom is not the only project facing challenges. Other projects in Saudi Arabia also need huge amounts of money as they move from the cheap early stages to the more expensive building phases.

Crown Prince Mohammed bin Salman introduced Vision 2030 in 2016 to change Saudi Arabia's economy within 14 years. Money has been spent on making the country a leader in sports, video games, and electric cars. The government has also encouraged more people to work in private companies, which made up 45% of the economy in 2023, compared to 40% in 2016.

Real estate projects are a big part of Vision 2030. Neom is the largest one, with plans for a car-free, vertical city in the desert for 1 million people by 2030—and 9 million later. The original cost was \$500 billion, but some former leaders said it might actually cost trillions.

Other big projects include a huge entertainment district near Riyadh with theme parks, the region's largest water park, and the world's tallest roller coaster. There's also a \$30 billion group of 50 resorts on the Red Sea and a massive Riyadh project centered around a giant cube building.

The list of projects keeps growing. In the last year, Saudi Arabia announced a \$100 billion plan for advanced electronics, new investments in artificial intelligence, and 11 futuristic stadiums for a World Cup bid, including one on top of a skyscraper at Neom.

Altogether, these projects could cost trillions of dollars if fully built—much more than the country's \$1 trillion wealth

fund, which includes investments that are hard to sell.

Saudi Arabia still has enough money for daily services and keeps debt low. But falling oil prices are a big problem. Oil sales make up half of Saudi Arabia's income, and oil prices are far below what the country needs to balance its budget. Saudi officials recently said there will be budget shortages in the coming years instead of surpluses.

[Middle-level comprehension questions]

사우디아라비아가 네옴과 비전 2030 프로젝트를 진행하면서 재정적 어려움을 겪는 주요 이유는 무엇인가요?

네옴의 더 라인 건설 초기 단계에서 어떤 변경 사항이 있었으며, 그것이 중요한 이유는 무엇인가요?

본문에서 설명된 사우디아라비아 비전 2030 계획의 주요 목표는 무엇인가요?

사우디아라비아의 현재 석유 수입 의존도가 비전 2030 프로젝트 자금 조달에 어떤 영향을 미치나요?

본문에서 언급된 네옴 외에 다른 주요 프로젝트 두 개를 이름과 특징과 함께 설명해주세요.

본문에 따르면, 사우디아라비아가 예산을 균형 있게 유지하는 데 있어 어떤 재정적 어려움에 직면하고 있나요?

[High-level reading passage]

Defying doubters, Saudi Arabia is moving ahead with hundreds of billions of dollars in projects at Neom, a completely new region the size of Massachusetts, featuring futuristic buildings, a dry ski resort, and a long list of impressive projects meant to attract a population bigger than New York City's.

None is more daring than a multitrillion-dollar pair of skyscrapers taller than the Empire State Building designed to stretch 105 miles and house nine million people, the main project called "The Line." Its supporter, Saudi Crown Prince and leader Mohammed bin Salman, has compared the project to Egypt's Great Pyramids.

The kingdom in recent months reduced the Line's first phase, facing the reality of costs at a time the country is spending much more than it is earning. Now organizers plan to start by building around 1.5 miles of the structure by 2030, rather than the roughly 10-mile first section that had earlier been planned, multiple people familiar with the plans said. Still, even that shorter section would be by far the world's largest building, equal to more than 60 Empire State Buildings in size.

The stakes for Saudi Arabia are as big as Mohammed's ambition. Neom is the ultimate symbol of his plans to change the kingdom's economy, reduce its reliance on oil money, and make it a center for money and talent from around the world. But he risks wasting much of the country's money on an entirely new kind of city-building that could turn out to be too hard to complete.

A large number of problems lie ahead. More than 100,000 extra construction workers must be housed in a remote part of the kingdom's vast desert, two hours' drive from any big city. Neom's needs for steel, exterior glass, and other materials are so massive they may push up global prices and be hard to get. Planners worry the special idea of the Line, a vertical city housed in twin skyscrapers the length of Delaware, might turn out to be an unattractive place to live.

At the same time, the reduced plans for the Line bring attention to Neom's huge costs for what is now expected to be a medium-sized city. The costs keep going up. The predicted cost of a ski resort in the region's dry mountains has more than doubled over two years to \$38 billion as of October, according to Neom documents seen by The Wall Street Journal. Real estate adviser Knight Frank estimates more than \$237 billion worth of building contracts have already been signed at Neom.

People working on the project say that the number is far too low. The first 1.5 miles of the Line alone is expected to cost more than \$100 billion, two people familiar with the plans said.

If it were fully built, Neom employees expect the real cost of the Line would be well over \$2 trillion. Building costs per square foot are more than double the usual amount for other tall buildings in the Middle East, they said.

This makes it unlikely Neom will get a lot of private funding to pay for future parts of the Line, they say. So far, it has been paid for by the Saudi government.

Seven years after starting, little has been finished other than Neom's film studios and a large new royal area that includes giant palaces, a golf course, and at least 10 helipads, satellite images show.

Architects soon decided the first phase should be built somewhere else, leaving the Line's first foundations abandoned for now, said people familiar with the matter.

For over a year, most of the work has been digging—the world's biggest, Neom says. Four-lane temporary construction roads are crowded with lines of dump trucks; diesel smoke from trucks and generators fills the air.

Once foundations are done, a big test will be if and when Neom signs the expensive contracts to start vertical construction—a key moment that makes it harder to stop the project.

Another question is height. Many people working on Neom have questioned the need for a 1,600-foot-tall building—

which comes with extra engineering problems, higher costs, and makes evacuation harder in an emergency.

[High-level comprehension questions]

What is the main goal of Saudi Arabia's Neom project?

- a) To build the world's tallest skyscraper
- b) To reduce the kingdom's reliance on oil and attract global talent
- c) To create a desert ski resort
- d) To compete with Egypt's Great Pyramids

What challenges are mentioned regarding the construction of Neom?

- a) Limited interest from international investors
- b) Difficulty sourcing massive amounts of materials like steel and glass
- c) Shortage of construction workers
- d) All of the above

What makes “The Line” an unusual and ambitious project compared to traditional city-building?

Why might the Neom project struggle to attract private funding for future developments?

Describe two specific reasons why the projected costs of Neom are significantly higher than usual construction projects.

Appendix C-1

Survey questionnaire items on the grammar-translation method

나는 작년 영어 수업시간에 잔 경험이 있다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

나는 작년 영어 수업에 전반적으로 만족한다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

작년 영어 수업을 통해 나의 영어 실력은 향상되었다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

작년 영어 수업은 내가 영어공부를 하도록 동기를 부여했다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

Appendix C-2

Survey questionnaire items on the new teaching method

나는 올해 영어 수업에 전반적으로 만족한다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

올해 영어 수업을 통해 나의 영어 실력은 향상되었다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

올해 영어 수업은 내가 영어공부를 하도록 동기를 부여했다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

자신의 수준에 맞는 지문을 스스로 선택하는 수업 방식이 좋다고 생각한다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

나는 올해 영어 수업이 재미있었다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

올해 영어 수업에서, 문법 설명을 최소화해서 독해에 필요한 문법만 배우는 방식이 좋다고 생각한다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다

자신의 수준에 맞는 지문을 스스로 선택하는 수업 방식이 나의 자존감을 낮추지 않았다고 생각한다

1. 매우 그렇다 2. 그렇다 3. 보통이다 4. 아니다 5. 매우 아니다