LINGUATECH
Seguridad, Ética y Excelencia Científica
ISSN 3121-2700

La enseñanza de lenguas basada en tareas y el desarrollo de la expresión oral de nivel A2 en estudiantes de octavo grado

Task-Based Language Teaching and A2 Speaking Development in Eighth-Grade Learners

FRANKLIN DAVID MOYA DAVILA*
https://orcid.org/0009-0003-7023-8188
franklin.moya@casagrande.edu.ec
Universidad Casagrande
Ecuador – Guayaquil

RAFAEL MANUEL FUNES GARCIA
https://orcid.org/0009-0002-2095-7208
rfunesg@est.ups.edu.ec
Universidad Politecnica Salesiana
Ecuador – Guayaquil

Article received: (22/06/2026) - Accepted for publication: 28/07/2026
Conflicts of interest: None to declare.

*Corresponding author

Suggested citation (APA, seventh edition)
Moya Davila, F., Funes Garcia, R. (2026). Task-based language teaching and A2 speaking development in eighth-grade learners. Linguatech, 2(2), pages 48 - 63.

RESUMEN

La enseñanza de lenguas basada en tareas (TBLT) organiza la instrucción en torno a resultados con un propósito que requieren que los estudiantes intercambien significado. Este estudio examinó los cambios en el desempeño oral de nivel A2 después de un programa basado en tareas de ocho semanas. Se utilizó un diseño cuantitativo de un solo grupo con pretest-postest con 28 estudiantes de octavo grado de 12 a 13 años en un colegio privado urbano de Ecuador. Los participantes completaron ocho sesiones de dos horas sobre presentaciones personales, rutinas, descripciones personales, hechos pasados, predicciones, pedidos de comida, direcciones e información de viaje. La expresión oral se evaluó antes y después de la instrucción con tareas de formato A2 Key y tres escalas analíticas de 0 a 5: Gramática y Vocabulario, Pronunciación e Interacción Comunicativa. Gramática y Vocabulario aumentó de M = 3.32 (DE = 0.94) a M = 4.25 (DE = 0.70); Pronunciación aumentó de M = 3.29 (DE = 0.98) a M = 4.29 (DE = 0.66); e Interacción Comunicativa aumentó de M = 3.25 (DE = 0.80) a M = 4.25 (DE = 0.70). Las pruebas de rangos con signo de Wilcoxon fueron significativas en las tres dimensiones tras la corrección de Holm (todas con p ajustada < .001). Todos los participantes mejoraron en el puntaje compuesto de 15 puntos, que pasó de M = 9.86 a M = 12.79. Los resultados indican una mejora intragrupo considerable asociada con la secuencia instructiva. Sin embargo, la ausencia de un grupo de comparación, la falta de reporte de la confiabilidad entre evaluadores y la ausencia de una evaluación diferida impiden establecer conclusiones causales y de retención definitivas.

Palabras clave: enseñanza de lenguas basada en tareas, desempeño oral, estudiantes de nivel A2, interacción oral, educación secundaria

ABSTRACT

Task-Based Language Teaching (TBLT) organizes instruction around purposeful outcomes that require learners to exchange meaning. This study examined changes in A2 speaking performance after an eight-week task-based program. A quantitative one-group pretest-posttest design included 28 eighth-grade learners aged 12-13 at an urban private school in Ecuador. Participants completed eight two-hour lessons on introductions, routines, personal descriptions, past events, predictions, food ordering, directions, and travel information. Speaking was assessed before and after instruction with A2 Key-format tasks and three analytic scales ranging from 0 to 5: Grammar and Vocabulary, Pronunciation, and Interactive Communication. Grammar and Vocabulary increased from M = 3.32 (SD = 0.94) to M = 4.25 (SD = 0.70); Pronunciation increased from M = 3.29 (SD = 0.98) to M = 4.29 (SD = 0.66); and Interactive Communication increased from M = 3.25 (SD = 0.80) to M = 4.25 (SD = 0.70). Wilcoxon signed-rank tests were significant for all dimensions after Holm correction (all adjusted $p < .001$). Every participant improved on the 15-point composite score, which rose from M = 9.86 to M = 12.79. The results indicate strong within-group improvement associated with the instructional sequence. Nevertheless, the absence of a comparison group, unreported interrater reliability, and lack of delayed assessment prevent definitive causal and retention claims.
Keywords: task-based language teaching, speaking performance, A2 learners, oral interaction, secondary education

