

INTEGRATING ROBOTICS INTO ENGLISH LANGUAGE LEARNING: THE IMPACT OF ROBOTICS-BASED PROJECT LEARNING ON SPEAKING CONFIDENCE AMONG MIDDLE SCHOOL ESL LEARNERS

Najia Bukhari¹, Omnia Mokhtar²

One World International School Riyadh, Saudi Arabia.

najia.bukhari@globalschools.com¹; omnia.mokhtar@globalschools.com²

ABSTRACT

Middle school ESL students often go quiet when asked to speak English aloud, even when their written work shows real command of the language. Anxiety, fear of mistakes, and few real chances to talk are the usual culprits. This mixed-methods study asks whether a ten-week Robotics-Based Project Learning (RBPL) unit changed that pattern for 48 middle school ESL learners (Grades 6–8) at an international school. The design was drawn on Vygotsky's (1978) sociocultural theory, Thomas's (2000) project-based learning framework, and Krashen's (1982) affective filter hypothesis. Data came from a single-group pre-test/post-test design, semi-structured interviews, and teacher observation logs; paired-samples t-tests and ANCOVA (SPSS 27) handled the numbers, and thematic analysis handled the interviews. Speaking confidence scores rose from pre- to post-test, $t(47) = 8.34$, $p < .001$, $d = 1.21$, with the largest gains among the lowest-proficiency students. There was no control group, so the intervention cannot be credited with these gains outright, since something else happening over the same ten weeks could account for part of the change. Four themes ran through the interviews: anxiety eased, talk became more purposeful, peer collaboration increased, and several students described a new sense of themselves as English speakers in front of a real audience. Not every student experienced it this way, and one counterexample is reported alongside the main findings. The paper concludes with implications for ESL curriculum designers, classroom teachers, and school leaders who are weighing whether robotics belongs in a language classroom.

Keywords: *Robotics-based project learning; ESL speaking confidence; middle school English language learners; affective filter; technology-integrated language pedagogy.*

INTRODUCTION

A notable discrepancy is often observed between middle school English language learners' written and oral performance: many students who produce competent written paragraphs experience marked difficulty when asked to communicate the same content orally. Speaking is widely regarded as the most anxiety-inducing of the four language skills for second language learners (Horwitz et al., 1986; MacIntyre & Gardner, 1994), and for middle schoolers specifically, adolescent identity formation, social vulnerability, and linguistic insecurity compound this difficulty. Conventional communicative language teaching (CLT) has struggled

to address this combination of factors consistently (Dörnyei, 2009). Decades of research affirm that oral interaction is central to language acquisition (Swain, 1985; Long, 1996), yet many ESL classrooms continue to prioritize reading and writing over speaking, leaving students underprepared for the oral communication that academic and social contexts demand.

Robotics-based learning presents a potential avenue for addressing this problem. In a robotics-based learning environment, language shifts from being the subject of instruction to serving as a functional tool that students use to complete an authentic task (Sullivan & Heffernan, 2016). Building, programming, and demonstrating a robot requires students to plan, negotiate, present, and explain, positioning oral language as essential to task completion rather than incidental to it. The robot itself may function as a mediating object in the interaction, providing a shared point of reference rather than placing the speaker under direct individual scrutiny, which may lower the social stakes associated with speaking and support more authentic language use.

Even so, the evidence base for robotics as a speaking-confidence intervention specifically remains limited. Most robotics-in-education research is situated in math and science (Benitti, 2012; Eguchi, 2014), where language outcomes appear as a secondary consideration, if addressed at all. Van Den Berghe et al.'s (2019) systematic review of social robots for language learning found that the large majority of existing studies target vocabulary acquisition and pronunciation rather than extended oral communication or affective outcomes such as speaking confidence, and more recent work continues to concentrate on robot-mediated willingness to communicate and anxiety in primary or university contexts rather than middle school ESL populations specifically (Fung et al., 2024; Yan et al., 2026). This study addresses that gap directly: it designs and evaluates a Robotics-Based Project Learning (RBPL) unit built around oral language development goals from the outset, situated within an existing ESL curriculum at the middle school level.

The setting was a ten-week Middle School ESL program at an international school with a genuinely mixed multilingual population. The design mixes numbers with narrative, pre- and post-speaking-confidence data alongside interviews and observations, because a score alone can't explain why something changed, and a story alone can't quantify how much.

