

UDK:3713

**THE IMPACT OF MULTIMODAL GENERATIVE ARTIFICIAL
INTELLIGENCE ON ORAL PROFICIENCY: A SURVEY OF GOOGLE
GEMINI IN SELF-DIRECTED EFL LEARNING**

Shamuratova Dilnoza Jamolovna

A senior teacher,

UzSWLU, the Department of English Applied Disciplines №1

Tel: 977676787

E-mail: dizamura5@gmail.com

Eshmuratova Aysuliy Rustamovna

Student of 2528 group, Uzbekistan state world languages university

Tel: 883490053

E-mail: djtustudent@gmail.com

ABSTRACT

English as a Foreign Language (EFL) learners often have ongoing fluency issues in self-directed learning settings, which are made worse by restricted access to native speakers. This article examines how Google Gemini, a multimodal Generative AI (GenAI) platform, can help overcome these challenges. Speaking performance could be revolutionized by recent developments in Gemini's capacity to mimic human-like interactions and adjust to individual demands (Widodo et al., 2026). This paper summarizes recent research on how well Gemini supports self-directed practice and task-based learning. According to quantitative data, speaking fluency has improved by an average of 13%, particularly in the areas of lexical diversity and decreased hesitation (Widodo et al., 2026). Increased speaker confidence and the advantages of a nonjudgmental, round-the-clock atmosphere are highlighted via qualitative analysis. The article concludes by examining the methodological challenges of integrating GenAI into formal curricula and the necessity for longitudinal research to substantiate long-term efficacy.

Keywords: *Google Gemini, Speaking Fluency, EFL Pedagogy, Task-Based Learning (TBL), Generative AI, Oral Proficiency.*

**ВЛИЯНИЕ МУЛЬТИМОДАЛЬНОГО ГЕНЕРАТИВНОГО
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА УРОВЕНЬ ВЛАДЕНИЯ
УСТНОЙ РЕЧЬЮ: ОБЗОР ИСПОЛЬЗОВАНИЯ GOOGLE GEMINI В
САМОСТОЯТЕЛЬНОМ ИЗУЧЕНИИ АНГЛИЙСКОГО ЯЗЫКА КАК
ИНОСТРАННОГО (EFL)**

АННОТАЦИЯ

Изучающие английский язык как иностранный (EFL) в условиях самостоятельного обучения часто сталкиваются с устойчивыми трудностями в развитии беглости речи, что усугубляется ограниченным доступом к носителям языка. В данной статье рассматривается роль Google Gemini — мультимодальной платформы генеративного ИИ (GenAI) — в преодолении этих барьеров. Последние достижения в способности Gemini имитировать человеческое взаимодействие и адаптироваться к индивидуальным потребностям могут произвести революцию в развитии речевых навыков (Widodo et al., 2026). В работе обобщены результаты актуальных исследований эффективности Gemini в поддержке самостоятельной практики и обучения на основе заданий (TBL). Согласно количественным данным, показатели беглости речи улучшились в среднем на 13%, особенно в таких аспектах, как лексическое разнообразие и сокращение пауз (Widodo et al., 2026). Качественный анализ подчеркивает рост уверенности учащихся и преимущества круглосуточной среды обучения, свободной от субъективного оценивания. В заключении статьи рассматриваются методологические проблемы интеграции генеративного ИИ в официальные учебные программы и указывается на необходимость долгосрочных исследований для подтверждения устойчивости полученных результатов.

Ключевые слова: Google Gemini, беглость речи, методика преподавания EFL, обучение на основе заданий (TBL), генеративный искусственный интеллект, уровень владения устной речью.

**MULTIMODAL GENERATIV SUN'IY INTELLEKTNING OG'ZAKI
NUTQ DARAJASIGA TA'SIRI: MUSTAQIL RAVISHDA INGLIZ TILINI
O'RGANISHDA (EFL) GOOGLE GEMINI IMKONIYATLARI TAHLILI**

ANNOTATSIYA

Ingliz tilini chet tili sifatida o'rganuvchilar (EFL) mustaqil ta'lim jarayonida ko'pincha nutq ravonligi bilan bog'liq muammolarga duch kelishadi, bu holat ona

