HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

EFFECTS OF MULTIMODAL SPEAKING TASKS ON YOUNG EFL LEARNERS' SPEAKING FLUENCY, PRONUNCIATION, AND WILLINGNESS TO COMMUNICATE

Sevara Alimurodova
English Language Teacher, School No. 1, Yangiyer, Sirdaryo, Uzbekistan
Doi: https://doi.org/10.5281/zenodo.22727692

Abstract: This quasi-experimental study examines the effects of multimodal speaking tasks on young EFL learners' speaking fluency, pronunciation, and Willingness to Communicate (WTC). Over 12 weeks, 60 fourth-grade students at School No. 1 in Yangiyer were assigned to experimental and control groups. The intervention combined visual flashcards, digital stories, audio-gestural prompts, and task-based interactions. Pre- and post-tests showed substantial gains in the experimental group: speaking fluency increased from 27.9 to 44.6 words per minute (WPM; $p < .001$), pronunciation scores improved from 2.55 to 3.80 on a 5-point scale, and WTC increased from 2.35 to 4.10. The findings suggest that multimodal speaking tasks can support oral fluency, pronunciation, and classroom participation by providing contextual and visual support for young EFL learners.

Keywords: Multimodal speaking tasks; young EFL learners; speaking fluency; pronunciation; willingness to communicate; visual scaffolding; task-based language teaching.

INTRODUCTION. In contemporary primary English as a Foreign Language (EFL) education, cultivating communicative competence—specifically fluent, spontaneous, and phonologically accurate oral expression—has emerged as a primary curriculum benchmark worldwide. In the context of Uzbekistan's ongoing educational reforms, aligning primary school language instruction with the Common European Framework of Reference for Languages (CEFR A1–A1+ descriptors) demands a shift from rote memorization toward authentic oral interaction. However, primary grade learners frequently encounter substantial cognitive, linguistic, and affective barriers during oral production.

Young EFL learners aged 8 to 10 years are situated in a unique developmental transition phase. While their cognitive capacity for abstract grammatical analysis remains nascent, their capacity for multi-sensory, experiential learning is at its peak. Traditional EFL instruction in primary public schools, which heavily relies on text-centric textbooks, teacher-fronted grammar explanations, and mechanical choral repetition, may not fully accommodate these developmental characteristics. Under traditional paradigms, young learners are forced to expend excessive cognitive energy on decoding written text and analyzing abstract rules, leaving insufficient working memory capacity for real-time speech encoding, lexical selection, and articulatory execution.

This instructional mismatch often results in higher levels of language anxiety, communication avoidance, and low Willingness to Communicate (WTC). Students frequently express reluctance to speak in front of peers, fearing negative evaluation or phonological errors. Consequently, oral proficiency development stalls, characterized by slow speech rates, frequent filled and unfilled pauses, monosyllabic answers, and mispronunciations of English segmental and suprasegmental features.

To overcome these persistent instructional barriers, contemporary applied linguistics highlights the strategic application of Multimodal Task-Based Language Teaching (TBLT). Multimodality—the purposeful integration of visual imagery, graphic organizers, auditory cues, tactile props, digital

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

animation, and kinesthetic-gestural scaffolding—can provide a pedagogical framework suited to young learners. By distributing cognitive processing across multiple sensory modalities (visual, auditory, tactile, and gestural), multimodal tasks diminish cognitive load, facilitate rapid dynamic concept retrieval, and provide rich contextual anchors for oral production.

While an extensive body of international literature documents the efficacy of multimodal instruction in tertiary and secondary language contexts, empirical research investigating its targeted impact on specific speaking sub-dimensions—namely speaking fluency, segmental/suprasegmental pronunciation, and situational WTC—among primary EFL learners in Uzbekistan and the broader Central Asian region remains scarce. To bridge this empirical gap, the present study investigates the exact pedagogical outcomes of integrating structured multimodal visual scaffolding and task-based strategies into primary school EFL classrooms in Yangiyer, Sirdaryo Region.

Specifically, this study addresses three fundamental research questions:

(1) How does multimodal task-based scaffolding affect the oral speaking fluency (measured in words per minute and unfilled speech pauses) of young EFL learners?

(2) What structural improvements are observed in learners' pronunciation accuracy and suprasegmental intonation?

(3) To what extent does multimodal engagement lower affective barriers and enhance learners' situational Willingness to Communicate in classroom settings?

