ISSN 1799-2591 (print); ISSN 2053-0692 (online)
Theory and Practice in Language Studies, Vol. 16, No. 11, pp. 3696-3706, November 2026
DOI: https://doi.org/10.17507/tpls.1611.03

The Effects of the Storytelling Teaching Approach on English Speaking Proficiency Among Primary Pupils in China: An Embedded Mixed-Methods Study

Luying Zheng
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Hanita Hanim Ismail*
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Harwati Hashim
Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Zuraina Binti Ali
Centre for Modern Languages, Universiti Malaysia Pahang, Malaysia

***Abstract*—This study examined the effects of the Storytelling Teaching Approach (STA) on Chinese primary pupils' English speaking proficiency and explored how classroom processes supported such development. An embedded mixed-methods quasi-experimental design was employed in a public primary school in China over a 13-week intervention period. Four intact Grade 5 classes participated, with two classes assigned to the experimental group and two to the control group. The quantitative strand was primary and involved a pretest-posttest comparison of pupils' speaking proficiency using an adapted Cambridge English: A2 Flyers speaking test rated analytically across four dimensions: pronunciation, vocabulary usage, grammatical accuracy, and fluency. Because most speaking variables did not satisfy normality assumptions, Mann-Whitney U tests were used for between-group comparisons and Wilcoxon signed-rank tests for within-group comparisons. To explain the quantitative results, qualitative data were collected through classroom observations and semi-structured interviews with selected pupils. The quantitative findings showed that the experimental group significantly outperformed the control group at posttest in overall speaking proficiency as well as in pronunciation, vocabulary usage, grammatical accuracy, and fluency. Both groups improved over time, but the gains in the experimental group were substantially greater. The qualitative findings further indicated that STA supported speaking development through multimodal scaffolding, rehearsal and retelling opportunities, embodied and creative oral production, peer collaboration, perceived progress, and a low-threat classroom environment. The study suggests that STA may be a pedagogically meaningful approach for supporting young learners' oral English development in Chinese primary classrooms.**

***Index Terms*—storytelling, English speaking proficiency, mixed-methods, primary pupils, China**

I. Introduction

Speaking remains one of the most difficult outcomes to develop in primary English education in China. Policy documents now give clear attention to communication and practical language use. However, many classroom routines still reward textbook coverage, vocabulary explanation, grammar accuracy in written work, and examination preparation. This pattern creates a familiar problem for English teachers. Pupils may know words and sentence patterns, but they hesitate when they need to use them orally (Hu, 2005; Pan & Block, 2011; Qi, 2016).

This problem has become more visible under recent reform. The English Curriculum Standards for Compulsory Education (2022 Edition) stresses meaningful language practice and oral communication. The 2021 Double Reduction policy also places renewed responsibility on schools. Schools now need to improve learning quality inside regular lessons rather than depend on homework or off-campus tutoring. In this context, teachers need an oral English pedagogy that works during ordinary class time. They cannot rely only on special projects, extra clubs, or private tutoring (Ministry of Education of the People's Republic of China, 2022; General Office of the Communist Party of China Central Committee & General Office of the State Council of the People's Republic of China, 2021).

The Storytelling Teaching Approach (STA) offers one possible response. In a story-based lesson, language is attached

* Corresponding Author. Email: hanitahanim@ukm.edu.my

© 2026 ACADEMY PUBLICATION

to characters, intentions, problems, actions, and consequences. This feature makes STA different from decontextualized speaking drills. In a drill, pupils may repeat a form without a strong reason to speak. In a story, pupils usually have a clearer reason. They describe what happened, explain what a character wants, continue a plot, or solve a problem. These story demands can move speaking from a display activity toward a meaning-making activity (Bruner, 1991; Egan, 1986).

STA is also suitable for young learners because children often respond well to concrete events and visible actions. A story gives pupils a path to follow. It gives them a beginning, a problem, and an ending. It also gives them emotional cues. These cues can help pupils understand why a character speaks and how a line should sound. For many Grade 5 pupils, this support matters. Pupils do not begin with an empty topic. They begin with a shared story world that they can remember, retell, and perform.

The theoretical case for storytelling is strong. Story activities can provide comprehensible input. They can create opportunities for interaction around meaning. They can also push learners to produce language that is longer and more purposeful than single-word answers (Krashen, 1982; Long, 1983; Swain, 1985). Narrative engagement may further support speaking because pupils do not need to invent all content in real time. They can rely on the plot, pictures, roles, and classroom rehearsal. These resources may reduce the burden of spontaneous oral production (Green & Brock, 2000; Zwaan, 2004).

Nevertheless, evidence from Chinese primary classrooms remains thinner than the policy enthusiasm for activity-based teaching might suggest. Many studies report livelier lessons or stronger interest after storytelling. Such findings are useful, but they do not fully answer whether pupils speak better. Speaking proficiency is not one simple outcome. It includes pronunciation, vocabulary usage, grammatical accuracy, and fluency. A useful classroom study therefore needs to examine these dimensions directly. It also needs to explain how classroom activity leads to change.

The present study addresses this need. The study compares an STA group with a regular-instruction group over 13 weeks. It assesses speaking through an adapted Cambridge English: A2 Flyers speaking test. It then uses observation and interview data to explain the classroom process. The focus is not only whether STA improves speaking. The study also asks how STA supports pupils when they try to pronounce words, retrieve vocabulary, form sentences, and keep speaking.

The study was guided by the following research questions:

RQ1: Are there significant differences in English speaking proficiency between Chinese primary pupils taught through STA and those receiving regular English instruction?

RQ1a: Are there significant differences in pronunciation?

RQ1b: Are there significant differences in vocabulary usage?

RQ1c: Are there significant differences in grammatical accuracy?

RQ1d: Are there significant differences in fluency?