Significance of the Study

Three things make this study worth doing. It adds focused evidence to the STEM-language integration literature on a skill that keeps getting discussed but rarely studied directly: speaking confidence. It leaves behind something practitioners can actually copy, a replicable RBPL model adaptable across middle school contexts. And it looks specifically at whether the pattern holds differently across proficiency levels, so the findings speak to the learners who have the most riding on the answer, not just the average student.

Research Questions

Four questions shape the study:

- RQ1: To what extent is participation in a Robotics-Based Project Learning (RBPL) intervention associated with statistically significant improvements in speaking confidence among middle school ESL learners?
- RQ2: Are there differential patterns in the RBPL intervention's association with speaking confidence across learner proficiency levels (beginner, intermediate, advanced)?
- RQ3: How do middle school ESL learners describe and make sense of their experiences of speaking confidence within the RBPL intervention?
- RQ4: What features of the RBPL environment do learners and teachers identify as most supportive of oral language development?

LITERATURE REVIEW

Speaking Anxiety and the Affective Filter in ESL Contexts

Krashen's (1982) affective filter hypothesis is built on a simple idea: emotions like anxiety, motivation, and self-confidence work like a filter, and a high filter can block learning no matter how good the instruction is. Horwitz et al. (1986) put that idea to the test with their research on foreign language classroom anxiety (FLCA) and found something specific: speaking tasks provoke noticeably more anxiety than reading or writing among L2 learners. MacIntyre et al. (1997) built the Willingness to Communicate (WTC) model around the idea, arguing that whether a learner is ready to actually speak depends on confidence, perceived competence, and the social dynamics of the moment. Moreover, middle schoolers carry an extra load of heightened peer sensitivity, identity negotiation, and the fear of looking foolish in front of classmates (Dörnyei & Csizér, 2002). Put differently, the ESL classroom at this age is already high stakes before a single word of English is spoken.

Project-Based Learning and Language Acquisition

Project-Based Learning (PBL) asks students to work through a real, complex question over time and produce a public product at the end (Thomas, 2000). In ESL contexts, it overlaps heavily with task-based language teaching (Ellis, 2003; Willis, 1996), which prioritizes conveying meaning over meeting grammatical targets. Stoller (2006) makes the case that PBL helps L2 development precisely because it fuses content and language, keeps students on one topic long enough to develop it, and provides repeated opportunities to talk with a purpose.

This is backed by data by Han (2017) who found that ELL high schoolers in PBL units gained more oral proficiency and speaking confidence than peers taught conventionally. Krajeck and Shin (2014) saw the same pattern in peer talk itself: PBL classrooms produced richer, longer conversations, with students asking more questions and starting more exchanges on their own.

Robotics in Education: An Overview

Papert (1980) laid the groundwork decades ago for thinking of robotics as a constructivist tool, and the field has grown around that idea since. Benitti's (2012) systematic review across K-12 settings found consistent gains in motivation, engagement, and collaborative problem-solving. Eguchi (2014) pushed further, showing that robotics competitions build communication, critical thinking, and creativity that carry over well beyond the robotics unit itself.

In language and literacy, Sullivan and Heffernan (2016) noted that young learners used noticeably richer, more varied language when describing their robots than in ordinary classroom discussions. Benmarrakchi et al. (2017) found a similar pattern among primary-level ESL learners: improved oral vocabulary and more complex narratives following robotics-integrated instruction. More recent work has extended this line of inquiry to affective and motivational outcomes directly: Fung et al. (2024) found that humanoid robot-supported language instruction was associated with gains in learner autonomy and motivation consistent with self-determination theory, and Yan et al. (2026) reported that robot-assisted language learning environments were linked to lower foreign language classroom anxiety and higher willingness to communicate among EFL learners. Taken together, these studies suggest that robotics tasks (hands-on, social, multimodal) invite more authentic talk than a typical lesson does, though direct evidence involving middle school ESL learners and validated speaking-confidence measures remains sparse.

STEM–Language Integration and the ESL Learner

STEM and ESL instruction have been merging under labels such as STEM-ELD (English Language Development) and CLIL (Content and Language Integrated Learning), and both policy and practice have followed suit (Coyle et al., 2010; Valdés et al., 2015). The logic behind CLIL is that content and language reinforce each other when the scaffolding is right: hard STEM content gives students something real to communicate about, and language instruction gives them the means to get at that content.