INTRODUCTION

Speaking is a central objective of foreign-language education because it requires learners to coordinate linguistic knowledge with real-time processing. A speaker must retrieve vocabulary, assemble grammatical forms, produce intelligible speech, interpret an interlocutor, and respond under temporal pressure. These processes are difficult for elementary learners, whose available language is limited and not yet automatic. Classroom instruction can therefore produce an apparent contradiction: students may recognize forms on a worksheet while remaining hesitant when they need to use the same resources in an exchange. Addressing this gap requires instruction that preserves explicit support while creating repeated reasons to communicate.

The Common European Framework of Reference for Languages describes A2 learners as basic users who can manage simple, routine exchanges and communicate about familiar matters when support is available (Council of Europe, 2020). At this level,

successful performance does not require sophisticated discourse. It requires sufficient control of simple language, intelligible pronunciation, and an ability to initiate or maintain brief interaction. These components are closely related but not interchangeable. A learner may have useful vocabulary yet be difficult to understand, or may pronounce familiar words clearly but depend heavily on prompts. A balanced speaking program should therefore integrate form, intelligibility, and interaction rather than equate oral ability with grammatical accuracy alone.

Task-Based Language Teaching (TBLT) offers a framework for creating this integration. In TBLT, tasks function as the primary unit of planning and require learners to use language to achieve an outcome (Ellis, 2003; Nunan, 2004). A task may involve finding differences, planning an itinerary, solving a practical problem, reaching agreement, obtaining information, or performing a service encounter. The defining feature is not the topic or resource but the relationship between language and purpose. Learners must understand and convey meaning in order to complete the activity, and success can be judged through the outcome as well as through the language produced.

This orientation distinguishes a task from an exercise. Exercises normally direct attention toward reproducing a target feature accurately, whereas tasks direct attention toward accomplishing something through language. The distinction does not mean that form-focused work is excluded. Ellis et al. (2020) describe TBLT as capable of incorporating preparation, feedback, and attention to form within a meaning-centered sequence. A teacher can preteach essential vocabulary, provide planning time, observe recurring difficulties, and conduct a post-task language focus. What changes is the organizing logic: linguistic support serves a communicative goal rather than replacing it.

The task cycle is particularly useful for A2 learners. A pre-task phase can activate topic knowledge, clarify the outcome, model useful expressions, and reduce uncertainty. During task performance, learners use available resources to exchange information or create a product. Planning and reporting then encourage them to organize what they want to say, while post-task feedback draws attention to forms that became relevant during communication (Willis & Willis, 2007). This sequence protects meaning while providing structure. For adolescents who are reluctant to speak publicly, rehearsal in pairs or small groups can also reduce the social risk of immediate whole-class performance.

Task difficulty must be carefully calibrated. Skehan (1998) argues that limited attentional capacity creates trade-offs among fluency, accuracy, and complexity. If a task contains unfamiliar content, many steps, and severe time pressure, elementary learners

may devote most of their attention to finding words and lose control of grammar or pronunciation. Conversely, a task that is fully scripted may produce accurate output without meaningful choice. Teachers can manage demand through familiar topics, visual support, role cards, planning time, repetition, and progressive increases in complexity. These adjustments do not remove the task outcome; they make the outcome attainable with the learners' developing resources.

Interaction is one mechanism through which tasks can support development. Collaborative work obliges learners to listen to one another, signal misunderstanding, request clarification, reformulate messages, and confirm decisions. Foster and Ohta (2005) showed that peer interaction involves more than formal negotiation routines; it also includes assistance, co-construction, and mutual support. Such exchanges can expand participation and make gaps in knowledge noticeable. For A2 learners, even short sequences such as asking for directions or ordering food can require multiple turns, responsive listening, and adjustment to a partner's contribution.