RQ2: How does STA support Chinese pupils' development in pronunciation, vocabulary usage, grammatical accuracy, and fluency in classroom practice?

II. Literature Review

Storytelling has long been used in education because it gives ideas a memorable shape. A story usually has characters, actions, problems, intentions, and outcomes. These elements help children follow meaning in a natural way. Bruner (1991) viewed narrative as a basic mode of meaning-making. Egan (1986) also argued that stories can organize learning content in ways that match children's imagination and cognitive development. In English language teaching, stories connect language with events, feelings, choices, and communication rather than presenting it as isolated words or rules.

The Storytelling Teaching Approach (STA) is not the same as simple story reading. In a traditional story-reading lesson, the teacher may read a text, explain vocabulary, and ask comprehension questions. In STA, the story frames the whole lesson, guiding pupils to listen, predict, rehearse, retell, interact, perform, and reflect, thereby moving the lesson from comprehension to oral use. STA gives them a reason to use English because the story requires them to describe events, express intentions, continue a plot, or solve a problem.

Second language acquisition theory helps explain why STA may support speaking development. Krashen's (1982) Input Hypothesis emphasizes the need for comprehensible input. Storytelling can make input more accessible through pictures, gestures, facial expressions, voice changes, and story events. These resources help pupils understand meaning even when some language is new.

STA also creates space for interaction. Long's (1983) Interaction Hypothesis highlights the value of meaning negotiation in language learning. In story-based lessons, pupils may discuss what happened, choose roles, ask classmates for help, or repair misunderstandings. These small exchanges give pupils chances to use English for real classroom purposes. Swain's (1985) Output Hypothesis further suggests that learners need opportunities to produce language. Through retelling, role-play, dialogue, and performance, pupils must retrieve vocabulary, form sentences, pronounce words, and continue speaking.

A sociocultural view offers another explanation. Vygotsky (1978) emphasized that learning develops through social support and mediated activity. STA provides different forms of support for speaking. Pictures, role cards, sentence frames, peer support, and teacher modeling provide layered scaffolding for pupils' oral production. These supports allow pupils to attempt speech before they can speak fully on their own. With repeated participation, pupils may gradually become more confident and independent speakers.

© 2026 ACADEMY PUBLICATION

Narrative processing theory also helps explain the role of stories in oral production. When pupils understand a story, they build a mental route through the events. Green and Brock (2000) described deep involvement in a story as narrative transportation. Zwaan (2004) also linked comprehension with mental simulation. These ideas are relevant to speaking because young learners often pause not only because they lack words, but also because they are unsure what to say next. A shared story reduces this burden.

The scenario cycle provides a practical structure for STA lessons. A story-based lesson can move through preparation, performance, and debriefing. During preparation, pupils learn the setting, characters, problem, and useful language before rehearsing with partners or groups. During performance, they retell, speak in roles, or act out scenes. During debriefing, the class reviews pronunciation, vocabulary choice, sentence accuracy, fluency, and communication strategies. This cycle connects input, interaction, output, and feedback. It also keeps storytelling from becoming only a pleasant classroom activity without clear language learning goals.

Speaking proficiency is treated as a multidimensional construct in this study. The adapted assessment focuses on pronunciation, vocabulary usage, grammatical accuracy, and fluency. This analytic view is necessary because STA may affect different aspects of speaking in different ways. A single holistic score can show overall improvement, but it may hide specific changes in pupils' speaking proficiency.

Pronunciation is important because clear speech supports intelligibility. Primary pupils may recognize a word in writing but still say it unclearly. STA may support pronunciation through repeated and meaningful exposure. The teacher models target words and expressions during storytelling. Pupils then imitate and practice them during rehearsal and performance. Storytelling also gives pronunciation a communicative purpose. This makes rhythm, stress, and intonation more natural than isolated repetition.

Vocabulary usage is also central to speaking. Many pupils memorize word meanings, but they may not retrieve words quickly in oral tasks. STA can help because story vocabulary is tied to people, objects, places, and actions. Vocabulary becomes part of story events rather than isolated word-list learning. This connection may make vocabulary easier to remember and use when pupils speak.

Grammatical accuracy is another key part of speaking proficiency. Pupils need to produce complete and understandable sentences. STA can support grammar without turning the lesson into grammar drilling. The teacher may provide sentence frames before speaking tasks. Pupils can reuse these patterns during retelling and role-play. The teacher can later address common errors during debriefing.

Fluency refers to the smoothness and continuity of speech. Young learners often pause because they need to search for ideas and language at the same time. STA reduces both demands. The story gives pupils something to say. Rehearsal gives them time to practice how to say it. Repeated retelling may help pupils move from hesitant speech to smoother expression.

Previous studies have reported positive effects of storytelling and digital storytelling on oral English learning. Hwang et al. (2016) found that technology-supported storytelling could support EFL speaking. Nair and Yunus (2021) reviewed digital storytelling studies and reported positive patterns in speaking development. Yang et al. (2022) linked digital storytelling with English speaking and creative thinking. More recently, Yao et al. (2025) found in a systematic review and meta-analysis that oral narrative interventions had positive effects on L2 learning outcomes, particularly speaking, grammar, and vocabulary. Macalister and Thao (2025) also showed that storytelling-based coursebook adaptation could promote young EFL learners' oral production. In the Chinese context, Zheng et al. (2025) reviewed empirical research on storytelling in English language education from 2014 to 2024 and highlighted the need for stronger classroom-based evidence on implementation and learning outcomes. These studies suggest that stories can support speaking by providing context, engagement, and repeated chances for expression.