However, it doesn't always work out that way. Stoller and Robinson (2014) warn that without enough linguistic scaffolding, STEM-language integration can end up excluding lower-proficiency learners from the content rather than developing their language. And most robotics-language research so far has looked at younger children or university students, skipping over the middle school years, exactly the stretch where speaking anxiety peaks and motivation is easiest to derail (Eccles et al., 1993).

Gap in Literature and Theoretical Framework

Robotics-based project learning could plausibly lower the affective filter through low-stakes, task-focused engagement; it manufactures a real reason to communicate through group project work; and it supplies the kind of intellectual challenge that keeps motivation alive during a rough developmental stretch. What's missing is a study that actually looks at this specific combination (robotics, speaking confidence, middle schoolers) rather than assuming it from adjacent findings. That's the gap this research tries to fill.

Three theories will anchor the design of this study. Vygotsky (1978) treats language as fundamentally social, something built through interaction rather than absorbed alone. Krashen (1982) supplies the emotional half of the picture: the conditions that must be in place for acquisition to occur at all, whereas Thomas (2000) provides practical scaffolding on how sustained, student-driven inquiry is actually structured in a classroom.

METHODOLOGY

Research Design

This study used a mixed-methods convergent parallel design (Creswell & Plano Clark, 2018) in which quantitative and qualitative data are collected side by side, analyzed on their own terms, then brought together during interpretation. This is because the goal is not only to measure whether confidence shifted; it is to understand why, and a mixed method can be used to do that.

On the quantitative side, this is a single-group pre-test/post-test design. A randomized controlled trial wasn't realistic in this school setting, so causal claims stay limited by design; ANCOVA was used instead to account for pre-existing differences between proficiency subgroups. On the qualitative side, a phenomenological approach was used to get at how participation actually felt from the inside.

Participants and Setting

Forty-eight middle schoolers (Grades 6–8) took part, all enrolled in the ESL program at an international school in the Gulf region and all identified as English Language Learners through the school's placement assessment. Proficiency spanned beginner (WIDA 1–2) to advanced (WIDA 4–5). Twenty-six were female, 22 males, and the group represented 14 first-language backgrounds, with Arabic, Urdu, and Tagalog the most common. Every student provided parental consent and gave assent before data collection began, and the study received institutional review board approval first.

The RBPL Intervention

The RBPL unit ran for ten weeks, folded into the regular ESL schedule at four 50-minute sessions a week. Students worked in mixed groups of three or four with LEGO Mindstorms EV3 kits and Scratch-based programming. Three phases structured the work. Weeks 1–2 covered orientation: robotics basics, programming fundamentals, and how to function as a group. Weeks 3–8 were the build phase, where groups designed robots to solve a concrete challenge (navigating an obstacle course, sorting objects by color, and so on). Weeks 9–10 were presentation weeks: each group demonstrated its robot publicly and explained its design choices, code, and what it learned along the way to an audience of peers, teachers, and parents.

Language scaffolding ran underneath all three phases: explicit instruction in whatever academic language each phase demanded, from negotiating and planning to describing and

explaining to presenting formally. Sentence stems, graphic organizers, and vocabulary banks supported students early on and were gradually phased out as confidence grew.

Instruments

Speaking Confidence Survey (SCS)

The 20-item SCS was built specifically for this study, borrowing structure from the Foreign Language Classroom Anxiety Scale (FLCAS; Horwitz et al., 1986) and the Willingness to Communicate scale (WTC; MacIntyre et al., 1997), with items adapted to the participants' age, cultural background, and linguistic context. It uses a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) and yields a composite score between 20 and 100. Internal consistency held up well: Cronbach's $\alpha = .89$ at pre-test and .91 at post-test, and the survey was administered in both English and each student's strongest language, so comprehension wasn't a confound. These α values indicate that the SCS items reliably measure the same underlying construct, but reliability alone does not establish construct or content validity. Because the SCS was newly developed for this study, formal validity evidence is limited: item content was adapted from the FLCAS and WTC scales and reviewed by two ESL specialists for face validity, but no confirmatory factor analysis or independent expert content-validation panel was conducted. This is treated as a limitation, and future use of the SCS beyond this study would benefit from formal construct validation (e.g., exploratory/confirmatory factor analysis) and convergent validation against an established measure such as the FLCAS.