Repetition and planning can further stabilize oral performance. Repeating a communicative task reduces the conceptual burden because learners already understand the procedure and content, allowing more attention to formulation and delivery. Research with child foreign-language learners has shown that task repetition can influence negotiation strategies and pair dynamics (García Mayo & Imaz Agirre, 2016). Effective repetition should preserve a genuine purpose while changing a condition, partner, information set, or audience. The learner is then able to reuse language without simply reciting an identical script.

TBLT also aligns with an action-oriented view of language proficiency. The CEFR Companion Volume treats learners as social agents who use language to act with others, not merely as recipients of linguistic content (Council of Europe, 2020). This view is relevant to adolescent classrooms because familiar life domains can be transformed into manageable tasks. Introductions, daily routines, descriptions, past experiences, future plans, food, directions, and travel information provide accessible content while supporting different interactional purposes. These topics can be sequenced from predictable exchanges toward decisions and short narratives.

Assessment must reflect the intended construct. The A2 Key Speaking assessment uses analytic criteria for Grammar and Vocabulary, Pronunciation, and Interactive Communication on a 0-5 scale (Cambridge Assessment English, 2020). The criteria provide a useful classroom framework because they separate linguistic range and control

from intelligibility and responsiveness. At the same time, scores depend on rating judgments, task sampling, and administration conditions. Bachman and Palmer (2010) emphasize that interpretations are strongest when assessment use is aligned with purpose and supported by evidence about consistency. When pretest and posttest ratings involve different evaluators, documentation of training and interrater agreement becomes especially important.

Classroom evaluations of TBLT often operate under practical constraints. Intact groups cannot always be randomly assigned, sample sizes may be small, and the same teacher may design and implement the intervention. East (2017) notes that the translation of TBLT from research into practice is shaped by teachers' beliefs, curriculum demands, and contextual realities. These conditions do not make classroom evidence uninformative, but they require claims to match the design. A one-group pretest-posttest study can identify systematic within-group change and generate a basis for further testing; it cannot, by itself, rule out maturation, test familiarity, concurrent instruction, or rater effects.

Research with school-age learners remains valuable because much task-based literature has focused on adults, university students, or tightly controlled performance tasks. Children and adolescents differ in cognitive maturity, classroom dependence, motivation, and experience with collaborative learning. The scholarship synthesized by Ahmadian and Long (2021) indicates that task-supported interaction can benefit school-age learners when tasks, support, and expectations are developmentally appropriate. Evidence from ordinary secondary classrooms can therefore clarify whether a structured task cycle is feasible and which speaking dimensions appear most responsive over a short instructional period.

The present study examined an eight-week speaking program for A2 eighth-grade learners. Eight lessons placed communicative outcomes at the center of work on familiar topics and were aligned with analytic speaking criteria. The objective was to determine whether learners demonstrated improvement in Grammar and Vocabulary, Pronunciation, and Interactive Communication from pretest to posttest. The null hypothesis stated that the median paired difference would be zero for each dimension; the alternative predicted positive change. A supplementary composite score was calculated to summarize performance across all three dimensions while retaining dimension-specific analyses.

METHODS

Design

A quantitative one-group pretest-posttest design was used. All participants completed an A2 speaking assessment before the intervention, received the same eight-week task-based program, and completed a posttest using the same analytic criteria. The design is commonly described as preexperimental or single-group quasi-experimental because exposure is manipulated but no randomized or non-equivalent comparison group is present. Accordingly, the analysis focuses on within-group change and avoids attributing every observed difference exclusively to TBLT.

The independent condition was participation in eight task-based speaking lessons. The primary dependent outcomes were scores for Grammar and Vocabulary, Pronunciation, and Interactive Communication. Each dimension was scored from 0 to 5. A composite speaking score, ranging from 0 to 15, was derived by summing the three dimension scores for each learner at each measurement point. This composite was supplementary because the analytic dimensions remain pedagogically distinct.