However, the existing evidence still has limitations. Some studies pay more attention to interest and engagement than to speaking proficiency. Some studies treat speaking as a broad outcome and do not examine pronunciation, vocabulary usage, grammatical accuracy, and fluency separately. Other studies are conducted outside Chinese primary classrooms. This matters because Chinese primary English lessons often face textbook pacing, examination pressure, large classes, and limited time for oral practice.

The present study addresses this gap by examining STA in an ordinary Chinese primary school classroom. It compares pupils taught through STA with pupils taught through regular English instruction. It uses pretest and posttest speaking measures to examine pupils' speaking development. It also uses classroom observations and pupil interviews to explain the results. In this way, the study does not only ask whether STA improves speaking. It also explores how STA helps pupils pronounce more clearly, use vocabulary more appropriately, form more accurate sentences, and speak more fluently.

III. Method

A. Research Design

The study used an embedded mixed-methods quasi-experimental design. The quantitative strand was primary. Pupils in the experimental group and control group completed a speaking pretest and posttest across a 13-week period. The qualitative strand was embedded to explain the classroom processes behind the quantitative pattern (Creswell & Plano Clark, 2018; Morgan, 2007).

This design was selected for practical and methodological reasons. The school had already placed pupils into intact classes. The researcher could not assign pupils randomly at the individual level. A quasi-experimental design was

© 2026 ACADEMY PUBLICATION

therefore suitable and ethically realistic (Dimitrov & Rumrill, 2003; Shadish et al., 2002). The pretest made it possible to examine baseline comparability. The posttest made it possible to compare outcomes after the intervention.

The qualitative strand had a clear explanatory purpose. Test scores could show whether pupils improved, but the scores could not show what happened during lessons. Observation and interview data therefore helped interpret the statistical pattern. These data showed how pupils used pictures, teacher modeling, peer assistance, rehearsal, and role-play when they tried to speak.

B. Participants and Context

The research took place in a public primary school in China during the second semester of the 2024-2025 academic year. The school followed the regular English curriculum. This context allowed the intervention to be examined under ordinary classroom conditions rather than in a laboratory-like setting.

Four intact Grade 5 classes participated in the study. Two classes were assigned to the experimental group (EG), and two classes were assigned to the control group (CG). The total sample included 160 pupils. Each group included 80 pupils. All classes followed the school timetable and studied comparable textbook content.

The qualitative strand was embedded within the same classroom population. Classroom observations were conducted during the intervention. Selected pupils joined semi-structured interviews after the intervention. Purposeful selection helped represent different participation patterns (Patton, 2007). Their comments helped explain how pupils experienced the story-based speaking lessons.

TABLE 1
STUDY OVERVIEW

Aspect	Description
Research design	Embedded mixed-methods quasi-experimental design
Setting	One public primary school in China (Grade 5)
Participants	160 pupils from four intact classes
Grouping	EG: 80 pupils (2 classes); CG: 80 pupils (2 classes)
Duration	13-week intervention
EG condition	Storytelling Teaching Approach (STA)
CG condition	Regular English instruction
Quantitative data	Adapted Cambridge English: A2 Flyers speaking test
Qualitative data	Classroom observations and semi-structured interviews

C. Instruments

Three instruments were used in the study: an adapted speaking test, classroom observation, and semi-structured interviews. The speaking pre-test and post-test generated the main quantitative evidence. The observation and interview data provided explanatory evidence about the classroom process.

Pupils' speaking proficiency was assessed with an adapted Cambridge English: A2 Flyers speaking test. This framework was selected because it is designed for young English learners. It uses visually supported and age-appropriate communicative tasks (Cambridge University Press & Assessment, 2024; Han et al., 2018; McKay, 2005). For the local school context, the researcher made minor procedural adjustments to prompts and instructions. The core elicitation logic was retained.

Pupils' speaking performances were rated across four dimensions: pronunciation, vocabulary usage, grammatical accuracy, and fluency. Each dimension was scored from 0 to 5. The total possible score was 20. An analytic rubric was preferred over a single holistic score because it made the pattern of oral development more visible (Luoma, 2004).

Classroom observation documented how story-based tasks were introduced, scaffolded, rehearsed, and performed. The observation focus included multimodal support, story sequencing, pupil rehearsal, peer interaction, and willingness to speak. Semi-structured interviews were conducted after the intervention. The interviews asked pupils how they perceived story-based speaking tasks, peer support, rehearsal, retelling, and classroom performance (Cohen et al., 2018; Creswell & Poth, 2018). Table 2 summarizes the analytic framework used to score pupils' speaking proficiency.

TABLE 2
ANALYTIC SPEAKING ASSESSMENT FRAMEWORK

Dimension	Operational focus	Score range
Pronunciation	Intelligibility and clarity of spoken production	0–5
Vocabulary usage	Appropriate and sufficient lexical use in context	0–5
Grammatical accuracy	Accuracy and completeness of sentence production	0–5
Fluency	Continuity and smoothness of spoken delivery	0–5
Overall FLSP	Sum of four analytic dimensions	0–20

D. Validity, Reliability, and Trustworthiness

The study established methodological rigor at both quantitative and qualitative levels. A pilot speaking assessment was conducted with 35 pupils from the same schooling stage but outside the final sample. The pilot helped identify procedural issues. It also provided evidence for the quality of the speaking assessment.

© 2026 ACADEMY PUBLICATION

Five experts in English education and language assessment reviewed the adapted speaking test package. The package included picture prompts, task instructions, sample questions, and follow-up prompts. The experts judged whether the materials were age-appropriate, clear, and able to elicit spoken responses from upper-primary pupils. Their feedback confirmed the suitability of the instrument. The researcher then revised several details before the main test.

The analytic rubric was also reviewed. The review checked whether the four dimensions matched the research objectives and whether the descriptors were suitable for primary learners. Inter-rater reliability was then checked during the pilot. Two trained raters scored the same speaking performances independently. Cohen's Kappa values were .844 for pronunciation, .905 for vocabulary usage, .817 for grammatical accuracy, and .810 for fluency. According to Landis and Koch (1977), these values indicate almost perfect agreement.