LITERATURE REVIEW. Cognitive Load Theory and Dual-Coding Frameworks in Early EFL. The theoretical rationale for multimodal scaffolding in primary language acquisition is rooted in Cognitive Load Theory [12, 13] and Paivio's Dual-Coding Theory [10]. According to Sweller, human working memory has strict processing limitations when handling novel information. In young L2 learners, speech production involves three concurrent cognitive operations: conceptualization (formulating thoughts), formulation (selecting lexical items and applying syntactic structures), and articulation (phonological execution). When instructional input is purely verbal or text-heavy, working memory becomes overloaded, leading to frequent pauses, speech breakdowns, and cognitive fatigue.

Paivio's Dual-Coding Theory asserts that human memory operates through two independent yet interconnected subsystems: a non-verbal (visual) system and a verbal (linguistic) system. When educational tasks present visual scaffolding—such as color-coded visual grammar cards, situational flashcard sequences, and iconographic prompts—alongside spoken L2 input, learners build dual mental representations. Visual cues can act as immediate mental shortcuts, bypassing the need for explicit L1 translation and enabling young learners to map target words directly to visual concepts. This dual-encoding mechanism may reduce unnecessary cognitive load, allowing working memory capacity to be redirected toward fluent speech planning and phonological refinement.

Multimodal Task-Based Language Teaching (TBLT) for Young Learners.Task-Based Language Teaching [4, 10] posits that language acquisition occurs most effectively when learners are engaged in meaningful, goal-oriented tasks rather than isolated language drills. Cameron [2] emphasizes that for young learners, tasks must be concrete, visually stimulating, and intrinsically motivating. Integrating multimodality into TBLT provides dynamic 'scaffolding'—a concept pioneered by Vygotsky [13] and expanded by Wood et al. [14]—which temporarily supports learners as they perform tasks slightly beyond their independent capacity.

In a multimodal TBLT framework, tasks are structured around interactive visual prompts (e.g., information-gap card trading, story-building cubes, interactive digital visual sequences, and gestural

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

drama prompts). These multimodal entry points provide clear communicative goals (e.g., solving a visual puzzle or telling a story based on cards) rather than focusing on abstract grammatical correctness. As Jewitt [5] notes, multimodality expands the dynamic landscape of the classroom, allowing young learners to leverage visual cues, facial expressions, and physical gestures as communicative bridges when their L2 lexical stores are momentarily exhausted.

Affective Filter and Willingness to Communicate (WTC). Krashen's [6] Affective Filter Hypothesis posits that emotional variables—such as anxiety, self-confidence, and motivation—act as an invisible barrier that can block language input from reaching the brain's acquisition mechanisms. In primary EFL classrooms, high anxiety severely restricts voluntary speech output. MacIntyre et al. [7] conceptualized Willingness to Communicate (WTC) as the operational readiness to enter into discourse at a specific moment using an L2. WTC is dynamic and context-dependent, heavily influenced by classroom climate, task type, and perceived task difficulty.

Multimodal speaking tasks function as a potential affective support. By redirecting student focus away from teacher evaluation and toward hands-on, visually engaging task completion, students experience lower levels of evaluative threat. Interactive visual cards and game-based task cards foster a playful classroom atmosphere where error making is perceived as a natural component of problem-solving rather than a personal failure. Consequently, students' contextual self-confidence increases, sparking higher spontaneous oral participation and sustained communicative eagerness.

Recent pedagogical work in Uzbekistan has highlighted the value of learner-centered and task-oriented approaches, while classroom implementation may still be constrained by class size and limited access to specialized visual materials. Investigating practical, low-cost multimodal scaffolding tools—such as flashcard-based task strategies, color-coded grammar visualizers, and interactive oral routines—can therefore provide useful evidence for strengthening communicative language teaching in local primary schools [1].

METHOD. Research Design and Participant Profile. This study employed a quantitative, quasi-experimental pre-test/post-test control group research design over a 12-week academic period during 2025–2026. The participant sample comprised 60 primary school EFL learners (aged 9–10 years; 32 females, 28 males) enrolled in two intact 4th-grade classes at School No. 1 in Yangiyer, Sirdaryo Region, Uzbekistan.

The two intact classes were assigned to the Experimental Group (EG, n = 30) and the Control Group (CG, n = 30), respectively. Prior to the intervention, a standardized diagnostic assessment (comprising a baseline vocabulary recognition test, simple listening comprehension, and an initial speaking warm-up) confirmed that all participants belonged to the CEFR A1 beginner proficiency level. Independent-samples t-tests revealed no statistically significant differences between the two groups regarding initial English language proficiency ($p > .05$), ensuring baseline homogeneity.

