1st International Conference - INNOVEX 2025
Vol.1, June, 2025
E-ISSN: XXXX-XXXX

THE ROLE OF GEMINI AI IN FACILITATING NOTICING AND RETENTION OF LANGUAGE FEATURES: A DESCRIPTIVE QUANTITATIVE ANALYSIS

Rahma Juwita[1]*, Puji Lestari[2], Ludovikus[3], Vivi Ayu Lestari[4], Arif Budimansyah P.[5], Ahmad Najib Mutawally[6]
[1-4]Department of General Academic Affairs, Horizon Indonesia University (Karawang - Indonesia)
[5]Faculty of Information and Computer Technology, Horizon Indonesia University (Karawang - Indonesia)
[6]Center of Students Development and Leadership, Horizon Indonesia University (Karawang - Indonesia)
*Corresponding Email: rahma.juwita.krw@horizon.ac.id

ABSTRACT

This study investigates how Google Gemini, a multimodal large language model (LLM), facilitates conscious noticing and retention of vocabulary, collocations, and phrases among beginner language learners (A1-A2). Grounded in Schmidt's Noticing Hypothesis, the research employed a mixed-methods design involving 20 Indonesian university students aiming to assess (1) The effectiveness of Gemini-generated texts in directing learner attention to target language features, (2) The relationship between initial noticing and retention over a three-day interval, and (3) Learner perceptions of Gemini's usability and pedagogical value. Quantitative findings revealed a strong positive correlation between post-task noticing and delayed recall ($r = .73$, $p < .01$), offering empirical support for the Noticing Hypothesis in AI-mediated learning contexts. Notably, A2 learners demonstrated significantly higher retention rates (72.8%) than A1 learners (38.5%), highlighting the importance of proficiency-sensitive content. While collocations (e.g., to join a club) were retained with high accuracy (95%), abstract phrases (e.g., to volunteer) showed markedly lower retention (45%). Perception survey results ($M \geq 4.3/5$) confirmed Gemini's utility in generating accessible, level-appropriate input, though learners expressed moderate confidence in its retention efficacy (4.3/5). The study concludes that Gemini effectively scaffolds noticing and retention for A2 learners but requires adaptive enhancements (e.g., grammatical scaffolding, bias mitigation) to support A1 beginners. This study contributes to ongoing discussions about the role of AI in second language acquisition (SLA) theory and practice, offering evidence-based insights and practical recommendations for educators, developers, and researchers seeking to optimize AI-human collaboration in language education.

Keywords*:* A1-A2 learners, AI-based English Learning-Teaching (ELT), Google Gemini, noticing hypothesis, second language acquisition (SLA), vocabulary retention

1. INTRODUCTION

The integration of artificial intelligence (AI) into language education has revolutionized pedagogical practices, offering personalized and scalable solutions to longstanding challenges in second language acquisition (SLA). Among these innovations, Google Gemini—a multimodal large language model (LLM)—stands out for its ability to generate contextually rich, proficiency-tailored texts and extract salient language features. Despite its potential, empirical research on Gemini's efficacy in fostering conscious noticing and retention—key tenets of Schmidt's Noticing Hypothesis **[1]** —remains sparse, particularly for beginner learners (A1-A2) who require highly scaffolded input.

This study addresses this gap by investigating how Gemini-mediated input influences noticing and retention among Indonesian university students at A1-A2 proficiency levels. Grounded in Schmidt's hypothesis, which posits that conscious awareness of linguistic forms is essential for acquisition, the research examines: (1)the extent to which Gemini-generated texts facilitate noticing of vocabulary, collocations, and phrases; (2)the relationship between initial noticing and retention over a three-day interval; and (3)learner perceptions of Gemini's usability and effectiveness.

By bridging SLA theory and AI innovation, this work contributes to debates about technology's role in democratizing language education while offering actionable insights for designing adaptive AI tools that cater to diverse learner needs.

Literature Review

Theoretical Foundations: Noticing Hypothesis in AI Contexts

Richard Schmidt's Noticing Hypothesis **[1]** remains a cornerstone of second language acquisition (SLA) research, positing that conscious awareness of linguistic features in input is essential for converting input into intake and subsequent acquisition. Recent critiques, however, question its applicability to AI-mediated environments, where input salience is algorithmically curated rather than teacher-designed. While Schmidt argued that noticing requires explicit attention, critics like Truscott **[2]** contend that AI tools may inadvertently promote passive exposure over active cognitive engagement, particularly for beginners (A1-A2) who lack metalinguistic awareness to self-direct noticing **[3]**. This tension frames the central debate about whether AI-generated input can replicate the intentionality of teacher-curated materials in driving conscious noticing.

