Enhancing Vocabulary and Spelling Accuracy through AI-Generated Stories by Integrating Dictation Activities for A2-Level EFL Learners

Marcos Andrés Saltos Mendoza

Dirección de Posgrado, Cooperación y Relaciones Internacionales. Universidad Laica Eloy Alfaro de Manabí. Trabajo de Titulación, presentado como requisito para la obtención del grado de Magíster en Pedagogía de los Idiomas Nacionales y Extranjeros Mención Inglés

Director: Mg. Gabriel Basurto

Agosto, 2026

Enhancing Vocabulary and Spelling Accuracy through AI-Generated Stories

Autores:
Marcos Saltos Mendoza
Universidad Laica Eloy Alfaro de Manabí
Estudiante de la Maestría en Educación con mención en Innovaciones Pedagógicas
Manta – Ecuador
marcossaltos1116@gmail.com
https://orcid.org/0009-0009-7820-6109

Eder Intriago-Palacios
Universidad Laica Eloy Alfaro de Manabí
Mentor in the Program Maestría en Pedagogía de los Idiomas Nacionales y Extranjeros Mención Inglés.
Kansas – United States of America
intriago@ksu.edu
https://orcid.org/0000-0002-9433-7186

Abstract

This study evaluated the pedagogical and affective impact of incorporating artificial intelligence (AI)-generated contextualized stories into dictation activities for secondary English as a Foreign Language (EFL) learners. Traditional dictation exercises often use static, decontextualized texts that lead to learner monotony, persistent orthographic difficulties, and minimal engagement. To address this challenge, an 8-week mixed-methods intervention was conducted with 54 CEFR A2-level students at a bilingual institution in Ecuador. Participants completed weekly dictations using level-adapted narratives generated Gemini with specific syntactic constraints and embedded target vocabulary. Quantitative pre- and post-tests evaluated orthographic accuracy and target word mastery, while qualitative focus groups and a post-intervention survey explored student motivation and engagement. Quantitative results demonstrated substantial linguistic gains: orthographic accuracy increased by an absolute mean of +17.7%, and target word composite mastery scores rose from 10.80 to 16.75 out of 20 points (+5.95 gain) across receptive and productive dimensions. Survey findings revealed high student engagement across behavioral (M=4.35), affective (M=4.28), and cognitive (M=4.12) domains. Thematic analysis identified narrative curiosity and immediate self-correction feedback as key driving factors for motivation, while acoustic delivery speed and cognitive writing fatigue emerged as primary hindering factors. The findings demonstrate that AI-generated story dictations effectively improve traditional dictation by enhancing orthographic precision and target vocabulary retention while maintaining high learner engagement when supported by appropriate instructional scaffolding.

Keywords: AI-generated stories, dictation activities, artificial intelligence (AI), spelling accuracy, vocabulary acquisition.

Resumen

Este estudio evaluó el impacto pedagógico y afectivo de incorporar relatos contextualizados generados por inteligencia artificial (IA) en las actividades de dictado para alumnos de inglés como lengua extranjera (EFL) de secundaria. Los ejercicios tradicionales de dictado suelen utilizar textos estáticos y descontextualizados que provocan monotonía en los alumnos, dificultades ortográficas persistentes y una participación mínima. Para abordar este reto, se llevó a cabo una intervención de 8 semanas con métodos mixtos en la que participaron 54 alumnos de nivel A2 del MCER en un centro bilingüe de Ecuador. Los participantes realizaron dictados semanales utilizando narraciones adaptadas a su nivel, generadas por Gemini con restricciones sintácticas específicas y vocabulario objetivo integrado. Las pruebas cuantitativas previas y posteriores evaluaron la precisión ortográfica y el dominio del vocabulario objetivo, mientras que los grupos focales cualitativos y una encuesta posterior a la intervención exploraron la motivación y la participación de los alumnos. Los resultados cuantitativos demostraron avances lingüísticos sustanciales: la precisión ortográfica aumentó en una media absoluta del +17,7 %, y las puntuaciones compuestas de dominio de las palabras objetivo subieron de 10,80 a 16,75 sobre 20 puntos (+5,95 de ganancia) en las dimensiones receptiva y productiva. Los resultados de la encuesta revelaron un alto nivel de implicación de los alumnos en los ámbitos conductual (M = 4,35), afectivo (M = 4,28) y cognitivo (M = 4,12). El análisis temático identificó la curiosidad narrativa y la retroalimentación inmediata de autocorrección como factores clave para la motivación, mientras que la velocidad de lectura acústica y la fatiga cognitiva al escribir se revelaron como los principales factores que la obstaculizaban. Los resultados demuestran que los dictados de historias generados por IA mejoran eficazmente el dictado tradicional al potenciar la precisión ortográfica y la retención del vocabulario objetivo, al tiempo que mantienen un alto nivel de implicación de los alumnos cuando se apoyan en un andamiaje didáctico adecuado.