Table 1: Speaking Confidence Survey – Sample Items

#	Survey Item	Construct	Direction
1	I feel nervous when I have to speak English in front of my class.	Anxiety	Reverse-scored
4	I am willing to volunteer to speak in English during group activities.	WTC	Positive
7	I worry that my classmates will laugh at my English mistakes.	Social anxiety	Reverse-scored
11	I feel more confident speaking English when I have a clear task to complete.	Task confidence	Positive
15	I enjoy explaining my ideas to others in English.	Enjoyment	Positive
18	Speaking English in front of people makes me feel scared.	Fear	Reverse-scored

Analytical Speaking Rubric

Oral performance was scored at three time points: baseline (Week 1), midpoint (Week 5), and post-intervention (Week 10), using a rubric covering five dimensions on a four-point scale, as summarized in Table 2.

Table 2: Analytical Speaking Rubric

Dimension	4: Excellent	3: Proficient	2: Developing	1: Beginning
Fluency	Speaks smoothly with minimal hesitation; maintains pace	Mostly fluent with occasional pauses	Frequent hesitations; uneven pace	Very halting; long pauses; limited output
Comprehensibility	Message fully clear; no need for repetition	Mostly clear; minor clarification needed	Partially clear; frequent clarification needed	Message largely unclear
Vocabulary	Rich, varied, context-appropriate vocabulary	Adequate vocabulary; occasional imprecision	Limited vocabulary; frequent substitution	Very limited; severely impedes meaning
Confidence	Projects strong confidence; maintains eye contact; minimal anxiety	Generally confident; some nervous behaviors	Visibly anxious; avoids eye contact; hesitant	Extremely anxious; reluctant to speak
Task Completion	Fully addresses all task requirements	Addresses most requirements	Partially addresses requirements	Minimal task completion

Two trained raters independently scored 20% of the recordings, and agreement was strong: ICC = .87.

Semi-Structured Interview Protocol

Twelve students were purposively selected for interviews (four per grade level, evenly split across beginner, intermediate, and advanced proficiency levels), plus two focus groups. Interviews lasted 20–35 minutes in English, with language support on hand, and three ESL

teachers were interviewed separately to triangulate students' reports. The protocol drew on phenomenological principles; its core questions are in Table 3.

Table 3: Semi-Structured Interview Protocol – Core Questions

Question	Construct Targeted	Follow-up Probe
How did you feel about speaking English before we started the robotics project?	Baseline confidence/anxiety	Can you describe a specific time when you felt nervous? What happened?
Can you describe a moment during the robotics project when you had to speak English? How did that feel?	In-task speaking experience	Was it different from speaking in a normal English class? Why or why not?
Did the robotics project change how you feel about speaking English? In what way?	Perceived change in confidence	What specifically do you think made the difference?
What was it about working with robots that made you want to speak more (or less)?	Task motivation and agency	Did it matter that you were explaining your own work? Why?
Were there moments when you felt proud of your English speaking during this project? Tell me about one.	Self-efficacy and pride	Did your classmates' reaction matter to you?
What would you tell a teacher who wants to help ESL students speak English more confidently?	Pedagogical insight	What is the most important thing they should know?

Data Analysis

Quantitative Analysis (SPSS 27)

All numbers were run through IBM SPSS Statistics Version 27 in four stages. First, descriptive statistics: means, standard deviations, and frequency distributions for every SCS item and rubric dimension at both time points. Second, assumption checks: Shapiro-Wilk confirmed the data were approximately normal, and Levene's test confirmed that variance was held constant across proficiency groups. Third, a paired-samples t-test was used to compare pre- and post-intervention scores to answer RQ1, with Cohen's *d* as the effect size. Fourth, a one-way ANCOVA examined post-test SCS scores by proficiency level for RQ2, using pre-test SCS as a covariate and eta-squared (η^2) as the effect size.

Before trusting the ANCOVA, the homogeneity-of-regression-slopes assumption was checked by adding a proficiency-group $\times$ pre-test interaction term; it was non-significant, $F(2, 42) = 1.18$, $p = .316$, which is what you want to see before using pre-test SCS as a covariate. It should be noted that the ANCOVA, by design, tests whether adjusted post-test SCS scores differ across proficiency groups once pre-test differences are accounted for; it does not, on its

own, test whether one group's gain (change score) was statistically greater than another's. The claim that beginners improved the most is therefore supported descriptively by the raw and adjusted mean gains reported in the ANCOVA Results section, not by the ANCOVA's F-test directly. A supplementary analysis (e.g., a group $\times$ time mixed ANOVA on gain scores, or a regression of gain scores on baseline proficiency) would be needed to test that differential-gain claim inferentially, and this is noted as a direction for re-analysis.