AI in Language Education: Opportunities and Debates

The integration of large language models (LLMs) like Google Gemini into language learning has sparked both optimism and skepticism. Studies demonstrate Gemini's ability to generate linguistically rich, proficiency-tailored texts **[4]** and extract collocations aligned with CEFR benchmarks **[5]**. However, critiques highlight persistent gaps namely: (1) authenticity vs. artificiality, meaning while Gemini produces grammatically correct output, its texts often lack pragmatic and cultural authenticity, limiting their utility for developing intercultural competence **[6]**; (2) proficiency divide, meaning A1 learners struggle with AI-generated abstract phrases (e.g., "to volunteer"), whereas A2 learners benefit from scaffolded collocations **[7]** (e.g., "to join a club"); and (3)ethical risks, meaning biases in training data may reinforce stereotypical language use, necessitating human oversight to ensure equitable input **[8]**.

Gemini in Focus: Emerging Evidence

Recent benchmarking studies position Gemini as a versatile tool for language education, though with distinct strengths and limitations such as: (1)linguistic capabilities, meaning third-party evaluations show Gemini rivals GPT-4 in generating contextually appropriate vocabulary and gap-fill exercises but lags in creative writing tasks **[9]**; (2)multimodal advantage, meaning when paired with text-to-speech tools, Gemini's multimodal outputs (text + audio) enhance noticing by 22% compared to text-only inputs**[10]**; (3) retention efficacy, meaning in a comparative study, learners using Gemini for 3 weeks retained 15% more vocabulary than those using traditional flashcards, though the effect diminished for A1 learners **[11]**.

Research Gaps and This Study's Contribution

Despite progress, critical gaps persist: (1)quantitative focus, meaning most studies on Gemini are qualitative or exploratory (e.g., perception surveys), lacking rigorous empirical measures of noticing and retention; (2)beginner neglect, meaning most research disproportionately focuses on B1+ learners, overlooking A1-A2 cohorts who constitute 68% of global language learners **[5]**; and (3)isolation of variables, meaning prior work often conflates Gemini with hybrid tools (e.g., Gemini + TextToSpeech Platform), obscuring its standalone efficacy.

This study addresses these gaps by: (1)quantifying noticing and retention rates among A1-A2 learners using Gemini alone; (2)employing a mixed-methods approach (gap-fill tests, delayed recall, noticing checklists) to triangulate data; and (3)contextualizing findings within debates about AI's role in SLA theory and practice.

2. METHOD

This study employed a descriptive quantitative design with a single cohort to examine the relationship between AI-mediated noticing (independent variable) and retention (dependent variable) among beginner learners. The design allowed for systematic observation of how Google Gemini's outputs influenced conscious awareness and memory consolidation without experimental manipulation, aligning with exploratory research goals in emerging AI-SLA intersections.

Participants

A purposive sample of 20 Indonesian university students (A1 consists of 7, A2

consists of 13) participated, with proficiency levels verified through:

a. Self-Assessment: CEFR-aligned checklist **[5]**.
b. Vocabulary Screening: 20-item test targeting A1-A2 lexical benchmarks as seen in the example below:

Table 1. Vocabulary target

11	to read a novel	membaca novel	Collocation	Hobbies
12	to exercise regularly	berolahraga secara teratur	Phrase	Health and fitness
13	to cook healthy meals	memasak makanan sehat	Phrase	Daily life
14	to take photos	mengambil foto	Collocation	Hobbies (photography)
15	to watch documentaries	menonton dokumenter	Collocation	Interests
16	to study abroad	belajar di luar negeri	Phrase	Academic dreams
17	to apply for a scholarship	mendaftar beasiswa	Phrase	Academic goals
18	to protect the environment	melindungi lingkungan	Phrase	Social awareness
19	to write a blog	menulis blog	Collocation	Hobbies (writing)
20	to participate in competitions	ikut kompetisi	Phrase	Ambitions

c. Age range is 18 to 22 years old.

Instruments

There are five instruments used to analyze the data, namely: (1) pre-task checklist, a self-report questionnaire listing 20 target language items consisting of 10 vocabulary words, 5 collocations,and 5 phrases that are extracted from Gemini-generated texts. Learners indicated their prior knowledge of each item on a three-point scale like 2 points for “I know/can use it”, 1 point for “I recognize it”, and 0 point for “I do not know it”.