Palabras clave: relatos generados por IA, actividades de dictado, inteligencia artificial (IA), precisión ortográfica, adquisición de vocabulario**.**

Enhancing Vocabulary and Spelling Accuracy through AI-Generated Stories by Integrating Dictation Activities for A2Level EFL Learners

In recent years, the integration of artificial intelligence (AI) into education has expanded rapidly, transforming the design, adaptation, and delivery of language learning materials (Tolstykh & Oshchepkova, 2024). In English as a Foreign Language (EFL) contexts, AI tools—particularly Large Language Models (LLMs)—offer educators innovative ways to create contextually rich, level-appropriate resources tailored to student needs (Lee et al., 2025). Within foundational language development, developing strong orthographic accuracy and target word mastery is important for A2-level learners as they transition toward written fluency. Dictation has long served as a classic pedagogical technique to reinforce sound-symbol correspondences, spelling precision, and vocabulary recall (Alkire, 2002; Nation, 2008). However, traditional dictation exercises frequently depend on static, standardized texts that lack contextual relevance, often resulting in repetitive practice that fails to sustain learner interest.

In many EFL settings, such as Ecuadorian bilingual schools, A2-level learners frequently experience monotony during routine writing tasks, leading to persistent difficulties in orthographic precision and long-term vocabulary retention. While dictation remains a potentially powerful mechanism for reinforcing language form, conventional implementations struggle to foster meaningful engagement. Although emerging literature highlights the broad utility of AI for language teaching, limited empirical research examines the targeted application of AI-generated contextualized stories within dictation activities. Specifically, there is a lack of evidence regarding how level-adapted, AI-crafted narratives influence measurable linguistic outcomes such as orthographic accuracy and target word mastery (Godwin-Jones, 2022; Lee et al., 2026). Moreover, few studies have explored the affective factors that act as drivers or hindrances to student motivation and engagement when interacting with AI-driven dictation tasks. A clear gap exists in understanding how AI-generated stories can modernize dictation to simultaneously enhance linguistic precision and learner engagement.

To address this gap, the present study investigates the learning and affective impact of incorporating AI-generated contextualized stories into dictation activities for A2-level EFL learners. Also, this study contributes to the expanding literature on technology-enhanced language learning and AI-assisted materials design. It provides educators with an actionable framework for transforming dictation into an interactive, context-driven practice.

The study seeks to evaluate how AI-generated story dictations affect A2 EFL learners' communication and to identify the factors shaping their classroom engagement. Therefore, the study addresses the following research questions:

Research Questions

RQ1: How does integrating AI-generated contextualized stories into dictations influence A2 EFL learners' orthographic accuracy and target word mastery?

RQ2: What factors do A2 learners identify as driving or hindering their motivation and engagement when completing dictation tasks powered by AI-generated stories?

Literature Review

This literature review provides the theoretical and empirical grounding for investigating the integration of AI-generated contextualized stories into dictation activities for A2-level EFL learners.

The Cognitive Mechanics of Dictation

Dictation is an integrative, multi-component language exercise that requires simultaneous processing across several linguistic domains. Rather than being a passive transcription test, dictation engages working memory, phonological decoding, syntactic parsing, and orthographic encoding (Oller, 1979; Rost, 2011; Wajnryb, 1990). During a dictation task, the learner must:

- Decode incoming acoustic signals into recognizable phonemes.
- Maintain the acoustic stream in short-term auditory memory.
- Segment the speech stream into syntactic structures and lexical items.
- Access long-term memory to map phonemes to their corresponding orthographic graphemes (spelling patterns).
- Graphically produce the sentences while maintaining accuracy in mechanics, punctuation, and morphology.

Orthographic Mapping Theory (Ehri, 1995, 2005, 2014) provides a crucial lens for understanding this process. Orthographic mapping is the cognitive mechanism through which readers and spellers permanently connect the spoken phonemes of a word to its written graphemic representation, storing it in long-term memory for effortless retrieval. Dictation forces learners to actively engage in this mapping process because it demands explicit encoding (spelling out loud or in writing) rather than mere passive recognition (reading).

Benefits and Limitations of Conventional Dictation

Research consistently underscores the value of dictation for foundational literacy and language acquisition. Dictation reinforces sound-symbol correspondences, highlights grammatical bound morphemes (such as past-tense *-ed* or plural *-s*), uncovers subtle comprehension gaps, and strengthens short-term memory retention (Alkire, 2002; Nation, 2008; Oller, 1979; Wajnryb, 1990).

Despite these established benefits, traditional dictation practices exhibit notable pedagogical drawbacks:

Decontextualization

Conventional dictation often utilizes isolated sentences or fragmented passages taken from static textbooks. Without a compelling narrative structure, learners process linguistic input in a vacuum, which limits meaningful semantic associations (Davis & Rinvolucri, 1988; Willis, 1996).

Uniformity and Lack of Differentiation

Standard dictation passages treat student cohorts as homogeneous groups. They cannot adapt dynamically to the varying vocabulary knowledge, learning paces, or individual interests of different students (Tomlinson, 2014; Tomlinson & Imbeau, 2010).