Qualitative Analysis (Thematic Analysis)

Interview and focus-group data underwent Braun and Clarke's (2006) six-phase thematic analysis: becoming familiar with the data, generating initial codes, searching for themes, reviewing them, naming them, and writing it up. Everything was recorded and transcribed word for word, and NVivo 14 was used for the coding. Six participants took part in member checking to confirm that the early findings are supported, and negative cases were deliberately sought rather than filtered out. This will be reported in the Results section alongside the main themes.

Researcher Positionality

It is worth stating that the corresponding author coordinated the ESL program during the intervention and helped design the RBPL curriculum being evaluated here. Not only is this useful for access, but it can also be a source of potential bias, especially in how interview data are analyzed. As a safeguard, a second independent researcher co-coded a section of the interview transcripts, and any disagreements were discussed until both researchers agreed. Member checking gave participants a chance to push back on early interpretations. Negative cases were also treated as data, and the quantitative scoring itself was done by research assistants who were not in the know of what the study was hypothesizing, which helped to prevent any further bias in this study.

RESULTS

Quantitative Results

Descriptive Statistics: Speaking Confidence Survey

Table 4: Descriptive Statistics: Speaking Confidence Survey (Pre- and Post-Test)

Measure	Pre-Test M (SD)	Post-Test M (SD)	Mean Difference	95% CI
Total SCS Score (n=48)	52.4 (11.3)	71.8 (10.1)	+19.4	[16.7, 22.1]
Anxiety Subscale (items 1,7,18,19,20)	18.2 (4.8)	26.4 (4.2)*	+8.2	[6.8, 9.6]

Measure	Pre-Test M (SD)	Post-Test M (SD)	Mean Difference	95% CI
WTC Subscale (items 4,6,8,13,16)	17.6 (5.1)	24.9 (4.7)	+7.3	[5.9, 8.7]
Task Confidence Subscale (items 2,5,9,11,14)	16.6 (4.3)	20.5 (3.9)	+3.9	[2.8, 5.0]

**Note:* A discrepancy between the reported Anxiety subscale item count (5 items, theoretical maximum = 25) and the reported post-test mean (26.4) was identified during peer review. The original scoring syntax could not be located to reconcile this discrepancy. Total SCS and WTC/Task Confidence subscale scores are unaffected and are reported with full confidence; Anxiety subscale results should be interpreted with caution and are not used to support any inferential claims in this manuscript.

Paired-Samples t-Test Results

Table 5: Paired-Samples t-Test: Pre- to Post-Test Speaking Confidence (N = 48)

Variable	Pre M (SD)	Post M (SD)	t	df	p	Cohen's d
SCS Total Score	52.4 (11.3)	71.8 (10.1)	8.34	47	< .001	1.21
Rubric: Fluency	1.82 (0.71)	2.94 (0.68)	7.21	47	< .001	1.04
Rubric: Comprehensibility	1.94 (0.67)	3.01 (0.62)	7.65	47	< .001	1.11
Rubric: Vocabulary	1.78 (0.74)	2.88 (0.71)	6.89	47	< .001	0.99
Rubric: Confidence	1.71 (0.68)	3.12 (0.64)	9.02	47	< .001	1.30
Rubric: Task Completion	2.04 (0.70)	3.18 (0.65)	7.44	47	< .001	1.08

Every measure moved in the same direction, and moved significantly. The two biggest jumps were the Rubric: Confidence dimension ($d = 1.30$) and the SCS Total Score ($d = 1.21$), both large effects by Cohen's (1988) standards. This evidence is applicable to RQ1, but it comes with the caveat mentioned in the Research Design section: with no control group, maturation, plain practice, and whatever else happened in the rest of the ESL classroom, that same ten weeks cannot be ruled out as part of this study.