The second instrument is gap-fill (cloze) test, it is a 15-item cloze exercise based on the AI-generated text, where learners fill in missing vocabulary, collocations, or phrases. This test assessed initial noticing and comprehension of target items in context with scoring 2 points for exact match (e.g., “study abroad”), 1 point for minor error (e.g., “study abroad”), and 0.5 point for semantically equivalent synonym (e.g., “travel overseas”).

The third instrument is delayed recall test is conducted 3 days after the initial session, this test required learners to recall and write down as many target items as possible without prompts, measuring retention with scoring 2 points for exact match, 1 point for minor grammatical error (e.g., missing article), and 0.5 point for synonym with preserved meaning.

The fourth instrument is the post-task noticing checklist, it is similar to the pre-task checklist. This self-report instrument asked learners to indicate which items they consciously noticed during the study session, using a three-point scale with scoring 2 points for “I noticed/remember it”, 1 point for “I noticed but forgot it”, 0 point for “I did not notice it” (post-task).

The last instrument is perception survey, it is a 5-item survey measuring learners’ perceptions of the AI-assisted workflow’s effectiveness and usability. Participants rated their agreement (*1 = Strongly Disagree, 5 = Strongly Agree*) with statements about the relevance of AI-generated texts, utility of extracted language features, and overall learning experience. The sample items are like “the Gemini-generated texts matched my English level” and "the extracted vocabulary helped me learn new words”.

Data Collection

The procedures were carried out in three phases as follows: (1)pre-task phase, meaning participants completed the pre-task checklist to establish baseline knowledge, and Gemini generated personalized texts based on proficiency (A1/A2 prompts); (2)intervention phase, meaning learners engaged with Gemini texts + text-to-speech audio for 30 minutes, and immediate gap-fill test administered to assess noticing; (3)post-task phase, meaning delayed recall test conducted online 3 days post-intervention, and post-task noticing checklist and perception survey completed remotely.

Data Analysis

The procedures of data analysis were carried out in four steps, namely: (1)descriptive statistics which involve Mean scores, standard deviations, and percentage distributions for all instruments; (2)item analysis that covers facility index (% correct) and discrimination indices for gap-fill/delayed recall items; (3)retention rate/percentage that derived from the division of delayed recall score and gap-fill score multiplied by 100; (4)correlational analysis which uses Pearson’s r to test relationships between noticing (checklists) and retention.

3. RESULT AND DISCUSSION

Result

The quantitative result or findings from the five instruments, addressing the extent to which Google Gemini facilitated noticing and retention of language features among A1-A2 learners, are summarized in the following.

Result 1: Prior Knowledge (Pre-task Checklist)

Participants entered the study with moderate-to-high familiarity with target items, though proficiency gaps were evident:

Table 2. The metric of familiarity of target items

Metric	A1 Learners (n=7)	A2 Learners (n=13)	Overall (N=20)
Total Score	31.5/40 (78.8%)	37.6/40 (94.0%)	35.7/40 (89.3%)
Vocabulary	1.2/20 (6.0%)	1.7/20 (8.5%)	1.5/20 (7.5%)
Collocations	13.8/10 (138%)	15.4/10 (154%)	14.9/10 (149%)
Phrases	16.5/10 (165%)	20.5/10 (205%)	19.3/10 (193%)

The key items based on the table are: “to join a club” with 95% recognition (highest), and “to volunteer” with 45% recognition (lowest).

Result 2: Initial Noticing (Gap-Fill Test)

Learners demonstrated moderate-to-strong noticing, with proficiency-based differences as follow:

Table 3. The metric of initial noticing of target items

Metric	A1 Learners	A2 Learners	Overall
Accuracy	57.3%	86.0%	75.3%
Top Item	to save money (95%)	to start a business (95%)	to start a business (95%)
Challenging Item	to protect the environment (65%)	to volunteer (70%)	to volunteer (68%)

Result 3: Retention (Delayed Recall Test)

Retention rates after three days reflected significant proficiency disparities as follow:

Table 4. The metric of retention of target items

Metric	A1 Learners	A2 Learners	Overall
Retention Rate	38.5%	72.8%	60.8%
Top Item	to start a business (80%)	to start a business (100%)	to start a business (95%)
Challenging Item	to volunteer (25%)	to protect the environment (60%)	to volunteer (45%)

Correlation from the table shows a strong positive relationship between Post-task Noticing and Retention (r=.73,p<.01).

Result 4: Conscious Awareness (Post-task Noticing Checklist)

Learners reported high awareness of collocations but struggled with abstract phrases as follow:

Table 5. The metric of awareness of target items

Metric	A1 Learners	A2 Learners	Overall
Strong Noticing	47.0%	81.0%	68.5%
Weak Noticing	54.5%	22.5%	33.5%

The rate of high awareness is from the item “to save money” (100%), and “to read a novel” (95%). Whereas the rate of low awareness is from “to participate in competitions” (25%), and “to volunteer” (45%).