The qualitative strand used several strategies to strengthen trustworthiness. The researcher triangulated observation and interview data. The observations followed a checklist. The interviews used a semi-structured guide. Parental consent and pupil assent were obtained. All interview records were anonymized (Lincoln & Guba, 1985; Shenton, 2004). The trustworthiness of the qualitative strand was further strengthened through classroom observation records and interviews with nine pupils selected to represent different participation patterns.

E. Intervention Procedure

The intervention lasted for 13 weeks. It was implemented within the normal school timetable. Both groups used the same core textbook, Compulsory Education Textbook: English (PEP Edition), Grade 5, Volume 2. Both groups also covered comparable unit content. This arrangement helped ensure that the main difference between the groups was the teaching approach rather than instructional time or curriculum exposure.

The experimental group received instruction through STA. The lessons drew on teaching-as-storytelling principles and a scenario-cycle logic. At classroom level, the cycle was translated into a sequence of story input, guided comprehension, language rehearsal, retelling, role-based performance, and debriefing. Each lesson aimed to move pupils from understanding a story to speaking about or through the story.

The teacher used textbook content as the foundation. The teacher did not replace the curriculum with unrelated stories. Instead, textbook sections were placed inside a story-based activity chain. For example, Let's try was used for prediction and situation building. Let's talk and Role-play were used for interaction and performance. Let's learn provided vocabulary and sentence-pattern support. Read and write and Let's wrap it up supported consolidation and reflection.

A typical STA lesson began with a contextual opening. The teacher introduced the setting, characters, and problem through pictures, short videos, gestures, and voice variation. Pupils then listened to the story and answered simple comprehension questions. After this stage, pupils practiced key words and sentence frames. They rehearsed with partners or groups. They retold parts of the story. They then performed short role-based tasks or reconstructed the story collaboratively.

The teacher used several forms of scaffolding. The teacher modeled target pronunciation. The teacher displayed key expressions. The teacher gave pupils preparation time. Group members helped one another recall words and plan lines. The teacher used delayed correction so that pupils could finish speaking before receiving feedback. This approach aimed to protect fluency and confidence while still giving attention to accuracy.

The control group followed the regular teaching approach used in the school context. These lessons emphasized vocabulary explanation, grammar instruction, reading comprehension, written exercises, and textbook-based question-and-answer routines. Some oral practice occurred in the control group. However, the lessons did not include the systematic storytelling cycle, role-based performance, or structured rehearsal process used in the experimental group.

TABLE 3
INSTRUCTIONAL CONTRAST BETWEEN THE EXPERIMENTAL AND CONTROL GROUPS

Aspect	EG: STA instruction	CG: Regular instruction
Lesson opening	story context, pictures, problem situation	textbook warm-up / teacher-led review
Input	story-based input with multimodal support	vocabulary and dialogue explanation
Speaking practice	rehearsal, retelling, role-play, performance	repetition, short answers, textbook dialogue reading
Peer interaction	group planning and peer support	limited pair/group work
Feedback	delayed feedback and debriefing	immediate correction / answer checking
Lesson closure	reflection on language and performance	exercise checking / summary

F. Data Analysis

The quantitative data were analyzed using IBM SPSS Statistics 25.0. Descriptive statistics were first calculated for overall English speaking proficiency and for the four speaking dimensions: pronunciation, vocabulary usage, grammatical accuracy, and fluency. Means and standard deviations were used to show pupils' pretest and posttest proficiency.

The Kolmogorov-Smirnov test was then used to examine normality because each group included more than 50 pupils (Massey, 1951). Since most variables did not meet the normality assumption, non-parametric tests were adopted. The Mann-Whitney U test was used to compare the experimental group and the control group at pretest and posttest (Nachar, 2008). The pretest comparison checked baseline equivalence, while the posttest comparison examined post-intervention between-group differences. The Wilcoxon signed-rank test was used to examine within-group changes from pretest to posttest (Sprent & Smeeton, 2007).

© 2026 ACADEMY PUBLICATION

The qualitative data from classroom observations and pupil interviews were analyzed thematically following Braun and Clarke's (2006) thematic analysis procedure. The analysis focused on repeated classroom patterns, including multimodal support, rehearsal, retelling, peer assistance, role-play, and pupils' oral participation. These themes were used to explain how STA supported pronunciation, vocabulary usage, grammatical accuracy, and fluency. The quantitative and qualitative findings were integrated during interpretation following mixed-methods integration principles (Fetters et al., 2013). The test results showed whether pupils improved, while the qualitative data explained how the improvement occurred in classroom practice.

IV. Results

A. Quantitative Findings on English Speaking Proficiency

This section reports the quantitative findings for Chinese pupils' English speaking proficiency. The analysis covered overall speaking proficiency and four dimensions: pronunciation, vocabulary usage, grammatical accuracy, and fluency. The results are presented through descriptive statistics, between-group comparisons, and within-group comparisons. Table 4 shows the descriptive pattern of pupils' pretest and posttest scores.