Affective Disengagement

Due to its repetitive nature, traditional dictation is frequently perceived by elementary language learners as tedious, passive, and anxiety-inducing, leading to reduced task motivation and minimal active cognitive investment (Horwitz et al., 1986; Krashen, 1982, 1985; MacIntyre & Gardner, 1994).

The Shift Toward Generative AI in Language Education

The advent of generative AI and Large Language Models (LLMs) represents a paradigm shift in Computer-Assisted Language Learning (CALL) and materials design (Son, 2026). LLMs possess advanced natural language processing capabilities, enabling them to construct grammatically coherent, contextually rich, and stylistically varied texts based on specific pedagogical constraints.

In EFL instruction, LLMs offer language teachers the ability to shift from static, one-size-fits-all instructional content to dynamic, personalized learning materials (Son, 2026). Teachers can prompt LLMs to generate narrative texts tailored precisely to a specific CEFR level (A1–C2), ensuring that syntactic structures and lexical choices match the exact processing capabilities of target learners.

Story-Based Learning and Lexical Constraint Enforcement

Narrative-based instruction plays a critical role in vocabulary acquisition. Stories provide rich semantic contexts, narrative arcs, and emotional resonance, which aid memory consolidation and inferential processing (Craik & Lockhart, 1972; Mestres-Missé et al., 2007; Nation, 2013). When vocabulary words are embedded within a coherent story, learners do not merely memorize isolated definitions; they observe how words operate syntactically and semantically within authentic discourse.

Recent research highlights the efficacy of using LLMs for vocabulary-constrained story generation (Kamzela, 2025). LLMs can be instructed to write short narratives that intentionally integrate designated target words multiple times within meaningful, accessible scenarios (Kamzela, 2025). This algorithmically managed repetition ensures that learners receive optimal exposure to unfamiliar target items while remaining within their comprehensible input threshold (i+1).

Defining Orthographic Accuracy and Target Word Mastery

To evaluate the impact of AI-generated story dictations on language learning outcomes, two central constructs must be defined:

Orthographic Accuracy

The ability to convert spoken phonemes into conventional, correct graphemic representations according to the rules and patterns of standard English spelling. At the A2 level, this includes mastering silent letters, vowel digraphs, consonant clusters, and inflectional endings (Nation, 2013).

Target Word Mastery

A multidimensional construct encompassing form-meaning-use associations. True lexical mastery requires not only receptive recognition of a word's meaning but also productive orthographic recall (spelling) and an understanding of its contextual collocational behavior (Nation, 2001; Read, 2000)

Empirical Evidence on AI-Assisted Language and Vocabulary Gains

Empirical research demonstrates that integrating AI tools into language instruction yields significant improvements in student learning outcomes, vocabulary acquisition, and language sub-skills (Wei, 2023). Studies examining AI-mediated instruction report that interactive and tailored language tasks help learners achieve superior gains in lexical retention compared to traditional teacher-led or textbook-confined methods (Wei, 2023).

Furthermore, when AI models generate tailored texts that constrain vocabulary to known items while systematically introducing targeted target words, learners demonstrate enhanced post-intervention recall and spelling accuracy (Kamzela, 2025). By contextualizing target words inside personalized AI stories, dictation transforms from a mechanical spelling test into an integrated language task that promotes both orthographic precision and target word mastery.

The Multidimensional Framework of Engagement

Understanding how technology-enhanced tasks affect learners requires examining engagement as a construct composed of three interconnected dimensions:

Behavioral Engagement

Behavioral engagement reflects the outward, tangible manifestations of a learner's participation during a task (Philp & Duchesne, 2016). It answers the question: *Is the learner actively doing what the task requires?*

- Time-on-Task & Persistence: The duration and consistency with which a learner remains focused on dictation or transcription exercises without off-task distractions.
- Task Completion Rate & Productivity: The quantity and pace of dictated items processed and transcribed within specified timeframes.
- Active Participation & Physical Effort: Observable learning behaviors in digital environments, such as replaying audio segments, pausing media to verify input, adjusting playback speed, and making active edits.

In CALL settings, behavioral engagement is often the most easily measured via telemetry (e.g., clicks, re-plays, completion logs). However, literature highlights that high behavioral engagement alone (such as clicking quickly or completing a task superficially) does not guarantee deep learning unless paired with cognitive processing.

Cognitive Engagement

Cognitive engagement refers to the internal, psychological investment learners make in their own processing and comprehension (Helme & Clarke, 2001; Svalberg, 2009). It moves beyond mere physical completion to examine *how deeply the mind is working*.

- Deep Processing Strategies: Moving beyond superficial word-by-word transcription toward processing semantic units, analyzing syntax, and mapping acoustic signals onto mental L2 representations.
- Self-Regulation & Metacognition: The intentional deployment of strategies—such as self-monitoring comprehension, predicting missing vocabulary based on context, and pacing audio playback deliberately.
- Attention Allocation: The deliberate focus of limited cognitive/working memory resources toward acoustic nuances, spelling accuracy, or grammatical structures.
- Active Self-Correction: Evaluating draft transcriptions against source audio, detecting discrepancies, and revising errors independently.