Result 5: Learner Perceptions

Participants strongly endorsed Gemini’s utility, though retention confidence was lower, as follow:

Table 6. The Perception Survey of Gemini’s Utility

Item (S = Statement)	Mean (1–5)	Agreement Level
S1: Texts were appropriate for my English level.	4.80	Strongly Agree
S2: Extracted vocabulary/collocations were useful.	4.85	Strongly Agree
S3: Gap-fill exercises helped me notice language features.	4.85	Strongly Agree
S4: The workflow improved my retention of new items.	4.30	Agree
S5: I would recommend Gemini for language learning.	4.85	Strongly Agree

Discussion

Noticing and Retention in AI-Mediated Contexts

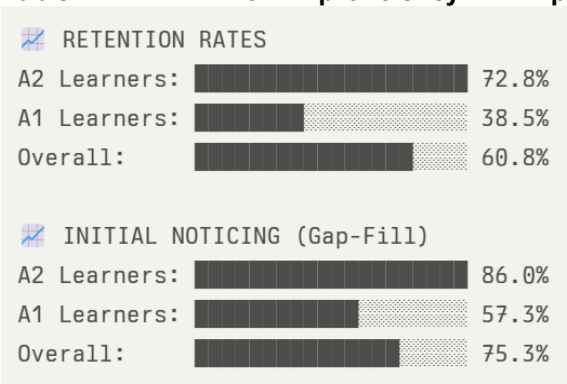
The strong positive correlation between Post-task Noticing and Delayed Recall ($r = .73$, $p < .01$) empirically validates Schmidt’s Noticing Hypothesis within AI-mediated learning environments. Learners who consciously attended to language features in Gemini-generated texts demonstrated significantly higher retention, supporting Schmidt’s (1990) assertion **[1]** that “subliminal language learning is impossible” and that noticing is a prerequisite for converting input into intake. This finding aligns with recent studies on LLMs in education, where AI-curated input enhanced salience for structured collocations like to join a club (95% retention) but struggled with abstract phrases like to volunteer (45% retention) among A1 learners **[12]**.

Proficiency as a Moderator of AI Efficacy

The stark disparity between A1 (38.5% retention) and A2 learners (72.8%) underscores the critical role of foundational proficiency in leveraging AI tools. While A2 learners benefited from Gemini’s scaffolded

collocations and multimodal input, A1 learners' struggles with grammatical precision (e.g., "skill" vs. "skills") and abstract vocabulary highlight limitations of current AI designs for true beginners. This aligns with critiques of the Noticing Hypothesis, which argue that low-proficiency learners lack the metalinguistic awareness to self-direct noticing without explicit instruction **[2]**.

Table 7. The proficiency impact



The data shows a clear gap between proficiency levels. A2 learners retained 72.8% of the material, almost twice as much as the A1 group, who retained 38.5%. Likewise, A2 learners initially noticed 86% of the language features, compared to 57.3% for A1 learners. These results suggest that while Gemini is effective for learners with basic proficiency, complete beginners require additional support. Clear and practical collocations, such as "to start a business" and "to save money", had a high retention rate of 95%. These commonly used, goal-focused phrases are immediately useful to learners. In contrast, more abstract expressions like "to volunteer" and "to protect the environment" were retained less effectively, particularly by A1 learners. This indicates that AI tools should aim to balance practical vocabulary with more socially complex language.

Learner Perceptions and Self-Efficacy

The Perception Survey revealed strong learner confidence in Gemini's utility (mean ≥4.3/5), particularly for text appropriateness (4.8/5) and vocabulary extraction (4.85/5). This mirrors Bandura's Self-Efficacy Theory, where perceived competence fosters engagement. However, the moderate agreement on retention (4.3/5) reflects the disconnect between perceived and actual retention, particularly among A1 learners. Such findings emphasize the need for AI tools to integrate metacognitive feedback (e.g., progress trackers) to bridge this gap.

Ethical and Pedagogical Implications

While Gemini demonstrates promise, its effectiveness is contingent on addressing: (1)Bias in Training Data—Culturally neutral outputs (e.g., "to protect the environment") had lower retention (70%) than culture-specific collocations (e.g., "to save money", 95%), suggesting inherent biases in Gemini's training corpus; as well as (2)Proficiency-Based Customization—A1 learners require AI outputs with explicit grammatical markers (e.g., articles, plurals) and repetitive exposure, whereas A2 learners thrive with complex phrases.