Table 4
Descriptive Statistics for English Speaking Proficiency

Variable	CG Pretest M (SD)	EG Pretest M (SD)	CG Posttest M (SD)	EG Posttest M (SD)
Pronunciation	2.875 (0.778)	2.870 (0.787)	3.780 (0.927)	4.090 (0.703)
Vocabulary usage	2.881 (0.764)	2.870 (0.770)	3.680 (0.911)	4.130 (0.679)
Grammatical accuracy	2.938 (0.769)	2.940 (0.752)	3.780 (0.941)	4.070 (0.745)
Fluency	2.819 (0.739)	2.810 (0.735)	3.840 (0.882)	4.120 (0.704)
Overall FLSP	11.512 (1.723)	11.488 (1.775)	15.075 (2.604)	16.413 (2.521)

As shown in Table 4, the two groups started from almost the same level before the intervention. In the control group, the pretest mean scores were 2.875 for pronunciation, 2.881 for vocabulary usage, 2.938 for grammatical accuracy, and 2.819 for fluency. The experimental group showed a very similar pattern, with mean scores of 2.870, 2.870, 2.940, and 2.810 respectively. The overall speaking scores were also close. The control group obtained 11.512, while the experimental group obtained 11.488. These values indicate that the two groups had comparable speaking proficiency at the beginning of the study.

After the 13-week intervention, both groups improved. The control group reached posttest mean scores of 3.780 for pronunciation, 3.680 for vocabulary usage, 3.780 for grammatical accuracy, and 3.840 for fluency. The experimental group achieved higher scores in all four dimensions. Its posttest means were 4.090 for pronunciation, 4.130 for vocabulary usage, 4.070 for grammatical accuracy, and 4.120 for fluency. The total score increased to 15.075 in the control group and 16.413 in the experimental group. This pattern suggests that regular instruction was associated with some progress, while the STA group showed a stronger overall gain.

Table 5 presents the between-group comparison results.

Table 5
Summary of Between-Group Results

Variable	Pretest U (p)	Posttest U (p)
Pronunciation	3188 (.966)	2533 (.017)
Vocabulary usage	3176 (.932)	2119.5 (< .001)
Grammatical accuracy	3198.5 (.996)	2572.5 (.025)
Fluency	3176 (.932)	2485 (.010)
Overall FLSP	3194 (.984)	2012 (< .001)

At pretest, the Mann-Whitney U test showed no significant differences between the two groups. The results were non-significant for pronunciation ($U = 3188$, $p = .966$), vocabulary usage ($U = 3176$, $p = .932$), grammatical accuracy ($U = 3198.5$, $p = .996$), fluency ($U = 3176$, $p = .932$), and overall speaking proficiency ($U = 3194$, $p = .984$). These findings confirm that the two groups were statistically comparable before the intervention. This baseline result is important because it makes the posttest comparison more meaningful.

At posttest, the results changed clearly. The experimental group performed significantly better than the control group in all areas of speaking. Significant differences were found in pronunciation ($U = 2533$, $p = .017$), vocabulary usage ($U = 2119.5$, $p < .001$), grammatical accuracy ($U = 2572.5$, $p = .025$), fluency ($U = 2485$, $p = .010$), and overall speaking proficiency ($U = 2012$, $p < .001$). These results indicate that the posttest advantage of the STA group was not limited to one speaking skill. Pupils in the experimental group spoke with clearer pronunciation, used vocabulary more appropriately, produced more accurate sentences, and delivered speech more smoothly.

The strongest posttest difference appeared in vocabulary usage and overall speaking proficiency. This result may be related to the nature of STA. In the experimental group, pupils met target words through story events, characters, pictures, actions, and role-based tasks. They did not only memorize vocabulary from the textbook. They used words to retell scenes, describe actions, and perform roles. This repeated use in meaningful contexts may have helped them retrieve vocabulary more effectively during speaking tasks.

© 2026 ACADEMY PUBLICATION

TABLE 6
SUMMARY OF WITHIN-GROUP RESULTS

Variable	CG Z (p)	EG Z (p)
Pronunciation	-5.511 (< .001)	-6.875 (< .001)
Vocabulary usage	-5.171 (< .001)	-6.865 (< .001)
Grammatical accuracy	-5.512 (< .001)	-6.354 (< .001)
Fluency	-6.856 (< .001)	-6.929 (< .001)
Overall FLSP	-7.474 (< .001)	-7.729 (< .001)

Note. U = Mann-Whitney U statistic; Z = Wilcoxon signed-rank test statistic.

Table 6 shows the within-group comparison results. The control group improved significantly from pretest to posttest in pronunciation ($Z = -5.511$, $p < .001$), vocabulary usage ($Z = -5.171$, $p < .001$), grammatical accuracy ($Z = -5.512$, $p < .001$), fluency ($Z = -6.856$, $p < .001$), and overall speaking proficiency ($Z = -7.474$, $p < .001$). These findings indicate that regular English instruction still contributed to pupils' speaking development during the semester.

The experimental group also showed significant improvement in all dimensions. The gains were significant for pronunciation ($Z = -6.875$, $p < .001$), vocabulary usage ($Z = -6.865$, $p < .001$), grammatical accuracy ($Z = -6.354$, $p < .001$), fluency ($Z = -6.929$, $p < .001$), and overall speaking proficiency ($Z = -7.729$, $p < .001$). The rank results further support this pattern. In the experimental group, most pupils improved in each speaking dimension. For overall speaking proficiency, 79 pupils improved, one pupil remained unchanged, and no pupil showed a decline.

Taken together, the quantitative findings provide a positive answer to RQ1. The two groups began with similar speaking levels, but the experimental group achieved significantly higher posttest scores after the intervention. Although both groups improved, the STA group showed stronger and more consistent gains across pronunciation, vocabulary usage, grammatical accuracy, fluency, and overall speaking proficiency. These results suggest that pupils in the STA group achieved stronger and more consistent speaking gains than those in the regular-instruction group.

B. Qualitative Findings Explaining Speaking Development

The qualitative data helped explain why pupils in the STA group showed stronger gains in English speaking proficiency. Classroom observations and pupil interviews pointed to five connected processes: multimodal scaffolding, rehearsal and retelling, peer collaboration, perceived progress, and a low-threat speaking environment. Together, these processes formed a classroom pathway through which pupils understood the story, prepared language, practiced speech, and spoke with more confidence.