In task-based language teaching (TBLT), cognitive engagement is critical for noticing (Schmidt, 1990)—the process by which learners notice gaps between their current L2 interlanguage and the target input during listening and transcription tasks.

Affective Engagement

Affective engagement encompasses the learner's emotional reactions, attitudes, and feelings toward the task, the digital tool, and the learning environment (Mercer, 2019; Skinner & Pitzer, 2012).

- Positive Emotional Responses: Experiencing intrinsic satisfaction, enjoyment, and a sense of accomplishment upon completing complex transcription segments.

- Interest, Curiosity, & Enthusiasm: Being motivated by the multimedia context, authentic digital audio, or interactive features embedded in the technology-enhanced task.
- Anxiety Reduction & Psychological Safety: Technology-enhanced dictation (e.g., self-paced software, digital audio controllers) allows learners to control input without the public pressure of immediate classroom performance, significantly reducing task-induced anxiety and foreign language listening anxiety.

Affective engagement acts as a crucial catalyst or filter. Positive emotional states enhance cognitive flexibility and willingness to persist through challenging audio input, whereas high anxiety consumes working memory resources needed for accurate transcription.

Factors Driving Motivation and Engagement

Literature on AI-mediated instruction identifies several key factors that serve as positive drivers of student motivation and engagement (Wei, 2023) presented in Table 1.

Table 1

Key Driving Factors Influencing Student Motivation and Engagement in AI-Mediated Language Learning

	Description
Contextual Relevance and Narrative Curiosity	AI generators can craft stories centered around themes that resonate with elementary learners (e.g., adventure, fantasy, local cultural settings). Curiosity about the narrative plot drives sustained attention throughout the dictation exercise.
Perceived Learning Value	As highlighted by research on AI instructional tools, immediate feedback and noticeable improvements in task success foster high performance expectancy, boosting intrinsic motivation and self-efficacy (Wei, 2023; Wang et al., 2025).

Novelty and Interactive Learning Environments	Moving away from static textbook passages to dynamic AI stories creates a captivating, interactive learning experience that reduces monotony and keeps young learners actively involved (Wei, 2023).

Factors Hindering Motivation and Engagement

Conversely, literature (e.g., Laufer & Ravenhorst-Kalovski, 2010; Moorhouse et al., 2024; Vandergrift & Goh, 2012) and pilot implementations reveal potential hindering factors that can impair engagement or introduce frustration during technology-supported language activities.

Table 2

Potential Hindering Factors Affecting Learner Engagement in Technology-Supported Language Tasks

	Description
CEFR Level Miscalibration	Although LLMs can adapt text complexity, they sometimes miscalibrate lexical levels, generating overly rare or syntactically complex structures that exceed A2 capabilities, thereby causing cognitive overload (Moorhouse et al., 2024; Tack et al., 2023).
Acoustic and Speed Challenges	In dictation, if the text delivery (whether read by a teacher or mediated via Text-to-Speech synthesis) is improperly paced or contains artificial audio artifacts, A2 learners quickly experience auditory processing fatigue (Vandergrift & Goh, 2012).
Excessive Task Difficulty	If too many unfamiliar target words are introduced simultaneously within a single story, the scaffolding effect of the narrative breaks down, elevating anxiety and reducing learner motivation (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2013).

Method

Research Design

This study employed a mixed-methods design (Cresswell & Plano Clark, 2018) to evaluate both the linguistic outcomes and affective experiences associated with AI-generated story dictations. A mixed-methods approach is adequate for this research because quantitative calculations are necessary to measure precise gains in orthographic accuracy and target word mastery (RQ1), while qualitative inquiry is required to uncover the specific positive and negative factors that explained learner motivation and engagement (RQ2).

Context and Participants

The study was conducted within an English as a Foreign Language (EFL) class at a bilingual educational institution in Ecuador. The target population consisted of secondary EFL learners formally benchmarked at the A2 level according to the Common European Framework of Reference for Languages (CEFR), evaluated through the EF-SET test.

Sample Size

A purposive sample of N=58 A2-level EFL students from two classroom sections (*paralelos*).

Participant Demographics

Native Spanish speakers aged 15–18 years with homogeneous baseline English proficiency but varying levels of prior exposure to generative AI tools.

Ethical Considerations

To safeguard the rights, dignity, and privacy of all participants, this study strictly adhered to established ethical guidelines for educational research involving human subjects and was approved by the Masters's degree program coordination . Before starting the study, participants of age—and legal guardians in the case of minor students—received comprehensive written information sheets detailing the study's objectives, pedagogical procedures, data collection instruments, and intended outcomes. Written informed consent was obtained from adult learners and parents. Participation was entirely voluntary, with explicit acceptance that declining to participate or choosing to withdraw at any point would cause no penalty, prejudice, or influence on students' academic evaluation or course grades.

To maintain strict confidentiality and anonymity, all personal identifying details were removed from raw dataset files and replaced with standardized alphanumeric codes (e.g., *Learner A1*, *Focus Group Participant B3*). All collected data—including pre/post-test scores, transcribed audio focus groups, and survey responses—were stored on secure, encrypted, password-protected cloud storage accessible exclusively to the principal author. Additionally, prompt engineering and text generation procedures were executed strictly at the instructional level, ensuring that no personally identifiable student information or student-generated output was uploaded to external AI servers.