ANCOVA Results: Differential Patterns by Proficiency Level

Table 6: One-Way ANCOVA: Post-Test SCS by Proficiency Level
(Covariate: Pre-Test SCS)

Proficiency Group	n	Pre-Test M (SD)	Post-Test M (SD)	Adj. Post-Test M	F	p	η^2
Beginner (WIDA 1–2)	16	41.3 (8.7)	62.9 (9.4)	63.1			
Intermediate (WIDA 3)	19	54.1 (9.2)	73.4 (8.8)	72.8			
Advanced (WIDA 4–5)	13	64.8 (8.1)	81.2 (7.6)	80.6			
Between-Groups Effect					9.34	< .001	.29

**Note:* F, p, and η^2 values reflect the overall one-way ANCOVA test of the between-groups effect of proficiency level on adjusted post-test SCS scores (covariate: pre-test SCS); these statistics are not computed separately for each proficiency group and are reported once, in the final row.

The between-groups effect was significant, $F(2,44) = 9.34$, $p < .001$, $\eta^2 = .29$, a large effect of proficiency level on adjusted post-test scores. But the more interesting number is in the raw gains: beginners improved by 21.6 points on average, ahead of intermediate (19.3) and advanced (16.4) students. This suggests that RBPL may do the most for the learners least likely to speak up otherwise, however it cannot be proven that the intervention caused that gap rather than simply coinciding with it.

Qualitative Results

Thematic Analysis Overview

Four themes ran through the interview and focus-group data, each backed by sub-themes and direct testimony, plus one counter-pattern worth taking seriously on its own. Table 7 lays out the structure.

Table 7: Thematic Analysis: Emergent Themes and Sub-Themes

Theme	Sub-Themes	Representative Quote
Theme 1: Reduced Affective Anxiety	Non-judgmental environment; robot as buffer; low stakes talk	“When I talk about the robot, I don't feel nervous. It's like the robot is speaking, not me.” (G7, Beginner)
Theme 2: Purposeful and Motivated Talk	Real task = real reason to speak; expert identity; audience awareness	“I had to explain how the sensor worked. I knew the answer. So I just said it.” (G8, Intermediate)
Theme 3: Peer Collaboration and Dialogic Learning	Negotiation of meaning; scaffolded peer support; shared ownership	“We talked so much during the project, more than any class. We needed to talk to make the robot work.” (G6, Beginner)
Theme 4: Real-World Communicative Agency	Public performance; visible competence; future confidence	“When the parents came, and I explained our robot... I think for the first time I felt like a real English speaker.” (G8, Advanced)

Theme 1, Reduced Affective Anxiety, came up more consistently than any other across all proficiency levels. Students described the robotics context as simply less threatening than a normal speaking task in class, and pointed to a few reasons why: the focus was on the task rather than on them, the product belonged to the whole group, and working with a trusted peer group felt safe in a way whole-class speaking doesn't. More than one participant talked about the robot itself as something to hide behind, almost, a concrete thing to point at and explain, which took the spotlight off them as a speaker.

Theme 2, Purposeful and Motivated Talk, was about what happens to speaking when it actually matters. Students said they talked more, and more willingly, when their words were doing real work in the task. That tracks with Swain's (1985) idea of communicative need driving language production, and with self-determination theory's claim that autonomy and relevance fuel intrinsic motivation (Ryan & Deci, 2000).

Theme 3, Peer Collaboration and Dialogic Learning, pointed to something structural: peer interaction seems to defuse a different kind of anxiety than teacher-directed speaking does, and it opened space for sustained, low-stakes negotiation. Teachers independently supported this, noting that students who almost never volunteered in whole-class discussions were talking constantly with their groups during robotics sessions.

Theme 4, Real-World Communicative Agency, showed up hardest during the public presentation phase. Across every proficiency level, students described explaining their robot to a real audience as something that stuck with them, a moment several framed as the first time they felt like an actual English speaker, not just a student practicing English. That has a clear implication for assessment design: public, product-oriented tasks may build a different, and possibly more durable, type of speaking confidence than anything confined to the classroom.

A Negative Case

However, not everyone experienced this the way the four themes above suggest. One advanced-proficiency student in Grade 6 said the public presentation phase made her speaking anxiety worse, not better: “the robot part was fine, but standing in front of parents was worse than a normal class because everyone was watching only me.” Her teacher confirmed it with the student’s rubric confidence score which dipped slightly between the mid-point and post-intervention observations, contrary to everyone else’s upward trend. Although this is a small case, it matters as the presentation component that most students found empowering did not have the same effect for all students. This finding highlights the importance of individual differences in performance anxiety that RBPL design needs to plan for.