First, multimodal scaffolding gave pupils clearer support before they were asked to speak. In the STA lessons, the teacher used pictures, short videos, gestures, facial expressions, voice changes, and key expressions to introduce the story. These resources helped pupils understand the setting, characters, actions, and problem. Observation data showed that pupils often used pictures or teacher gestures to recall content, suggesting that multimodal input reduced idea-generation burden and connected language with meaning.

This support was especially useful for pronunciation and vocabulary usage. Pupils heard target words several times during storytelling before they had to produce them. The teacher's oral modeling helped them notice pronunciation, stress, and intonation. At the same time, words were linked to visible characters, actions, and events. This made vocabulary easier to recall during retelling and role-play.

Second, rehearsal and retelling gave pupils repeated opportunities to turn input into output. In the STA lessons, pupils did not move directly from listening to public speaking. They first practiced key words and sentence frames. They then rehearsed with partners or groups. After that, they retold parts of the story or performed short role-based tasks. This gradual sequence allowed pupils to try the same language more than once. It also helped them reduce hesitation before speaking in front of others.

The rehearsal process appeared closely related to fluency. Several pupils reported that they could speak more smoothly after practicing the story several times. They also said that the plot helped them remember the order of ideas. The story provided an idea sequence, while rehearsal allowed pupils to prepare the language.

Third, peer collaboration supported pupils during speaking preparation and performance. In group work, pupils discussed story events, divided roles, planned lines, and helped one another recall words or sentence patterns. Stronger pupils often gave quick support to less confident classmates. These small forms of peer help allowed weaker speakers to continue instead of stopping completely.

Peer collaboration also reduced the individual pressure of speaking because pupils prepared with classmates before public performance. As a result, some pupils became more willing to try longer utterances. This process helps explain how STA supported vocabulary retrieval, sentence organization, and speaking confidence at the same time.

Fourth, pupils became aware of their own progress. Interview data showed that many pupils felt they could speak more clearly, use more words, or produce longer sentences than before. Some pupils noticed that they made fewer pauses during retelling. Others felt more comfortable when performing a role. Such perceived progress appeared to encourage further participation.

This sense of improvement matched the quantitative pattern. Pupils' comments also aligned with the four assessed dimensions, including clearer pronunciation, richer vocabulary use, better sentence formation, and smoother speech. These perceptions suggest that the improvement was not only visible in test scores. Pupils also experienced the change in

© 2026 ACADEMY PUBLICATION

classroom practice.

Finally, STA created a lower-threat environment for oral participation. In regular lessons, pupils often had to answer teacher questions individually. This format could make speaking feel risky, especially when pupils were unsure about grammar or pronunciation. In STA lessons, speaking was embedded in stories, roles, group rehearsal, and performance. Pupils could speak as characters and rely on group preparation before public performance.

The interviews showed that several pupils felt less afraid of making mistakes in story-based lessons. Some pupils reported focusing more on the story than on perfect sentences, while others felt safer after group rehearsal. This atmosphere encouraged pupils to attempt speech, sustain speech, and repeat speech. It also helped them take risks that they might avoid in teacher-fronted questioning. Table 7 summarizes the qualitative themes.

TABLE 7
SUMMARY OF QUALITATIVE THEMES

Theme	Classroom Process	Contribution to Speaking
Multimodal scaffolding	Pictures, video, gestures, voice, and key expressions prepared pupils before speaking.	Pupils had clearer content and stronger language support.
Rehearsal and retelling	Pupils practiced story language before performance.	Pupils produced smoother and longer utterances.
Peer collaboration	Group members planned lines and helped each other repair breakdowns.	Less confident pupils had support to continue speaking.
Perceived progress	Pupils noticed clearer and longer speech over time.	Pupils became more willing to participate.
Low-threat atmosphere	Role-play and group performance reduced the risk of speaking alone.	Pupils attempted speech with less fear of mistakes.

Overall, the qualitative findings answer RQ2 by showing how STA supported speaking development in classroom practice. Multimodal scaffolding helped pupils understand the story and prepare language. Rehearsal and retelling gave them repeated chances to practice oral output. Peer collaboration provided immediate support when pupils forgot words or lost confidence. Perceived progress encouraged continued participation. A low-threat environment reduced fear and made speaking more manageable. Collectively, these themes explain how STA supported pupils' development in pronunciation, vocabulary usage, grammatical accuracy, and fluency through linguistic scaffolding, repeated oral production, and a supportive interactional environment.

C. Integration of Quantitative and Qualitative Findings

The quantitative and qualitative findings converged. The posttest results showed that pupils in the STA group performed better than those in the regular-instruction group in overall speaking proficiency and in all four assessed dimensions. The classroom data helped explain this pattern. Repeated teacher modeling and story-based oral practice supported pronunciation. Story contexts helped pupils remember and use vocabulary. Sentence frames, peer rehearsal, and delayed feedback supported grammatical accuracy. Rehearsal and retelling helped pupils speak more smoothly because the story provided a clear sequence of ideas. Taken together, the findings suggest that STA supported speaking through a classroom chain of story input, guided preparation, peer support, speaking performance, and reflection.

V. DISCUSSION

The findings suggest that pupils taught through STA achieved stronger posttest speaking outcomes than those receiving regular instruction. This result is meaningful in the Chinese primary classroom context. In many regular lessons, speaking is still a small part of the lesson. Pupils may repeat after the teacher, read textbook dialogues, or answer short questions. These activities can provide practice, but they do not always give pupils enough time to prepare, organize meaning, and speak for a real purpose. The STA lessons appeared to reorganize this pattern by placing speaking inside a story-based context. Pupils spoke to retell events, play roles, solve problems, and move the story forward. This gave oral English a clearer classroom function.