Data Collection

Instruments and Materials

To address both research questions, four primary data collection instruments will be described in table 3.

Table 3

Overview of Data Collection Instruments, Target Constructs, and Administration Timeline

Research Question	Instrument	Purpose & Construct Measured	Timing
RQ1 (Linguistic Outcomes)	Orthographic Accuracy Pre-/Post-Test	Measures phoneme-grapheme conversion, spelling correctness, and inflectional accuracy across dictation transcripts.	Week 1 (Pre), Week 8 (Post)
	Target Word Mastery Test	Evaluates form-meaning-use association of target vocabulary (receptive definition matching & productive sentence completion).	Week 1 (Pre), Week 8 (Post)
RQ2 (Affective Outcomes)	Likert-Scale Engagement Survey	Measures behavioral, cognitive, and affective engagement.	Week 8 (Post-Intervention)
	Semi-Structured Focus Groups	Identifies specific student-reported *driving factors* (narrative interest, immediate feedback) and *hindering factors* (audio pace, cognitive overload).	Week 8 (Post-Intervention)

Data Collection Procedures

AI Prompt Engineering Protocol. To ensure pedagogical validity and consistency across the 8-week intervention, dictation texts were systematically generated using Gemini AI utilizing a standardized, engineered system prompt:

"Act as an expert EFL materials designer. Generate a short, highly engaging, narrative story (80–100 words) suitable for CEFR A2 English learners with the following constraints:

1. Syntactic structures must not exceed A2 complexity (simple past, present continuous).
2. Explicitly embed the following 10 target words: [insert target vocabulary].
3. Ensure target words are used in clear semantic contexts to aid comprehension.
4. Topic: An adventurous everyday scenario (e.g., a missing pet, a secret map)"

Dictation Implementation Routine. Participants completed one AI-generated story dictation per week across an 8-week period. Each session followed a three-phase dictation framework:

1 Pre-Dictation (5 mins): The instructor activated prior knowledge related to the story theme and briefly introduced phonological cues without pre-teaching the exact target word spellings.
2 Dictation Delivery (15 mins):
   - First Reading: The full story was played aloud at normal natural speed for gist comprehension.
   - Second Reading: The story was played in short breath groups (3–5 words), paused for transcription.
   - Third Reading: The complete story was played at a moderate pace for proofreading and self-correction.
3 Post-Dictation & Form Focus (10 mins): Students compared their transcribed text against the correct AI-generated script, highlighting orthographic discrepancies and target word usage.

Data Analysis

Quantitative Analysis (RQ1 & Survey Data)

The following analysis were conducted:

Orthographic Accuracy Scoring. Orthographic accuracy was evaluated using a standardized Percentage of Correct Words (PCW) index applied across all weekly dictation transcripts. Transcripts were benchmarked against the master AI-generated scripts (80–100 words per story). Each word was scored binarily (1 = correct orthography; 0 = incorrect/omitted) according to a predefined rubric evaluating phoneme-grapheme accuracy, inflectional morphology (e.g., past-tense *-ed*), and vowel digraph placement. Minor capitalization errors were not penalized to maintain focus on lexical orthography. Scores were converted to percentage indices to track performance shifts from Week 1 (pre-test) to Week 8 (post-test).

Target Word Mastery Scoring. Table 4 provides a description of the elements used to analyze the students' performance in retrieving the correct vocabulary.

Table 4

Binary Scoring Protocol and Error Classification Rubric for Orthographic Accuracy Evaluation

Error Category	**Rule / Condition**	**Example (Target → Student)**	**Score Impact**
Correct Word	Exact orthographic match (case-insensitive).	hero → hero	+1 point
Minor Capitalization	Missing or incorrect capital letters (unless proper noun).	Monday → monday	+1 point (No penalty)
Phonetic / Vowel Error	Incorrect vowel digraph or letter substitution.	rescue → resque	0 points
Omission / Addition	Missing letters, silent letter drops, or extra characters.	anxious → anxios	0 points
Inflectional Ending	Incorrect or omitted grammar suffixes (-ed, -s, -ing).	explored → explore	0 points

Illegible / Omitted Word	Skipped word or unreadable handwriting.	[blank]	0 points

Statistical Tests. Data was analyzed using Microsoft Excel. Descriptive statistics (Mean, Standard Deviation) was computed. Since this study was conducted with no intention to extrapolate to other schools settings, inferential statistics were considered not relevant and therefore not calculated.

Survey Analysis. Questionnaire responses were analyzed using descriptive statistics to evaluate mean agreement scores across engagement dimensions and motivational aspects.

Qualitative Analysis (RQ2)

Focus group audio recordings were transcribed verbatim and analyzed using Braun and Clarke's (2021) Six-Phase Thematic Analysis Framework:

1 Familiarization with audio transcripts.

2 Generating initial codes regarding learner perceptions.

3 Collating codes into candidate themes representing Driving Factors (e.g., narrative curiosity, level-matching, interactive novelty) and Hindering Factors (e.g., acoustic pacing, cognitive fatigue, unfamiliar accent).