Context and participants

The study was conducted in an urban private school in Ecuador. The accessible population comprised one intact eighth-grade cohort, and all 28 enrolled learners participated. Students were 12-13 years old and had been placed at approximately A2 proficiency through the school's English program. Because the full cohort was included, no random sampling or allocation was performed. Sex-disaggregated participant information was not available in the study records used for this analysis.

Table 1
Study design and participant profile

Characteristic	Description
Design	One-group pretest-posttest
Participants	28 eighth-grade learners
Age range	12-13 years
Initial level	Approximately CEFR A2
Intervention	Eight task-based lessons across eight weeks
Instructional time	Two hours per lesson; 16 hours total
Outcome scale	Three analytic dimensions, each scored 0-5

Note. The intact cohort constituted the full accessible population.

Instructional intervention

The program consisted of eight two-hour lessons delivered over approximately eight weeks. Each lesson was organized around a familiar communicative domain and a practical outcome. Activities used flashcards, images, maps, short videos, worksheets, story prompts, menus, role-play materials, and travel information. The sequence progressed from predictable interpersonal exchanges toward narration, decision making, and information exchange.

Across lessons, the instructional cycle included orientation to the outcome, activation or modeling of useful language, pair or group task performance, and feedback. Grammar, vocabulary, and pronunciation support were embedded where needed, but learners were expected to use language to complete an exchange rather than only reproduce isolated forms. Interactive communication was present throughout the program and was the principal target in lessons on introductions, directions, and travel information.

Table 2

Eight-week task-based speaking sequence

Week	Communicative focus	Principal outcome	Targeted dimensions
1	Greetings and introductions	Exchange personal details and introduce another person	Interactive communication
2	Daily routines	Compare routines and describe current activities	Grammar/vocabulary; interaction
3	Personal appearance	Identify and describe people from visual information	Grammar/vocabulary; interaction
4	Past events	Construct and share a short past-time narrative	Grammar/vocabulary; interaction
5	Predictions and plans	Discuss future intentions and agree on likely outcomes	Grammar/vocabulary; interaction
6	Food and meals	Complete a restaurant role-play and express preferences	Grammar/vocabulary; interaction
7	Directions	Guide a partner to a destination using a map	Interactive communication
8	Travel information	Request and provide information about places and travel	Interactive communication

Note. Each lesson lasted two hours. Pronunciation was addressed through modeling, rehearsal, performance, and feedback across the sequence.

Speaking assessment

Speaking performance was assessed with tasks modeled on the A2 Key for Schools Speaking test and evaluated with the published A2 analytic scales. The scale includes six score points from 0 to 5, with descriptors anchored at selected bands. Grammar and Vocabulary addresses control of simple grammatical forms and the appropriateness and range of language used. Pronunciation addresses control of phonological features and overall intelligibility. Interactive Communication addresses the ability to maintain exchanges, respond, and require an age- and level-appropriate amount of prompting or support (Cambridge Assessment English, 2020).

Baseline scores originated from a mock-examination process conducted under standardized conditions and involving certified examination personnel. Posttest performances were recorded and rated by experienced teachers using the same published rubric. The available records did not report whether the same raters scored both occasions, whether ratings were blind to measurement time, or whether a second independent rating was obtained. These omissions are treated as limitations because changes in rater severity could influence observed score differences.

Procedure and ethics

The baseline assessment was completed before the instructional sequence. Learners then participated in the eight lessons during regular English instruction. The posttest followed the intervention and used the same three analytic criteria. Scores were recorded without names in the analytic dataset. The instructional program was reviewed before implementation for alignment with language-learning theory, CEFR A2 expectations, and the intended assessment criteria.

Institutional authorization and informed permission procedures were completed before participation. Learners received information about the instructional activities, confidentiality, and voluntary participation. They could decline specific activities without academic penalty. Results were handled in aggregate, and no learner, teacher, or institution is identified in this article. Classroom activities were designed to provide equal access, culturally respectful content, and additional scaffolding when required.