The posttest results showed that the STA group performed better in pronunciation, vocabulary usage, grammatical accuracy, fluency, and overall speaking proficiency. This broad improvement suggests that STA did not only make pupils more active. It also supported the quality of their speech. The finding is consistent with SLA perspectives which argue that language development needs comprehensible input, interaction, and pushed output (Krashen, 1982; Long, 1983; Swain, 1985). In the STA lessons, these conditions appeared together. Pupils first received story-based input. They then rehearsed with peers. After that, they produced spoken language through retelling and role-play. This sequence gave pupils several chances to hear, practice, and use English.

The improvement across the four speaking dimensions can be interpreted through the classroom processes observed in STA lessons. Pronunciation gains may be linked to repeated oral modeling and meaningful imitation, as pupils heard and practiced key expressions before performance. Vocabulary gains may be related to the story context, where words were connected with characters, actions, and events rather than memorized only as textbook items. Grammatical accuracy may have been supported by sentence frames, model turns, peer rehearsal, and delayed feedback, which allowed pupils to focus on meaning first and form later. Fluency gains appeared to be connected with the story sequence and repeated retelling, which reduced the need to search for ideas and language at the same time. These patterns are consistent with

© 2026 ACADEMY PUBLICATION

work on vocabulary use, scaffolding, and repeated task performance (Nation & Newton, 2009; Gibbons, 2002; Bygate, 2001).

The qualitative findings add depth to the statistical results. They suggest that STA may have supported speaking development through a classroom chain rather than through one single technique. Multimodal input helped pupils understand the story. Rehearsal helped them prepare language. Peer collaboration helped them continue when they forgot words. Performance gave them a reason to speak. Debriefing helped them notice what could be improved. A low-threat atmosphere also mattered. Pupils were more willing to try when they could speak as characters and prepare with classmates. This finding echoes Vygotsky's (1978) view that learning develops through supported participation.

Overall, the findings suggest that STA is useful because it reorganizes the speaking ecology of the classroom. It does not simply add more speaking time. It changes how speaking is prepared, supported, performed, and reflected on. For Chinese primary English classrooms, this point is practical. Teachers do not need to abandon textbook content. They can place textbook vocabulary and sentence patterns inside story-based tasks. They can also use pictures, role cards, sentence frames, group rehearsal, and delayed feedback to make speaking more manageable. The study therefore provides evidence that STA may be a feasible approach for supporting oral English development among primary pupils. Especially when teachers aim to balance meaning, accuracy, fluency, and learner confidence.

Several limitations should be acknowledged. First, the study was conducted in one public primary school in China, and all participants were Grade 5 pupils from a relatively homogeneous instructional context. Therefore, the findings should be interpreted with caution when applied to pupils from other regions, school types, grade levels, or proficiency backgrounds. Second, the intervention lasted 13 weeks, and no delayed posttest was conducted; therefore, the long-term sustainability of pupils' speaking gains remains to be examined.

VI. Conclusion

This study examined STA in relation to Chinese primary pupils' English speaking proficiency. The findings showed that pupils taught through STA achieved significantly better posttest results than those taught through regular English instruction. The advantage appeared in overall speaking proficiency and in all four assessed dimensions: pronunciation, vocabulary usage, grammatical accuracy, and fluency.

The qualitative findings helped explain how this improvement was supported in classroom practice. STA gave pupils a meaningful story context, clear language support, repeated rehearsal, peer assistance, and opportunities for role-based performance. These classroom conditions appeared to help pupils understand what to say and how to say it. They also reduced the pressure of speaking in front of others. This may have contributed to pupils' greater willingness to use English and their ability to sustain oral expression.

The study suggests that STA may be a practical and pedagogically meaningful approach for supporting primary English speaking instruction in China. It does not require teachers to abandon the textbook. Instead, it helps teachers reorganize textbook content into story-based tasks that support oral use. Future studies may examine STA in different regions, grade levels, and school contexts. Longer interventions may also show whether pupils' speaking gains can be sustained over time.

Acknowledgements

The authors gratefully acknowledge the participating school, teachers, pupils, and supervisors.

References

[1] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[2] Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, *18*(1), 1-21. https://doi.org/10.1086/448619

[3] Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Longman.

[4] Cambridge University Press & Assessment. (2024). *Pre A1 Starters, A1 Movers and A2 Flyers handbook for teachers*. Cambridge University Press & Assessment.

[5] Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

[6] Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.

[7] Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

[8] Dimitrov, D. M., & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159-165. https://doi.org/10.3233/WOR-2003-00285

[9] Egan, K. (1986). *Teaching as story telling: An alternative approach to teaching and curriculum in the elementary school*. Althouse Press.

[10] Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, *48*(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

[11] General Office of the Communist Party of China Central Committee, & General Office of the State Council of the People's Republic of China. (2021). *Guanyu jinyibu jianqing yiwu jiaoyu jieduan xuesheng zuoye fudan he xiaowai peixun fudan de yijian* [Opinions on Further Reducing the Burden of Homework and Off-Campus Training for Students in Compulsory Education]. Retrieved March 31, 2026, from https://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html