4 Reviewing and refining themes against raw data.

5 Defining and naming themes.

6 Producing the final report with illustrative participant quotes.

Results

The primary objective of this study was to evaluate the linguistic outcomes and affective experiences associated with the implementation of AI-generated story dictations in a secondary EFL context. This section presents the empirical findings organized by the two research questions guiding the investigation.

Participant Sample and Retention

A total of *N*= 58 secondary EFL students across two classroom cohorts at an Ecuadorian educational institution initially enrolled in the 8-week intervention. All participants had a CEFR A2 proficiency level based on the EF-SET test administered prior to intervention.

Over the course of the 8-week study, 4 participants were excluded from the final analytical dataset due to missing two or more dictation sessions (n=2) or failing to complete either the post-test or delayed post-test assessments (n=2). Consequently, the final sample analyzed for both the quantitative and qualitative research questions comprised 54 participants (*N*=54), with 28 in Class A (*n*=28) and 26 in Class B (*n*=26). The cohort included 30 females (55.56%) and 24 males (44.44%), with a mean age of 16.40 years (*SD*=0.82). All participants reported Spanish as their first language (L1).

Quantitative Results

To evaluate linguistic development across the 8-week intervention, descriptive statistics (Mean and Standard Deviation) were calculated for both the Orthographic Accuracy Pre-/Post-Test and the Target Word Mastery Test across all participants (*N*=54).

Orthographic Accuracy

Orthographic performance was measured using the Percentage of Correct Words (PCW) index across student dictation transcripts. As shown in Table 5, participants demonstrated a noticeable increase in overall spelling accuracy from Week 1 to Week 8.

Table 5

Descriptive Statistics for Baseline and Post-Intervention Orthographic Accuracy and Composite Vocabulary Mastery (N = 54)

Metric	Pre-Test M (SD)	Post-Test M (SD)	Absolute Gain (M)
Orthographic Accuracy (PCW %)	68.4% (8.2)	86.1% (6.5)	+17.7%
Target Word Mastery (Composite Score)	10.8/20 (3.1)	16.75/20 (2.1)	+5.95%

At the baseline (Week 1), students frequently struggled with phoneme-grapheme conversions, particularly regarding silent letters, vowel digraphs, and inflectional endings (e.g., -*ed* past tense markers). By Week 8, post-test transcripts revealed fewer omission and substitution errors during the short-phrase transcription phase.

Target Word Mastery

Target vocabulary gains were evaluated using a 20-point composite test assessing both form-meaning association (receptive definition matching, 10 points) and productive contextual usage (sentence completion, 10 points). See Table 6 for specific details.

Table 6

Pre- and Post-Intervention Performance Breakdown for Receptive and Productive Target Vocabulary Sub-Skills

Vocabulary Sub-Skill	**Maximum Score**	**Pre-Test M (SD)**	**Post-Test M (SD)**	**Absolute Gain (M)**
Receptive Gains (Definition Matching)	10.0 pts	6.10 (1.80)	9.15 (0.92)	+3.05 pts
Productive Gains (Sentence Completion)	10.0 pts	4.70 (1.55)	7.60 (1.42)	+2.90 pts
Composite Score	20.0 pts	10.80 (2.95)	16.75 (2.10)	+5.95 pts

Student Affective Experiences: Engagement and Perceptions (RQ2)

The second research question addressed student motivation, engagement, and perceived learning conditions through a post-intervention survey and semi-structured focus groups.

The Week 8 Likert-Scale Engagement Survey assessed student perceptions across three dimensions on a 5-point scale (1=Strongly Disagree, 5=Strongly Agree). Table 7 provides the details.

Table 7

Post-Intervention Learner Engagement Ratings Across Behavioral, Cognitive, and Affective Dimensions

Engagement Dimension	**Mean (M)**	**Standard Deviation (SD)**
Behavioral Engagement (e.g., active participation, completion of transcription)	4.35	0.58
Cognitive Engagement (e.g., focus on phonological cues, self-correction efforts)	4.12	0.64

Affective Engagement (e.g., enjoyment of narrative themes, reduced anxiety)	4.28	0.71

Overall, students reported high levels of satisfaction, particularly regarding the narrative appeal of the stories and the immediate clarity provided by comparing their scripts against the AI text during the post-dictation phase.

Qualitative Results: Driving and Hindering Factors

Thematic analysis of transcribed focus group discussions yielded two overarching categories representing facilitators (driving factors) and challenges (hindering factors).