tilida so‘zlashuvchilar bilan muloqot qilish imkoniyati cheklanganligi sababli yanada murakkablashadi. Mazkur maqolada Google Gemini multimodal generativ sun‘iy intellekt (GenAI) platformasining ushbu muammolarni bartaraf etishdagi o‘rni tadqiq qilingan. Geminining insonga xos muloqotni imitatsiya qilish va individual ehtiyojlarga moslashish borasidagi so‘nggi yutuqlari og‘zaki nutq ko‘rsatkichlarini tubdan o‘zgartirishi mumkin (Widodo va boshqalar, 2026). Maqolada Geminining mustaqil amaliyot va topshiriqlarga asoslangan o‘qitish (TBL) metodikasini qanchalik samarali qo‘llab-quvvatlashi haqidagi so‘nggi tadqiqotlar umumlashtirilgan. Miqdoriy ma‘lumotlarga ko‘ra, og‘zaki nutq ravonligi o‘rtacha 13 foizga yaxshilangan, ayniqsa leksik xilma-xillik va nutqdagi tutilishlarning kamayishi sezilarli darajada kuzatilgan (Widodo va boshqalar, 2026). Sifat tahlili esa so‘zlovchi ishonchining ortishi hamda sun‘iy intellektning 24/7 rejimida ishlovchi, tanqid va tazyiqlardan xoli muhitining afzalliklarini ko‘rsatib beradi. Maqola yakunida generativ sun‘iy intellektni rasmiy o‘quv dasturlariga integratsiya qilishdagi metodologik qiyinchiliklar va uning uzoq muddatli samaradorligini tasdiqlash uchun davomiy (longityud) tadqiqotlar o‘tkazish zarurligi muhokama qilinadi.

Kalit so‘zlar: Google Gemini, nutq ravonligi, EFL pedagogikasi, topshiriqlarga asoslangan o‘qitish (TBL), generativ sun‘iy intellekt, og‘zaki nutq darajasi.

INTRODUCTION

With the advent of Large Language Models (LLMs), the field of Computer-Assisted Language Learning (CALL) has seen a dramatic change from static, pre-programmed replies to dynamic, generative interactions. Among these developments, Google Gemini is a key advancement in multimodal AI, providing an advanced interface that surpasses the constraints of conventional text-based teaching resources (Al-Obaydi, 2026; Kawinkoonlasate, 2026). Gemini offers a scalable solution for the "interaction gap" in English as a Foreign Language (EFL) contexts, whereas the pursuit of oral proficiency has traditionally been limited to the physical classroom—often hindered by high student-to-teacher ratios and limited opportunities for individual output (Baskara, 2025; Widodo et al., 2026).

Improving speaking abilities is a particularly difficult task that calls for phonological accuracy, real-time cognitive processing, grammatical precision, and the capacity to control communication tactics under duress. In the past, "foreign language speaking anxiety," a psychological obstacle that frequently prevents the spontaneous production of speech, has plagued students in non-native settings (Widodo et al., 2026). By providing a low-stakes, round-the-clock setting where

students may participate in "risk-free" trial and error, Gemini's integration into the learning ecosystem tackles this. In contrast to text-only models, Gemini employs a multimodal framework that processes text, audio, and visual input simultaneously, more closely emulating the multisensory aspect of human communication (Al-Obaydi, 2026; Setyawan et al., 2026).

Additionally, the need for tools that can replicate real-world, situational discussion has grown due to the theoretical shift towards Communicative Competence and Task-Based Learning (TBL) (Widodo et al., 2026; Younas, 2025). Gemini's capacity to take on distinct personalities, such as a casual peer or a professional interviewer, enables a highly personalized "scaffolding" of the learning process. For self-directed learners who need quick, diagnostic feedback to close the gap between their current interlanguage and target-level fluency, this flexibility is essential (Kawinkoonlasate, 2026). Understanding the methodological implications of such tools is crucial as we navigate the educational landscape of 2026 in order to create curriculum that are both pedagogically sound and technologically sophisticated.

METHODS

This assessment synthesizes data from recent experimental research and longitudinal educational interventions carried out between 2024 and 2026 to provide a thorough picture of how Google Gemini aids speaking enhancement. Three separate layers of implementation comprise the methodological framework: The Framework for Task-Based Language Teaching (TBLT) The main methodological strategy found in the literature is incorporating Gemini into cycles of task-based learning. These studies use Gemini's conversational mode as the main interlocutor and assign learners real-world communicative goals, including negotiating a commercial contract or defending a thesis (Widodo et al., 2026). The pre-task (brainstorming vocabulary with Gemini), task cycle (the actual spoken conversation), and post-task (examining Gemini's transcribed feedback for language faults) are the three main components of this approach.

Gemini's multimodal nature enables Digital Scenario-Based Teaching, in contrast to earlier versions of AI. To encourage spontaneous oral production, researchers have used techniques in which Gemini produces high-fidelity visual and aural cues, such as a busy airport or a medical emergency (Younas, 2025). This approach assesses the learner's capacity to sustain "contextual fluency," moving the emphasis from basic sentence formation to pragmatic competency and situational appropriateness (Kawinkoonlasate, 2026).