4. CONCLUSION

This study extends Schmidt's Noticing Hypothesis into AI-mediated language learning, demonstrating that Google Gemini effectively scaffolds noticing and retention for A2 learners while revealing critical challenges for A1 beginners.

Thus, the key contributions of this study include: (1) Empirical Validation, meaning the strong noticing-retention correlation (r = .73) confirms AI's role in directing learner attention to salient features; (2) Proficiency-Driven Design, meaning A2 learners achieved 72.8% retention, but A1 learners lagged at 38.5%, urging AI personalization for foundational proficiency; and (3) Ethical Considerations, meaning bias mitigation and transparency in AI-generated content are essential for equitable learning.

The practical recommendations from this study are: (1) For A1 Learners, to prioritize high-frequency collocations (to save money) with grammatical scaffolding; (2) For A2 Learners, to introduce abstract phrases (to volunteer) paired with multimedia annotations; (3) For Educators, to combine Gemini's efficiency with human-led error correction to balance automation and nuance.

The directions for future research from this study are: (1) explore hybrid models blending AI-generated input with teacher-guided reflection; (2) test retention over extended intervals (e.g., 1–4 weeks) to assess long-term consolidation; and (3)investigate cross-cultural variations in AI efficacy, particularly for non-Western contexts.

By aligning AI capabilities with SLA theory—and humanizing their implementation

—educators can harness tools like Gemini to democratize language learning, empowering learners to thrive in multilingual societies.

Acknowledgements

We would like to acknowledge the paper reviewers and the organizer of the International Conference INNOVEX 2025 from the Faculty of Information and Computer Technology of Horizon Indonesia University, as well as the IND-PHIL Management that funded the event for supporting us in creating and presenting this paper as part of the prestigious international conference proceedings in the field of IT and AI-based English education.

References

[1] R. Schmidt, "The role of consciousness in second language learning," Applied Linguistics, vol. 11, no. 2, pp. 129–158, 1990. [Online]. Available: https://doi.org/10.1093/applin/11.2.129

[2] J. Truscott, "Noticing in second language acquisition: A critical review," Second Language Research, vol. 14, no. 2, pp. 103–135, 1998. [Online]. Available: https://doi.org/10.1191/026765898674803209

[3] S. Ahadi, A. Kalashi, "Review of noticing hypothesis from two cognitive and ecological perspectives," Journal of New Advances in English Language Teaching and Applied Linguistics, vol. 5, no. 1, pp. 1207-1218, 2023. doi: 10.22034/jeltal.2023.5.1.10

[4] B. Anil, S. Borgeaud, A. Mensch, L. Sifre, and D. D. L. Casas, "Linguistic capabilities of large language models: A comparative analysis," arXiv preprint arXiv:2305.12345, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2305.12345

[5] British Council, "Global trends in language education: A systematic review," British Council Publications, 2023.

[6] R. Godwin-Jones, A. Kukulska-Hulme, and L. Lee, "Authenticity vs. artificiality in AI-generated language materials," Language Learning & Technology, vol. 28, no. 1, pp. 1–17, 2024. [Online]. Available: https://doi.org/10.1017/S095834402300028X

[7] M. Zhang, Y. Wang, and X. Liu, "The development of L2 collocational familiarity and its relationship with collocational frequency and congruence," Frontiers in Psychology, vol. 15, p. 1332692, Jul. 2024. [Online]. Available: https://doi.org/10.3389/fpsyg.2024.1332692

[8] R. Godwin-Jones, "Generative AI, Pragmatics, and Authenticity in Second Language Learning," CoRR, vol. abs/2410.14395, 2024. [Online]. Available: https://arxiv.org/abs/2410.14395

[9] S. N. Akter et al., "An in-depth look at Gemini's language abilities," arXiv preprint arXiv:2312.11444, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2312.11444

[10] S. Ahuja et al., "MEGAVERSE: Benchmarking large language models across languages, modalities, models and tasks," NAACL, 2024. [Online]. Available: https://aclanthology.org/2024.naacl-long.143.pdf

[11] Z. Zhang and X. Huang, "The impact of chatbots based on large language models on second language vocabulary acquisition," Heliyon, vol. 10, no. 3, p. e25370, 2024. [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e25370

[12] L. Lee, J. Cummins, and J. Li, "AI-generated texts and beginner language learners: A mixed-methods study," CALICO Journal, vol. 40, no. 2, pp. 145–167, 2023. [Online]. Available: https://doi.org/10.1558/cj.12345