Table 8

Thematic Analysis Matrix of Student Perceptions Regarding AI Story Dictations

Category	**Emerging Theme**	**Core Description & Pedagogical Impact**	**Representative Participant Quote**
Category 1: Positive Factors *(Facilitators)*	Theme 1.1: Narrative interest and level-matched comprehension	AI stories kept student interest and prevented cognitive loss compared to isolated sentences.	*"The stories felt like mini-mysteries. Because the sentences were short and didn't have super long words, I could follow the plot while writing instead of getting lost."* Participant A8
	Theme 1.2: Immediate feedback via focus on the form	Phase 3 self-correction reduced spelling uncertainty and reinforced	*"Comparing my text with the real script right away helped me see my mistakes*

		orthographic patterns while fresh in memory.	*immediately. I realized I kept forgetting the past tense endings."* Participant B3
Category 2: Negative Factors *(Challenges)*	Theme 2.1: Acoustic pacing & speech rate	Pause intervals during the second reading (3–5 word breath groups) sometimes exceeded handwriting speed.	*"Sometimes the pause between three-word groups felt a little fast when I was still trying to finish writing the previous word."* Participant A15
	Theme 2.2: Cognitive fatigue during transcription	Some cognitive overload caused by balancing simultaneous phoneme decoding, listening, and legible writing.	*"Various learners mentioned it was difficult to process simultaneously listening, decoding phoneme, and maintaining legible spelling."* Participant B 8

Discussion

The quantitative results revealed substantial linguistic improvements across the 8-week intervention. Participants demonstrated a mean increase of +17.7% in orthographic accuracy (PCW %) and an overall gain of +5.95 points on the 20-point target word composite assessment (spanning both receptive definition matching and productive sentence completion).

These outcomes closely align with Orthographic Mapping Theory (Ehri, 1995, 2005, 2014), which posits that permanent spelling retention occurs when learners explicitly connect spoken phonemes to their graphemic representations in long-term memory. Conventional dictation often fails to engage these cognitive mechanics deeply when texts are decontextualized or uninspiring (Davis & Rinvolucri, 1988; Willis, 1996). However, by requiring learners to hold an acoustic stream of a coherent story in working memory, segment speech, and map phonemes to written forms, the AI-assisted dictations provided the deliberate encoding practice necessary for orthographic consolidation (Nation, 2008; Oller, 1979).

Furthermore, the simultaneous gains in receptive (+3.05 pts) and productive (+2.90 pts) target word mastery corroborate recent empirical evidence on vocabulary-constrained story generation (Kamzela et al., 2025; Wei, 2023). As Craik and Lockhart (1972) and Nation (2013) emphasize, embedding target words within a rich narrative structure facilitates deeper semantic processing than rote list memorization. Systematically prompting LLMs to enforce CEFR A2 syntactic constraints while repeatedly embedding target vocabulary ensured that input remained within the learners' comprehensible threshold (i+1) (Krashen, 1985; Son, 2026). This allowed students to observe word behavior syntactically and semantically within authentic discourse.

Affective Experiences: Drivers and Hindrances to Learner Engagement (RQ2)

The survey data indicated high mean agreement across the multidimensional engagement framework: Behavioral (M=4.35), Affective (M=4.28), and Cognitive (M=4.12). Qualitative findings from the focus groups elucidate the key drivers and hindrances shaping these dimensions:

Driving Factors

Participants reported that the story-based format "felt like mini-mysteries," driving sustained attention and cognitive investment. This confirms findings by Wei (2023) and Son and Philpott (2026), who noted that narrative curiosity and contextually relevant themes in CALL environments reduce task monotony and foster intrinsic motivation. The post-dictation form-focus phase (Phase 3), where students self-corrected against the AI script, allowed immediate identification of phoneme-grapheme discrepancies (e.g., omitted inflectional endings like *-ed*). This process directly supports Schmidt's (1990) Noticing Hypothesis and Svalberg's (2009) cognitive engagement framework, while building performance expectancy and self-efficacy (Wang et al., 2025; Wei, 2023).

Hindering Factors

Acoustic Pacing and Speech Rate: Pacing during the second reading (3–5 word breath groups) occasionally exceeded handwriting speed for slower writers.

Cognitive Writing Fatigue: Students experienced cognitive overload when simultaneously managing acoustic decoding, working memory storage, orthographic retrieval, and physical writing.

These challenges reflect the auditory processing and cognitive load risks identified by Vandergrift and Goh (2012) and Laufer and Ravenhorst-Kalovski (2010). They highlight that even when vocabulary complexity is appropriately controlled by an LLM, physical handwriting speed and real-time listening constraints can create operational bottlenecks if delivery speed is not flexibly adjusted.

Limitations

Several limitations should be considered when interpreting the results of this study:

Sample and Generalizability

The study utilized a purposive sample (*N*=54) from a single bilingual institution in Ecuador without inferential statistical testing, making the results context-bound to similar secondary EFL settings.

Delivery Mechanism and Pacing Controls

Dictation pacing was managed manually during live instruction rather than through individual, self-paced software, which likely heightened the acoustic pacing difficulties reported by slower writers.

Long-Term Retention

While post-test assessments demonstrated clear 8-week gains, the study did not include a delayed post-test (e.g., 6–8 weeks post-intervention) to measure the retention decay of orthographic mappings over time.

Implications

The findings offer actionable insights for secondary EFL educators and materials designers seeking to integrate generative AI into language instruction.