Many of the surveyed research use quantitative comparative analysis to guarantee objective outcomes. These techniques include:

- A/B testing involves contrasting an experimental group (Gemini-mediated practice) with a control group (conventional classroom interaction).
- Suprasegmental Analysis: Measuring changes in the experimental group's mean length of run (MLR), frequency of non-functional pauses, and speech rate (syllables per second) using acoustic software (Setyawan et al., 2026).
- Rubric-Based Expert Evaluation: To validate AI-reported gains, human raters who are blind to the intervention grade the participants on "Interactivity" and "Coherence" using CEFR-aligned rubrics (Widodo et al., 2026).

Because speaking is both a language and a psychological act, researchers have used Likert-scale surveys and narrative inquiry. These qualitative approaches monitor shifts in "Foreign Language Virtual Anxiety" (FLVA), investigating how a learner's desire to communicate (WTC) is affected over time by Gemini's perceived "patience" and round-the-clock availability (Aruta, 2026; Widodo et al., 2026).

RESULTS

The incorporation of Google Gemini into EFL curricula produces notable, quantifiable benefits in both language and psychological areas, according to a synthesis of empirical data from 2024–2026. Four main impact areas are identified from the results:

When learners use Gemini for self-directed practice, intervention studies typically show a statistically significant gain in oral fluency. In particular, over a twelve-week period, longitudinal data reveals an average improvement in core speaking metrics of 13% (Widodo et al., 2026). These improvements are especially noticeable in "temporal fluency," which includes a decrease in the "Mean Length of Pause" and an increase in speaking pace (measured in syllables per minute). Students showed that they could now keep their speech flowing without using a lot of non-functional fillers like "um" or "uh"(Setyawan et al., 2026).

Gemini can make subtle changes that conventional software cannot because of its advanced Large Language Model (LLM) design. The capacity to couple words organically, like "heavy rain" rather than "strong rain," has significantly improved, according to the results (Aruta, 2026). Additionally, the data indicates that exposure to Gemini's output encourages learners to use more complex grammatical structures, like third conditionals and passive voice, and more sophisticated "hedging" words, like arguably, potentially, and seemingly, which are frequently avoided in spontaneous speech (Kawinkoonlasate, 2026).

The effect of AI on the psychological condition of the learner is among the most notable qualitative findings. High levels of "Foreign Language Speaking Anxiety" are frequently triggered in traditional classroom settings because of the fear of instructor correction or peer criticism. However, polls show that the Gemini interface's "unlimited patience" and "non-judgmental" features considerably reduce the Affective Filter (Widodo et al., 2026). Students practice "Willingness to Communicate" (WTC) up to four times more frequently than they would in a typical lab setting as a result of this decrease in apprehension.

The efficacy of feedback has been redefined by Gemini's capacity to process and react through audio and imagery, in contrast to text-only models. The findings indicate that students react more favorably to multimedia-based feedback, such as when Gemini provides a phonetic transcription for pronunciation correction or an image to clarify a vocabulary problem (Kawinkoonlasate, 2026). When compared to typical written corrections given by a teacher after the speaking exercise was completed, real-time audio feedback from Gemini was found to be 22% more helpful for pronunciation retention (Al-Obaydi, 2026).

ANALYSIS

According to an analytical synthesis of available data, Google Gemini's effectiveness in improving speaking is based on its capacity to close the "Output Gap"—the difference between a learner's receptive knowledge and their production competence. The results' critical analysis reveals the following themes:

Gemini's multimodal architecture is a key component of its success. Conventional CALL (Computer-Assisted Language Learning) technologies frequently approach speaking as a result of reading, isolating skills. According to analysis, Gemini produces a "rich-input environment" that promotes deeper cognitive learning by processing visual, literary, and aural information simultaneously (Kawinkoonlasate, 2026). Gemini performs better than monomodal models by offering source integration that replicates real-world complexity, especially in resource-constrained or rural situations where genuine English exposure is restricted (Al-Obaydi, 2026; Kawinkoonlasate, 2026).

The analysis shows that the nature of error treatment has changed. While delayed correction may result in the "fossilization" of mistakes, fast correction might disturb "flow" in human-led classrooms. Gemini's algorithmic feedback strikes a compromise by producing an accurate, behavior-based transcript for post-task analysis while permitting "uninterrupted output" throughout the task (Widodo et al., 2026). According to data, there is a strong correlation ($p < .05$) between the quick

development of academic speaking competency and this "on-demand" diagnostic input (Setyawan et al., 2026).

Gemini redefines the power dynamics of the language-learning relationship from a sociolinguistic standpoint. According to an analysis of student narratives, students saw Gemini as a "feedback teammate" or "partner in discourse" rather than as a "examiner" (Agostini, 2026). As the characteristic of true fluency, "extemporaneous speech"—prepared speech that is not memorized—is encouraged by this change in agency. AI-mediated benefits must be properly transferred into human-to-human environments, according to academics, who warn that this can result in "technology scaffolding" where a learner's performance might drastically decline when the AI is removed (Tien, 2026).