Instructional Interventions and Experimental Procedures. Both groups received identical instructional volume: 3 academic hours of English per week for 12 weeks (totaling 36 hours of classroom instruction). Both cohorts followed the national 4th-grade English curriculum topics (*e.g., Module 1: Animals & Emotions, Module 2: Daily Routines & Time, Module 3: Family, Hobbies & Professions, Module 4: My Classroom & Town*). However, the pedagogical execution differed fundamentally between the two groups:

Control Group (CG - Traditional Method): Instruction relied on traditional textbook-centered methodologies. Lessons followed a presentation-practice-production (PPP) structure dominated by

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

static textbook dialogues, teacher-fronted grammar explanations, translation into L1 (Uzbek), mechanical repetition drills, and individual workbook exercises. Oral practice was largely restricted to reading written dialogues aloud.

Experimental Group (EG - Multimodal TBLT Method): Instruction incorporated structured multimodal speaking tasks and visual scaffolding strategies. Lessons utilized color-coded visual grammar cards, thematic storyboards, audio-gestural prompts, pair-based visual information-gap tasks, and interactive speaking routines ("Magic Animals & Emotions" demo structures). Textbooks were augmented with tactile visual cards, digital visual overlays, and peer-to-peer interactive games.

Table 1. Sample 15-Minute Multimodal Speaking Routine ("Magic Animals & Emotions")

Phase & Time	Multimodal Scaffolding Tools	Student Task & Oral Output
Phase 1: Visual Warm-Up (3 mins)	Digital flashcards + Gestural mimics (Teacher shows happy lion/sad monkey).	Students identify animal and emotion by combining words: 'Happy Lion!', mimicking sound and gesture.
Phase 2: Information-Gap Task (7 mins)	Pair-work visual cards (Student A has animal card, Student B has emotion card).	Students ask/answer in pairs using visual prompts: 'Is your bear angry?' - 'No, my bear is sleepy!'
Phase 3: Story Cube Production (5 mins)	Tactile story cubes + Visual grammar cards (Color-coded Subject-Verb-Adjective).	Students roll cubes and generate 2-3 sentence oral micro-stories: 'The green monster is happy today.'

Data were collected at pre-test (Week 1) and post-test (Week 12) using three measures. First, participants completed a two-minute audio-recorded picture-description task; recordings were analyzed for words per minute (WPM) and unfilled pauses longer than 2 seconds. Second, two trained EFL evaluators independently rated pronunciation on a five-point analytic rubric covering segmental accuracy, word stress, and sentence intonation/rhythm; inter-rater reliability was 0.89. Third, a child-adapted 10-item Willingness to Communicate questionnaire used a five-point Likert scale, supplemented by classroom observations of voluntary oral responses. Data were analyzed in IBM SPSS Statistics 26.0 using descriptive statistics, independent-samples t-tests for baseline equivalence and post-test comparisons, paired-samples t-tests for within-group changes, and Cohen's d for effect sizes; statistical significance was set at $p < .05$.

RESULTS. Statistical analyses confirmed that prior to the intervention, no statistically significant differences existed between the Control Group (CG) and Experimental Group (EG) across any measured parameters.

Table 2. Pre- and Post-Test Descriptive and Inferential Results

Measure	Group	Pre-test M (SD)	Post-test M (SD)	Mean diff.	t	p	d
Fluency (WPM)	CG	28.4 (4.2)	32.1 (4.8)	+3.7	3.12	< .01	0.82
Fluency (WPM)	EG	27.9	44.6	+16.7	13.45	< .001	3.47

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

		(4.5)	(5.1)				
Unfilled pauses (>2 s)	CG	6.8 (1.4)	5.9 (1.2)	−0.9	2.64	< .05	0.69
Unfilled pauses (>2 s)	EG	7.1 (1.5)	2.3 (0.8)	−4.8	15.20	< .001	3.98
Pronunciation (1–5)	CG	2.60 (0.45)	2.95 (0.50)	+0.35	2.85	< .01	0.73
Pronunciation (1–5)	EG	2.55 (0.42)	3.80 (0.48)	+1.25	10.88	< .001	2.76
WTC (1–5)	CG	2.40 (0.50)	2.70 (0.52)	+0.30	2.28	< .05	0.58
WTC (1–5)	EG	2.35 (0.48)	4.10 (0.45)	+1.75	14.62	< .001	3.75

DISCUSSION. The empirical findings of this 12-week study indicate that the experimental group improved more than the control group across the measured speaking outcomes. These results are consistent with the theoretical rationale of multimodal scaffolding and task-based learning.