Data analysis

Participant-level scores were reconstructed from the pretest-posttest dataset. Means, standard deviations, medians, interquartile ranges, minimums, maximums, and percentages were calculated for each dimension. The percentage of learners scoring in

Bands 4-5 was used as a descriptive indicator of movement toward the upper part of the classroom rating scale. Change direction was summarized as improved, unchanged, or decreased.

Because the outcomes were bounded ordinal ratings and the paired difference distributions departed from normality, Wilcoxon signed-rank tests were used. Tests were two-sided and excluded zero differences from the signed-rank calculation. The family of three dimension-specific tests was controlled with the Holm procedure; adjusted probabilities remained below .001. Standardized paired change was reported descriptively as dz, calculated by dividing the mean difference by the standard deviation of individual difference scores. A supplementary Wilcoxon test was also applied to the 15-point composite. Statistical significance was evaluated at alpha = .05, but interpretation emphasized score patterns, improvement counts, and design limitations rather than p values alone.

RESULTS AND DISCUSSION

Descriptive change

Performance increased in all three analytic dimensions. Grammar and Vocabulary rose from M = 3.32 (SD = 0.94) to M = 4.25 (SD = 0.70), a mean gain of 0.93 points. Pronunciation rose from M = 3.29 (SD = 0.98) to M = 4.29 (SD = 0.66), a one-point gain. Interactive Communication also gained one point, increasing from M = 3.25 (SD = 0.80) to M = 4.25 (SD = 0.70). The median moved from 3 to 4 for every dimension.

Dispersion decreased after instruction. Standard deviations fell by 0.24 for Grammar and Vocabulary, 0.32 for Pronunciation, and 0.10 for Interactive Communication. This pattern indicates that higher posttest means were accompanied by somewhat greater concentration in the upper score bands. No dimension had a posttest score below 3, whereas baseline scores extended to 2. The data therefore show both upward movement and reduced representation at the lower end of the scale.

Table 3

Pretest and posttest descriptive statistics by speaking dimension

Dimension	Time	M	SD	Median	IQR	Min.-Max.	Mean gain
Grammar and Vocabulary	Pretest	3.32	0.94	3	3-4	2-5	
Grammar and Vocabulary	Posttest	4.25	0.70	4	4-5	3-5	0.93

Dimension	Time	M	SD	Median	IQR	Min.-Max.	Mean gain
Pronunciation	Pretest	3.29	0.98	3	2.75-4	2-5	
Pronunciation	Posttest	4.29	0.66	4	4-5	3-5	1.00
Interactive Communication	Pretest	3.25	0.80	3	3-4	2-4	
Interactive Communication	Posttest	4.25	0.70	4	4-5	3-5	1.00
Composite (0-15)	Pretest	9.86	1.65	9	9-11	8-14	
Composite (0-15)	Posttest	12.79	1.26	13	12-14	11-15	2.93

Note. n = 28. IQR = interquartile range. The composite is the sum of the three 0-5 analytic ratings.

Individual improvement patterns

The direction of change was uniformly favorable. In Grammar and Vocabulary, 20 learners (71.4%) improved and 8 (28.6%) remained unchanged. In both Pronunciation and Interactive Communication, 21 learners (75.0%) improved and 7 (25.0%) remained unchanged. No participant decreased in any analytic dimension. Because different learners could improve in different dimensions, all 28 participants increased their composite score by at least one point.

The proportion of learners scoring 4 or 5 also increased substantially. Grammar and Vocabulary moved from 42.9% at baseline to 85.7% after instruction. Pronunciation increased from 42.9% to 89.3%, and Interactive Communication increased from 46.4% to 85.7%. These shifts are pedagogically meaningful because they show that the higher means did not result from a few exceptional gains. Most learners moved into or remained within the upper portion of the classroom scale.