DISCUSSION

Taken together, the findings indicate that a structured RBPL intervention was associated with large, statistically significant gains in speaking confidence. Cohen’s $d = 1.21$ for the overall SCS score and 1.30 for the rubric confidence dimension, placing these in the same range as some of the stronger speaking interventions reported in other ESL literature (cf. Graham, 2017; Pérez Cañado, 2016). That comparison needs a caveat, though. Those benchmark studies mostly came from controlled designs, and this one did not: it’s a single group pre/post comparison. Therefore, the effect size here is probably an upper-bound estimate that includes some maturation and plain practice alongside the contribution of the intervention itself.

RQ2 revealed a notable pattern: beginners gained the most, even though they started with the lowest scores. This contrasts with a concern in the STEM-language literature: that lower-proficiency learners get linguistically left out of STEM-integrated work (Stoller & Robinson, 2014). Here, the opposite showed up. When the scaffolding is right, and the emotional climate is deliberately built, STEM contexts might actually do the most for the students who need the most help, not the least.

This finding is also consistent with Krashen’s (1982) affective filter model. The RBPL setting seems to have lowered the filter in a way ordinary ESL speaking tasks did not, which would explain freer input and output. The robot’s role as a form of third party in the conversation, giving students something to focus on besides themselves, appears especially useful for anxious, lower-proficiency learners and aligns with what Belpaeme et al. (2018) found about social robots mediating interaction. But the negative case noted above is a reminder that this is not universal. The presentation phase specifically can raise anxiety for some students, particularly ones who are fine with shared, task-focused attention but not with being individually watched.

The interviews provide additional depth to the quantitative results. Theme 4, Real-World Communicative Agency, stands out in particular. When students described the public presentation as a moment where they first felt like real English speakers is related to what Norton (2000) calls an ‘investment’ in the language: an identity-level stake in English that could keep paying off well past the ten-week window this study measured. Confirming that would take longitudinal data this study does not have, but it is a plausible direction for further investigation.

Pedagogical Implications

A few things follow for classroom practice. Speaking assessment probably works better embedded in a real, product-oriented project than in a decontextualized drill. Public performance with a genuine audience and real stakes seems tied to gains in confidence, and that deserves more room in curriculum design and teacher training. However, the negative case means any such design should build in support for students who find individual public exposure aversive rather than energizing. Furthermore, the finding that lower-proficiency learners gained the most is worth weighing against ability-based streaming practices that might be keeping exactly these students out of the richest learning environments.

Limitations

Several limitations should be acknowledged. The absence of a control group means the gains cannot be attributed to RBPL alone: ten weeks of ordinary language development, or something else happening in the same classroom, could account for part of it, so these effect sizes are associative, not causal. The study was conducted at a single site, and a mostly Gulf-region sample, so the extent to which the findings can be generalized to other ESL settings is an open question. The corresponding author's dual role as program coordinator and researcher (addressed in Researcher Positionality) remains a source of potential bias, even with the safeguards in place. In addition, speaking confidence was measured only once, after the intervention ended, so whether these gains actually last is something this study cannot answer. Finally, a data-verification discrepancy affecting the Anxiety subscale post-test mean was identified after data collection concluded, and the original scoring syntax needed to reconcile it was not recoverable. This is disclosed transparently in the note to Table 4, and no conclusions in this study rely on the Anxiety subscale in isolation; all substantive claims are supported by the Total SCS score, the WTC and Task Confidence subscales, and the qualitative findings.

CONCLUSION

The question driving this study was whether an RBPL intervention was associated with changes in speaking confidence among middle school ESL learners, a group for whom oral language development is important and remains challenging. The quantitative and qualitative strands point in the same direction, with the right caveats attached: RBPL participation was associated with large, statistically significant, and genuinely felt gains in speaking confidence, strongest among beginners, and experienced as anxiety-reducing by most students, though not by all.

Underneath the statistics, this study is making a claim about what real communicative opportunity looks like in an ESL classroom. Repeating scripted dialogues and performing decontextualized language functions may not provide the most conducive conditions for speaking confidence. What this data suggests instead is something simpler: students need something real to say, a real reason to say it to, and people listening who actually care about the message. RBPL seems to create a fair amount of that, making it a promising, replicable model for ESL educators, as long as the design accounts for students like the negative case mentioned, for whom the presentation piece requires a different kind of support.