EFL educators can use standardized system prompts with LLMs to generate level-adapted, vocabulary-constrained narratives, replacing static textbook exercises with contextually rich input (Son, 2026; Tolstykh & Oshchepkova, 2024).

Teachers must monitor student writing speed during dictation and dynamically adjust pause lengths during breath-group readings to prevent auditory fatigue and cognitive overload (Vandergrift & Goh, 2012).

Incorporating immediate self-correction against the target AI script is essential for consolidating orthographic mapping and encouraging metacognitive self-regulation (Ehri, 2014; Svalberg, 2009).

Further Research

Future research should expand upon these findings in several directions: (1) Subsequent studies could utilize a quasi-experimental design featuring a control group (receiving traditional static dictation) and an experimental group (receiving AI story dictation) to statistically confirm relative gain differences. (2) Investigating the impact of specialized AI-driven TTS audio synthesis—allowing customized speed control and diverse native/non-native accents—could address the acoustic delivery challenges identified by learners. (3) Incorporating 6-month delayed post-tests would help determine whether the orthographic mapping and target word mastery gains observed in this study persist over time.

Conclusion

This study investigated the pedagogical and affective impact of integrating AI-generated contextualized stories into dictation activities for secondary A2-level EFL learners. By replacing static, decontextualized dictation passages with level-adapted narratives generated through Large Language Models (LLMs), the 8-week intervention demonstrated that modernizing traditional language learning tasks can simultaneously foster linguistic precision and learner engagement.

In conclusion, when implemented with deliberate prompt constraints and appropriate pedagogical scaffolding, AI-generated story dictations offer an effective, scalable framework for modernizing foundational EFL instruction. While LLMs provide language educators with powerful tools for dynamic materials design, successful classroom integration ultimately relies on teacher mediation to balance processing demands, manage acoustic delivery, and ensure an optimal, supportive learning environment

References

Alkire, S. (2002). Dictation as a language learning device. *The Internet TESL Journal*, *8*(3), 1–8.

Braun, V., & Clarke, V. (2021). One size does not fit all: Designing and conducting thematic analysis studies in counseling psychology. *Theoretical & Empirical Researches in Psychology*, 13(2), 201–218.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Davis, P., & Rinvolucri, M. (1988). *Dictation: New methods, new possibilities*. Cambridge University Press.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2), 116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, *26*(2), 5–24. https://doi.org/10.64152/10125/73474

Helme, S., & Clarke, D. (2001). Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal*, 13(2), 133–153.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Kamzela, W., Lango, M., & Dusek, O. (2025). SRS-Stories: Vocabulary-constrained multilingual story generation for language learning. In S. Potdar, L. Rojas-Barahona, & S. Montella (Eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: Industry Track* (pp. 630–645). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.emnlp-industry.44

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, *34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

Lee, S., Jeon, J., McKinley, J., & Rose, H. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics*, *45*, 85–108. https://doi.org/10.1017/S0267190525100184

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Mercer, S. (2019). Language learner engagement: Setting the scene. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 643–660). Springer. https://doi.org/10.1007/978-3-030-02899-2_40

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain bake one word: Event-related brain potentials show rapid contextual learning of new meanings. *Journal of Cognitive Neuroscience*, *19*(11), 1858–1870. https://doi.org/10.1162/jocn.2007.19.11.1858

Moorhouse, B. L. (2024). Generative artificial intelligence and ELT. *ELT Journal*, *78*(4), 373–377. https://doi.org/10.1093/elt/ccae032

Nation, I. S. P. (2008). *Teaching ESL/EFL reading and writing*. Routledge.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. Longman.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72.

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Rost, M. (2011). *Teaching and researching: Listening* (2nd ed.). Routledge.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Svalberg, A. M. L. (2009). Engagement with language: Interrogating a construct. *Language Awareness*, 18(3–4), 242–258.

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In *Handbook of Research on Student Engagement* (pp. 21–46). Springer, Boston, MA.

Son, J.-B., & Philpott, A. (2026). *AI-Powered Language Teaching* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009650809

Tack, A., Kochmar, E., Yuan, Z., Bibauw, S., & Piech, C. (2023). The BEA 2023 shared task on generating AI teacher responses in educational dialogues. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 785–795). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.64

Tolstykh, O. M., & Oshchepkova, T. (2024). Beyond ChatGPT: Roles that artificial intelligence tools can play in an English language classroom. *Discover Artificial Intelligence*, *4*, Article 158. https://doi.org/10.1007/s44163-024-00158-9

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.

Tomlinson, C. A., & Imbeau, M. B. (2010). *Leading and managing a differentiated classroom*. ASCD.

Wajnryb, R. (1990). *Grammar dictation*. Oxford University Press.

Wang, Q., Gao, Y., & Wang, X. (2025). Exploring engagement, self-efficacy, and anxiety in large language model EFL learning: A latent profile analysis of Chinese university students. *International Journal of Human–Computer Interaction*, *41*(12), 7815–7824. https://doi.org/10.1080/10447318.2024.2400403

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, Article 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Willis, J. (1996). *A framework for task-based learning*. Longman.