Table 4

Direction of change and upper-band performance

Dimension	Improved n (%)	Unchanged n (%)	Decreased n (%)	Bands 4-5 pretest	Bands 4-5 posttest
Grammar and Vocabulary	20 (71.4%)	8 (28.6%)	0 (0.0%)	12 (42.9%)	24 (85.7%)
Pronunciation	21 (75.0%)	7 (25.0%)	0 (0.0%)	12 (42.9%)	25 (89.3%)
Interactive Communication	21 (75.0%)	7 (25.0%)	0 (0.0%)	13 (46.4%)	24 (85.7%)

Dimension	Improved n (%)	Unchanged n (%)	Decreased n (%)	Bands 4-5 pretest	Bands 4-5 posttest
Composite score	28 (100.0%)	0 (0.0%)	0 (0.0%)	N/A	N/A

Note. Bands 4-5 apply only to the analytic scales. Composite improvement indicates a higher posttest sum.

Inferential results

Wilcoxon signed-rank tests rejected the null hypothesis of zero median change for every dimension. For Grammar and Vocabulary, W = 0 and the unadjusted two-sided p value was .000042. Pronunciation and Interactive Communication each produced W = 0 and p = .000026. Holm adjustment for the three primary tests did not alter the conclusion; every adjusted p value remained below .001. The zero W values reflect the absence of negative differences among learners whose scores changed.

Standardized paired changes were large and nearly identical across the dimensions: dz = 1.30 for Grammar and Vocabulary, 1.30 for Pronunciation, and 1.30 for Interactive Communication after rounding. The composite increased by M = 2.93 points (SD of gain = 1.51), with W = 0, p = .000003, and dz = 1.93. These statistics describe strong within-group change. They should not be interpreted as controlled treatment effects because the design cannot estimate how much change would have occurred without the program.

Table 5

Wilcoxon signed-rank tests and standardized paired change

Outcome	Mean gain	Positive	Zero	Neg.	W	Two-sided p	dz
Grammar and Vocabulary	0.93	20	8	0	0	< .001	1.30
Pronunciation	1.00	21	7	0	0	< .001	1.30
Interactive Communication	1.00	21	7	0	0	< .001	1.30
Composite (0-15)	2.93	28	0	0	0	< .001	1.93

Note. Primary dimension tests remained p < .001 after Holm correction. Positive, zero, and negative refer to paired posttest-pretest differences. dz = mean paired difference divided by its standard deviation.

Interpretation of the findings

The results show a consistent upward pattern across linguistic control, intelligibility, and interaction. This breadth matters because a narrow intervention might improve rehearsed vocabulary without changing pronunciation or responsiveness. Here, the mean gains were similar across the three scales, and no dimension showed a decline for any

learner. The outcome is compatible with the program's integrated design: familiar communicative tasks required learners to prepare language, listen to partners, formulate messages, and respond within the same lesson cycle.

The sequence from introductions to travel information may have supported development through graduated demand. Early lessons relied on predictable personal content and short exchanges. Later lessons required narration, prediction, role performance, route explanation, and information exchange. This progression is consistent with Skehan's (1998) argument that task conditions shape learners' allocation of attention. Familiarity and support can preserve participation at the beginning, while additional reasoning or interactional requirements can be introduced once learners understand the task cycle.

Pronunciation produced the highest posttest mean, although the difference from the other dimensions was small. Repeated oral performance may have increased opportunities to monitor intelligibility and receive feedback. Modeling, rehearsal, and partner response can draw attention to whether a message is understandable without turning pronunciation into isolated mimicry. Nevertheless, the analytic rating does not reveal which phonological features changed. Claims about segmental accuracy, stress, rhythm, or intonation would require recorded-speech analysis or criterion-level rater comments that were not available.

The gain in Interactive Communication is also coherent with task-based principles. Directions, restaurant exchanges, introductions, plans, and travel questions require reciprocal turns rather than individual display. Learners must respond to content, not only produce a memorized sentence. Foster and Ohta's (2005) account of peer assistance helps explain how pair work can make elementary interaction possible: learners co-construct utterances, repeat, prompt, and provide time for one another. Such support should be interpreted as part of developing interactional competence rather than as evidence that a learner failed to perform independently.