The experimental group increased its speaking rate by 16.7 WPM, while the control group increased by 3.7 WPM. Unfilled pauses longer than two seconds decreased from 7.1 to 2.3 per two-minute sample in the experimental group. These changes are consistent with the potential contribution of visual and verbal scaffolding to speech planning and retrieval.

The substantial increase in Willingness to Communicate (WTC) highlights the notable affective change facilitated by multimodal TBLT. In primary education, language learning anxiety is heavily tied to fear of negative peer evaluation and formal teacher correction. Multimodal tasks transform the classroom dynamics from a high-stakes evaluative environment into an interactive, game-based problem-solving space. Visual cards, story cubes, and gestural prompts shift student attention away from linguistic self-monitoring toward task achievement. As Krashen's Affective Filter [6] is lowered, students develop psychological safety, leading to spontaneous participation and willingness to risk phonological errors.

In the present context, multimodal task strategies appear to support the development of CEFR A1-level oral communication skills. The findings suggest that contextualized, multisensory scaffolding may be particularly useful when young learners have limited linguistic resources.

PEDAGOGICAL IMPLICATIONS & PRACTICAL RECOMMENDATIONS. Based on the empirical insights gained from this investigation, several actionable recommendations are offered for primary EFL teachers, textbook authors, and curriculum designers in Uzbekistan and similar EFL contexts:

1. Redesign Primary Classroom Materials: Educational authorities and textbook authors should enrich primary school EFL textbooks with visual grammar cards, color-coded sentence builders, and

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

storyboards. Static text-only exercises should be systematically replaced with visual information-gap tasks.

2. Adopt 15-Minute Multimodal Speaking Routines: EFL teachers should incorporate structured, 15-minute multimodal speaking routines (such as the 'Magic Animals & Emotions' model) into daily lesson plans to ensure consistent, low-anxiety oral practice.

3. Emphasize Suprasegmental Audio-Visual Modeling: Pronunciation instruction for young learners should prioritize suprasegmental features (stress, rhythm, intonation) using visual pitch contours, gesture prompts, and audio-rhythmic chants rather than mechanical phonetic drills.

4. Teacher Training and Professional Development: Pre-service and in-service teacher training programs should train educators in multimodal task design, dynamic visual scaffolding techniques, and affective filter management strategies tailored to young learners.

CONCLUSION. Overall, the 12-week quasi-experimental study indicates that multimodal speaking tasks can support young EFL learners' oral development. Compared with the control group, the experimental group showed greater gains in speaking fluency, pronunciation, and Willingness to Communicate. These findings suggest that visual, audio-gestural, and task-based scaffolding can provide useful contextual support for oral interaction. Because the study involved 60 fourth-grade learners in one school, further research with larger and more diverse samples is needed to examine the durability and transferability of these effects.

Ethics and limitations. As the participants were fourth-grade learners, the final submission should state the actual parental/guardian consent, school authorization, and any institutional ethics approval obtained for the study. The single-school sample and 12-week duration also limit the generalizability of the findings.

References

1. Alimurodova, S. (2024). Enhancing Speaking Skills in Young EFL Learners through Multimodal Visual Scaffolding. Journal of Foreign Language Teaching and Applied Linguistics, 8(2), 45-58.
2. Cameron, L. (2001). Teaching Languages to Young Learners. Cambridge: Cambridge University Press.
3. Dörnyei, Z. (2005). The Psychology of the Language Learner: Individual Differences in Second Language Acquisition. Mahwah, NJ: Lawrence Erlbaum Associates.
4. Ellis, R. (2003). Task-based Language Learning and Teaching. Oxford: Oxford University Press.
5. Hasanova, D. (2018). English language teaching in Uzbekistan: Evolution and current trends. Asian EFL Journal, 20(4), 112-129.
6. Jewitt, C. (2008). Multimodality and literacy in school classrooms. Review of Research in Education, 32(1), 241-267.
7. Krashen, S. (1985). The Input Hypothesis: Issues and Implications. London: Longman.
8. MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. The Modern Language Journal, 82(4), 545-562.
9. Mayer, R. E. (2009). Multimedia Learning (2nd ed.). Cambridge: Cambridge University Press.
10. Nunan, D. (2004). Designing Tasks for the Communicative Classroom. Cambridge: Cambridge University Press.
11. Paivio, A. (1986). Mental Representations: A Dual Coding Approach. Oxford University Press.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

12. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257-285.
13. Sweller, J. (2011). Cognitive load theory. Psychology of Learning and Motivation, 55, 37-76.
14. Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.
15. Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89-100.