Lastly, the data reveals that Gemini's greatest advantage over earlier text-based LLMs is its capacity to improve "Prosody" (intonation and stress). Gemini can identify and guide students on "word stress" and "sentence rhythm," which are frequently the most challenging concepts for L2 (second language) learners to acquire through written study alone, by using suprasegmental analysis (Setyawan et al., 2026; Widodo et al., 2026). This implies that AI's capacity to serve as a phonetic coach in addition to a linguistic one will be crucial for improving speaking in the future.

DISCUSSION

A comprehensive analysis uncovers a number of subtle issues that call for caution, notwithstanding the educational benefits of Google Gemini in promoting oral proficiency. The problem of algorithmic feedback variability is at the heart of the present scholarly discussion. There is an inherent risk of "hallucinations" or grammatical irregularities that could unintentionally foster non-standard usage because Gemini's responses are created based on statistical patterns from large web-based datasets (Kawinkoonlasate, 2026). Because of this, learners must possess a high degree of "AI literacy" in order to discern between possible machine faults and excellent verbal output.

It's also important to distinguish between pragmatic competence and surface-level fluency. Although Gemini is quite good at boosting a learner's "speech pace" and "lexical diversity," its ability to mimic the social subtleties of human interaction is still being studied (Widodo et al., 2026). It is challenging for existing GenAI models to fully replicate higher-order communicating tactics, such as the delicate art of "negotiation of meaning," "nuanced turn-taking," and the understanding of cultural idioms in real social discussions (Agostini, 2026). Additionally, there is the "Isolation Risk": if students only use AI to practice speaking, they would find it difficult to deal

with the unexpected nature of human-to-human communication, where body language and emotional intelligence are crucial (Tien, 2026).

Finally, the ethics of data privacy and the "digital reliance" of learners in poor places must be addressed. Researchers wonder if the benefits of AI-mediated practice are actually internalized or if they are merely a transient "performance boost" that disappears without the interface, as Gemini turns into a "cognitive crutch" for oral production (Al-Obaydi, 2026; Kawinkoonlasate, 2026).

CONCLUSION

As a potent, multimodal addition to conventional EFL training, Google Gemini is a revolutionary advancement in the field of computer-assisted language learning. Its main benefit is that it democratizes advanced speaking practice by providing a round-the-clock, accepting setting for students who encounter psychological, financial, or geographic obstacles to conventional dialogue (Widodo et al., 2026). Gemini bridges the gap between theoretical knowledge and spontaneous oral production by greatly increasing lexical diversity and decreasing the "hesitation patterns" linked to speaking anxiety.

However, rather than being replacement-based, the incorporation of such technology should be seen as complimentary. In order to provide AI with the sociocultural context and ethical guidance it now lacks, human educators continue to play a crucial role. In order to ascertain the long-term "retention rate" of AI-mediated fluency increases, longitudinal designs with a variety of demographic cohorts must be given top priority in future research. Furthermore, creating more complex, human-centered educational interactions will need improving AI's "turn-taking mechanics" and emotional reactivity (Kawinkoonlasate, 2026; Widodo et al., 2026). In the end, the effective use of Gemini for speaking improvement hinges on a harmonious fusion of cutting-edge technology and tried-and-true teaching methods.

REFERENCES

1. **Agostini, D. (2026).** When ChatGPT joins the team: a mixed-methods study of AI-mediated collaborative lesson design. *Frontiers in Education*.
2. **Al-Obaydi, L. H. (2026).** The use of Copilot, Gemini and ChatGPT in the context of foreign language learning and teaching: An academic technology review. *Contemporary Educational Technology*, 18(1).
3. **Aruta, M. G. H. (2026).** Exploring The Impact of AI Chatbots on Vocabulary Development of Teacher Education Students: A Narrative Inquiry. *International Research Journal of Multidisciplinary Scope (IRJMS)*, 7(2), 608.

4. **Baskara, R. (2025).** Navigating the AI Frontier in EFL: A Review of Current Trends and Tools. *Journal of Digital Learning*.
5. **Kawinkoonlasate, P. (2026).** A comparative study of Google Gemini and ChatGPT in enhancing English language learning for EFL learners. *Pedagogical Research, 11(2)*.
6. **Setyawan, S., et al. (2026).** Speech fluency and accuracy as indicators of academic speaking proficiency. *JEAPCO*.
7. **Tien, C. Y. (2026).** Exploring AI-mediated linguistic and cognitive support for non-native English speakers in English-only higher education. *ERIC*.
8. **Widodo, A., et al. (2026).** Task-Based Learning Reimagined: How Gemini AI Enhances EFL Students' Speaking Fluency in Self-Directed Learning. *JOLLT Journal of Languages and Language Teaching, 14(2)*.
9. **Younas, M. (2025).** Evaluating the effectiveness of digital scenario-based English teaching at the university level using the artificial intelligence generated content. *Frontiers in Education*.