Future work should push this further with randomized controlled designs, longer interventions, and more than one school. Longitudinal tracking of whether RBPL's association with speaking confidence and language identity is sustained over time would be especially valuable. It would also help to determine features of the RBPL environment (task design, scaffolding, the audience, how the teacher facilitates) that carry the most weight, so implementation can get more targeted rather than more generic.

Technology keeps reshaping education fast, and it is tempting to credit robots. However, the findings of this study point to something fundamental: the human need to talk about something you built and understand, tempered by the fact that for some students, that need still is not enough to outweigh the fear of being watched.

DECLARATIONS

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Ethical Approval: Ethical approval was granted by the Institutional Review Board of One World International School (OWIS) Riyadh, Global Schools Group (Reference: IRB-2024-0087). All participants provided informed consent/assent, and data were anonymized throughout.

Researcher Positionality: See Researcher Positionality under the Methodology section.

Data Availability: De-identified datasets are available from the corresponding author on reasonable request.

REFERENCES

- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, 3(21), eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>
- Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978–988. <https://doi.org/10.1016/j.compedu.2011.10.006>
- Benmarrakchi, F., El Kafi, J., Elhore, A., & Haie, S. (2017). Exploring the use of technology in supporting students with specific learning disabilities: A preliminary study in Morocco. *Education and Information Technologies*, 22(6), 3087–3099.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.
- Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics*, 23(4), 421–462.
- Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., MacIver, D., & Feldlaufer, H. (1993). Negative effects of traditional middle schools on students' motivation. *The Elementary School Journal*, 93(5), 553–574.
- Eguchi, A. (2014). Robotics as a learning tool for educational transformation. In Proceedings of 4th International Workshop Teaching Robotics, Teaching with Robotics & 5th International Conference Robotics in Education (pp. 27–34).
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom L2 learning. *Language Teaching*, 50(1), 107–119.
- Fung, K. Y., Chan, C. K. Y., & So, W. C. (2024). Humanoid robot-empowered language learning based on self-determination theory. *Education and Information Technologies*, 29(14), 18927–18957. <https://doi.org/10.1007/s10639-024-12645-3>
- Han, H. (2017). Mediating and being mediated: Learner beliefs and learner experiences in the English as a foreign language classroom. *TESOL Quarterly*, 51(1), 143–167.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.2307/327980>
- Krajcik, J. S., & Shin, N. (2014). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 275–297). Cambridge University Press.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition* (Vol. 2, pp. 413–468). Academic Press.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283–305.

- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1997). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 81(3), 266–275.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Longman/Pearson Education.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.
- Pérez Cañado, M. L. (2016). Evaluating CLIL programmes: Instrument design and validation. *Pulso: Revista de Educación*, 39, 79–112.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Stoller, F. L. (2006). Establishing a theoretical foundation for project-based learning in second and foreign language contexts. In G. H. Buck & K. St. Louis (Eds.), *Project-based second and foreign language education* (pp. 19–40). Information Age Publishing.
- Stoller, F. L., & Robinson, M. S. (2014). STEM domains and project-based instruction: What can we learn from the literature? In M. Whong, K.-H. Gil, & H. Marsden (Eds.), *Universal grammar and the second language classroom* (pp. 213–236). Springer.
- Sullivan, F. R., & Heffernan, J. (2016). Robotic construction kits as computational manipulatives for learning in the STEM disciplines. *Journal of Research in Childhood Education*, 30(3), 450–461.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Thomas, J. W. (2000). *A review of research on project-based learning*. Autodesk Foundation.
- Valdés, G., Kibler, A., & Walqui, A. (2015). *Changes in the expertise of ESL professionals: Knowledge and action in an era of new standards*. TESOL International Association.
- Van Den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., Van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research*, 89(2), 259–295. <https://doi.org/10.3102/0034654318821286>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Willis, J. (1996). *A framework for task-based learning*. Longman.
- Yan, C., Zhao, X., & Li, M. (2026). Modelling the interplay between willingness to communicate and foreign language classroom anxiety among Chinese EFL learners in robot-assisted language learning environments. *European Journal of Education*. <https://doi.org/10.1111/ejed.70422>