Grammar and Vocabulary improved by 0.93 points, slightly less than the one-point gains in the other dimensions. This result challenges the assumption that meaning-centered instruction necessarily neglects form. The lesson sequence repeatedly recycled simple structures and high-frequency vocabulary within different outcomes. After learners encountered a need during task performance, feedback could connect form to a communicative consequence. Ellis et al. (2020) emphasize that TBLT can include

planned and incidental focus on form, particularly when attention is brief and tied to meaning.

The movement into Bands 4-5 is noteworthy but should not be equated directly with certification outcomes. The analytic rubric supports classroom judgment about performance qualities, while an official qualification result depends on standardized administration, trained examiners, test specifications, and score conversion procedures. The present study used A2 Key-format tasks and published descriptors as an instructionally aligned measure. It therefore provides evidence of classroom speaking development on those criteria, not a substitute for an official examination result.

The composite score adds a useful whole-performance summary. Every learner improved, and the mean rose by nearly three points on a 15-point range. Yet summing the dimensions assumes that each contributes equally and that distances between rating bands can be treated as approximately comparable. For this reason, the composite was supplementary and did not replace the analytic results. Reporting both views is more informative: the composite shows broad movement, while the dimensions identify the areas in which that movement occurred.

The very small p values require contextual interpretation. Statistical significance reflects the consistency of paired improvement, not proof that the intervention alone caused it. Without a comparison group, alternative explanations include maturation, increased familiarity with the speaking format, concurrent English instruction, seasonal changes in motivation, and differences between pretest and posttest raters. The fact that no learner declined also raises a measurement question: true broad improvement is plausible, but changes in scoring severity cannot be excluded when interrater reliability and rater blinding are undocumented.

A stronger future evaluation would retain the same task-based sequence while adding a comparison condition. Intact classes could be matched on baseline proficiency, instructional time, and teacher experience. If random assignment were not feasible, an analysis of covariance could compare posttest scores while controlling for baseline differences. Audio recordings should be independently scored by at least two trained raters who are blind to testing occasion, and interrater agreement should be reported for each analytic dimension. Parallel task forms would reduce simple test familiarity.

Retention also remains unknown. The posttest indicates performance after the eight-week sequence, but no delayed measure determined whether gains persisted. A follow-up after four to eight weeks could examine whether learners maintained the language

without continued rehearsal. Transfer tasks would be equally valuable. Students could complete an unfamiliar service encounter or collaborative decision task that was not practiced during the program. Successful performance under new conditions would provide stronger evidence that learning extended beyond the taught scenarios.

The intervention nevertheless has practical value. Eight lessons and 16 hours constitute a feasible unit that can be embedded within a secondary English course. Materials were accessible and did not depend on specialized technology. Teachers can adapt the content to local routines and student interests while preserving the core architecture: a clear outcome, preparation, meaningful exchange, reporting, and feedback. This adaptability is important because TBLT works as a design framework rather than a fixed collection of activities (Bygate, 2020).

For implementation, participation structures deserve deliberate attention. Pair rotation, assigned roles, individual planning notes, time limits, and visible success criteria can prevent confident learners from controlling the task. Brief self-assessment can ask whether the learner conveyed the message, responded to a partner, used target expressions, and repaired misunderstanding. Feedback should identify one or two high-value language points instead of interrupting every error. At A2, communicative confidence and intelligibility can grow when learners experience successful exchanges while still receiving precise guidance.

Overall, the findings support TBLT as a promising approach for elementary adolescent speaking development. The strongest evidence is not any single mean but the convergence of three patterns: most learners improved in every analytic dimension, upper-band participation nearly doubled, and every composite score increased. The design limits causal inference, but the consistency of change justifies a controlled replication and supports continued use of carefully scaffolded speaking tasks in comparable classrooms.