© 2026 ACADEMY PUBLICATION

[12] Gibbons, P. (2002). *Scaffolding language, scaffolding learning: Teaching second language learners in the mainstream classroom*. Heinemann.
[13] Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701-721. https://doi.org/10.1037//0022-3514.79.5.701
[14] Han, F., Huang, J., & Chen, Q. (2018). Strategic processing of Chinese young English language learners in an international standardized English language test. *Frontiers in Psychology*, *9*, Article 1020. https://doi.org/10.3389/fpsyg.2018.01020
[15] Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language Policy*, *4*(1), 5–24. https://doi.org/10.1007/s10993-004-6561-7
[16] Hwang, W.-Y., Shadiev, R., Hsu, J.-L., Huang, Y.-M., Hsu, G.-L., & Lin, Y.-C. (2016). Effects of storytelling to facilitate EFL speaking using web-based multimedia system. *Computer Assisted Language Learning*, *29*(2), 215-241. https://doi.org/10.1080/09588221.2014.927367
[17] Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
[18] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310
[19] Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
[20] Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*(2), 126-141. https://doi.org/10.1093/applin/4.2.126
[21] Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
[22] Macalister, J., & Thao, L. H. P. (2025). Promoting young EFL learners' oral production through storytelling: Coursebook adaptation in the Vietnamese classroom. *International Review of Applied Linguistics in Language Teaching*, *63*(2), 915-945. https://doi.org/10.1515/iral-2023-0045
[23] Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68-78. https://doi.org/10.1080/01621459.1951.10500769
[24] McKay, P. (2005). *Assessin gyoung language learners*. Cambridge University Press. https://doi.org/10.1017/CBO9780511733093
[25] Ministry of Education of the People's Republic of China. (2022). *Yiwu jiaoyu yingyu kecheng biaozhun (2022 nian ban)* [English Curriculum Standards for Compulsory Education (2022 Edition)]. Retrieved March 31, 2026, from https://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582349487953.pdf
[26] Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76. https://doi.org/10.1177/2345678906292462
[27] Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 13-20. https://doi.org/10.20982/tqmp.04.1.p013
[28] Nair, V., & Yunus, M. M. (2021). A systematic review of digital storytelling in improving speaking skills. *Sustainability*, *13*(17), 9829. https://doi.org/10.3390/su13179829
[29] Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge. https://doi.org/10.4324/9780203891704
[30] Pan, L., & Block, D. (2011). English as a "global language" in China: An investigation into learners' and teachers' language beliefs. *System*, *39*(3), 391–402. https://doi.org/10.1016/j.system.2011.07.011
[31] Patton, M. Q. (2007). Sampling, qualitative (purposeful). In *The Blackwell encyclopedia of sociology*. Blackwell.
[32] Qi, G. Y. (2016). The importance of English in primary school education in China: Perceptions of students. *Multilingual Education*, *6*, Article 1. https://doi.org/10.1186/s13616-016-0026-0
[33] Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
[34] Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*(2), 63–75. https://doi.org/10.3233/EFI-2004-22201
[35] Sprent, P., & Smeeton, N. C. (2007). *Applied nonparametric statistical methods* (4th ed.). Chapman & Hall/CRC.
[36] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.
[37] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
[38] Yang, Y.-T. C., Chen, Y.-C., & Hung, H.-T. (2022). Digital storytelling as an interdisciplinary project to improve students' English speaking and creative thinking. *Computer Assisted Language Learning*, *35*(4), 840-862. https://doi.org/10.1080/09588221.2020.1750431
[39] Yao, T., Xi, X., & Prieto, P. (2025). The effects of oral narrative interventions on second language learning: A systematic review and meta-analysis. *System*, *134*, Article 103835. https://doi.org/10.1016/j.system.2025.103835
[40] Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35-62). Elsevier. https://doi.org/10.1016/S0079-7421(03)44002-4
[41] Zheng, L., Ismail, H. H., Hashim, H., & Ali, Z. B. (2025). Storytelling in English language education in China: A systematic review of empirical research from the past decade (2014–2024). *Forum for Linguistic Studies*, *7*(2), 280–295. https://doi.org/10.30564/fls.v7i2.8314

Luying Zheng received the bachelor's degree in education from Wuchang Institute of Technology, China, and the master's degree in teaching and learning from Taylor's University, Malaysia. She is currently pursuing the Ph.D. degree in education at Universiti Kebangsaan Malaysia, Bangi, Malaysia.

She has been engaged in English language teaching and research. Her research interests include English language teaching, storytelling teaching approach, young learners, speaking development, foreign language speaking motivation, and foreign language

© 2026 ACADEMY PUBLICATION

speaking anxiety.

Ms. Zheng may be contacted at p119571@siswa.ukm.edu.my; her ORCID iD is 0009-0003-5818-0999.

Hanita Hanim Ismail received the Bachelor of Human Sciences degree in English Language and Literature from the International Islamic University Malaysia, the Master of Arts degree in English Literature from the University of Malaya, and the Ph.D. degree in English Literature from Universiti Putra Malaysia. She is currently affiliated with the Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

She has been engaged in teaching, supervision, and research in English literature, literature in education, literacy, and English language education. Her research interests include English literature, teaching English literature, literacy, identity, and language education.

Dr. Ismail may be contacted at hanitahanim@ukm.edu.my; her ORCID iD is 0000-0003-3121-8822.

Harwati Hashim received the Bachelor of Education degree in Teaching English as a Second Language from the University of Malaya, and the Master of Education and Ph.D. degrees in Teaching English as a Second Language from Universiti Kebangsaan Malaysia. She is currently an Associate Professor at the Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

She has been involved in teaching, research, postgraduate supervision, and academic leadership in English language education. Her research interests include technology-enhanced language learning, mobile learning, mobile-assisted language learning, English language teaching, and digital pedagogy.

Associate Professor Hashim may be contacted at harwati@ukm.edu.my; her ORCID iD is 0000-0002-8817-427X.

Zuraina Binti Ali received the Ph.D. degree in Teaching English as a Second Language from Universiti Putra Malaysia.

She has been engaged in English language teaching, research, and postgraduate supervision. Her research interests include Teaching English as a Second Language, vocabulary learning, language education, social sciences, and language teaching and learning.

Associate Professor Ali may be contacted at zuraina@umpsa.edu.my; her ORCID iD is 0000-0002-1527-6320.

© 2026 ACADEMY PUBLICATION