CONCLUSIONS

An eight-week task-based speaking program was associated with substantial improvement in the A2 classroom speaking performance of 28 eighth-grade learners. Mean scores increased from 3.32 to 4.25 in Grammar and Vocabulary, from 3.29 to 4.29 in Pronunciation, and from 3.25 to 4.25 in Interactive Communication. The median rose from Band 3 to Band 4 in every dimension, and the proportion scoring in Bands 4-5 increased to at least 85.7% across all three scales.

Paired analyses confirmed that the pattern was highly consistent. Between 71.4% and 75.0% of learners improved in each analytic dimension, no learner declined, and all 28 improved on the 15-point composite. Wilcoxon tests remained significant after correction for multiple comparisons, and standardized paired changes were large. The results therefore reject the null hypotheses of zero median change and identify the program as a credible instructional approach for further evaluation.

The pedagogical contribution lies in the integration of meaningful outcomes with structured support. Familiar topics enabled learners to practice introductions, routines, descriptions, narratives, plans, service encounters, directions, and travel exchanges. Preparation, modeling, rehearsal, interaction, and feedback linked linguistic form with communicative purpose. The similar gains across grammar/vocabulary, pronunciation, and interaction suggest that task-based work can support a balanced conception of speaking rather than prioritize fluency at the expense of accuracy or intelligibility.

The findings must remain bounded by the design. A single intact cohort, absence of a control group, short duration, unreported rater agreement, and lack of a delayed posttest prevent definitive causal or retention claims. Future studies should use matched or randomized conditions, parallel speaking tasks, blinded dual ratings, interrater reliability statistics, delayed measurement, and transfer tasks. Within these limits, the program provides a feasible model for expanding purposeful oral language use among A2 secondary learners.

REFERENCES

Ahmadian, M. J., & Long, M. H. (Eds.). (2021). The Cambridge handbook of task-based language teaching. Cambridge University Press. https://doi.org/10.1017/9781108868327

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice. Oxford University Press.

Bygate, M. (2016). Sources, developments and directions of task-based language teaching. The Language Learning Journal, 44(4), 381-400. https://doi.org/10.1080/09571736.2015.1039566

Bygate, M. (2020). Some directions for the possible survival of TBLT as a real-world project. Language Teaching, 53(3), 275-288. https://doi.org/10.1017/S0261444820000014

Cambridge Assessment English. (2020). A2 Key for Schools handbook for teachers. https://www.cambridgeenglish.org/images/504506-a2-key-for-schools-handbook-2020.pdf

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment - Companion volume. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

East, M. (2017). Research into practice: The task-based approach to instructed second language acquisition. Language Teaching, 50(3), 412-424. https://doi.org/10.1017/S0261444816000409

Ellis, R. (2003). Task-based language learning and teaching. Oxford University Press.

Ellis, R., Skehan, P., Shintani, N., Lambert, C., & Révész, A. (2020). Task-based language teaching: Theory and practice. Cambridge University Press. https://doi.org/10.1017/9781108643689

Foster, P., & Ohta, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. Applied Linguistics, 26(3), 402-430. https://doi.org/10.1093/applin/ami014

García Mayo, M. P., & Imaz Agirre, A. (2016). Task repetition and its impact on EFL children's negotiation of meaning strategies and pair dynamics: An exploratory study. The Language Learning Journal, 44(4), 451-466. https://doi.org/10.1080/09571736.2016.1185799

Long, M. H. (2015). Second language acquisition and task-based language teaching. Wiley Blackwell.

Nunan, D. (2004). Task-based language teaching. Cambridge University Press.

Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young beginner learners. TESOL Quarterly, 49(1), 115-140. https://doi.org/10.1002/tesq.166

Skehan, P. (1996). A framework for the implementation of task-based instruction. Applied Linguistics, 17(1), 38-62. https://doi.org/10.1093/applin/17.1.38

Skehan, P. (1998). A cognitive approach to language learning. Oxford University Press.

Van den Branden, K. (2016). The role of teachers in task-based language education. Annual Review of Applied Linguistics, 36, 164-181. https://doi.org/10.1017/S0267190515000070

Willis, J., & Willis, D. (2007). Doing task-based teaching. Oxford University Press.