ISSUES

BRIDGING LANGUAGE GAPS in REAL-TIME: INVESTIGATING UNIVERSITY STUDENTS' SELF-INITIATED USE of SPEECH-TO-TEXT TRANSLATION in ENGLISH LANGUAGE CLASSROOMS

Napattanissa Sangkawong[1], Junifer Leal Bucol[1][a], Rozelda Luciano[2]

[1] Center for Language Teaching and Learning, Walailak University, [2] School of Graduate Studies, Philippine Christian University

Teaching English with Technology

This study investigates the self-initiated use of real-time speech-to-text translation (STTT) tools among Thai university students in English language classrooms, and how they perceive these tools to support their comprehension, motivation, engagement, and participation. Adopting a mixed-methods research design, data were collected from 224 students across various faculties. Quantitative findings indicated a two-construct model comprising comprehension and engagement/participation, each validated through exploratory and confirmatory factor analyses. Students reported that STTT tools supported their comprehension, particularly in vocabulary recognition, teacher clarity, and lecture following, and were associated with active classroom engagement, increased confidence, reduced hesitation, and more collaborative participation. Both scales demonstrated excellent internal consistency (α = .95), with overall favorable perceptions (M=4.19 and SD=0.572). Thematic analysis of open-ended responses suggested that students were motivated by perceived gains in comprehension, vocabulary acquisition, increased confidence, and greater classroom participation. However, reported challenges included translation inaccuracies, technological limitations, accent-related recognition issues, and internet instability. Grounded in cognitive load theory and sociocultural theory, this research highlights the nuanced interplay between learner agency, assistive technology use, and classroom interaction. The findings contribute to the underexplored area of STTT use in low-resource language settings and offer actionable insights for educators and policymakers aiming to integrate assistive technologies into inclusive pedagogical practices. This underscores the transformative potential of STTT tools to bridge linguistic gaps and promote more equitable and effective English language learning experiences in EFL contexts.

1. Introduction

The integration of technology into language learning environments has ushered in a new era of pedagogical possibilities, particularly in the realm of second language acquisition (Ivenz & Poláková, 2024). In Thailand, where English is taught as a foreign language, the challenges of bridging the

---

a Corresponding author

linguistic gap between Thai and English are ever-present (Kawinkoonlasate, 2020). Communicative language teaching methodologies, which emphasize real-world communication and interaction, have gained traction in Thai classrooms, yet students still require additional support to navigate the complexities of English. The increasing recognition of technology as a bridge between formal and informal learning environments has enabled students to actively and effectively engage with target language learning tools both within and beyond the classroom (An et al., 2021).

Lim and Toh (2024) highlight that the integration of technology in language education not only transforms the qualitative nature of learning but also fosters positive learner attitudes toward its use. Speech-to-text translation (STTT) emerges as a promising tool that could potentially provide such support, enabling students to overcome communication barriers and enhance their language learning experience (Lyons, 2016). The effectiveness of collaborative learning, which is a component of communicative language teaching, has been tested in Thailand and proven to be effective in enhancing English communicative competence among students (Rodphotong, 2018). Given the ubiquity of mobile devices and the advancements in speech recognition and machine translation, it is increasingly common for students to independently utilize STTT tools to facilitate their understanding and production of English (Zhu, 2020).

This paper aims to explore the self-initiated use of STTT from English to Thai in real-time among Thai university students during English language instruction. This investigation seeks to understand the extent of tool adoption, the pedagogical purposes for which students employ these tools, and the perceived affordances and limitations associated with their use. Through the examination of these factors, this research aims to illuminate the potential of STTT as a supplementary pedagogical tool for language acquisition, particularly given documented challenges Thai students face in developing oral English proficiency despite years of formal instruction (Dwijayani & Musigrungsi, 2022). The findings aim to contribute to a comprehensive understanding of how such technologies might be effectively integrated into English language instruction to support students' learning needs.

Moreover, findings of this study hold implications for pedagogical practice and language policy development in the EFL context and other low-resource language settings, potentially contributing to more effective and inclusive English language instruction. Through examining the intersection of language and technology, the study provides a foundation for educators and policymakers to make evidence-based decisions regarding the integration of technology in language education, with the objective of optimizing instructional outcomes and cultivating a more engaging and supportive learning environment. Furthermore, through systematic examination of the

affordances and constraints of STTT, educators may design instructional activities and assessment strategies that leverage the potential of these tools while mitigating potential drawbacks, such as over-reliance on technology and the reinforcement of inaccurate language skills (Bensalah, 2024; Choemue & Muljani, 2021). This research contributes to the broader discourse on technology-mediated language learning across diverse educational contexts, informing the development of effective technology-enhanced language learning frameworks in varied educational settings.

2. Literature review

2.1. Evolution and advancement of speech-to-text and machine translation

Speech-to-text (STT) and machine translation (MT) technologies have significantly reshaped human-language interactions, with transformative implications for education, accessibility, and cross-lingual communication. STT converts spoken language into text (Reddy et al., 2023), whereas MT facilitates translation across languages (Kenny, 2018; Wang et al., 2022). These technologies have progressed from rule-based systems to advanced neural architectures and end-to-end models (Stahlberg, 2020).

The origins of STT trace back to mid-20th-century automatic speech recognition, exemplified by Bell Labs' Audrey in 1952 and IBM's Shoebox in 1962, which laid foundational groundwork (Furui, 2010). Significant academic and commercial advancements emerged recently, driven by STT's interdisciplinary potential, fostering innovative systems that redefine educational engagement (Lampou, 2023). Key developmental milestones include the 1990s' adoption of hidden Markov models (HMMs) and dynamic time warping (DTW) algorithms (Hirai & Kovalyova, 2024; Rabiner, 1989), the 2000s' integration of Gaussian mixture models (GMMs) for enhanced noise robustness (Deng & Yu, 2014), the subsequent application of deep learning techniques such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for improved handling of accents and contextual nuances (Ravanelli et al., 2017; Seligman & Waibel, 2019), and the current dominance of transformer-based models (Vaswani et al., 2017) with large-scale pretraining, as seen in OpenAI Whisper (Radford et al., 2023), achieving near-human transcription accuracy across diverse contexts.

Concurrently, machine translation (MT) has undergone parallel advancements, evolving from inflexible rule-based systems requiring extensive linguistic expertise (Z. Yang et al., 2024) to more dynamic frameworks. Initially dominated by rule-based approaches in the late 20th century (Hutchins, 1995; Jolley & Maimone, 2022), MT progressed to statistical machine translation (SMT) in the 1990s and 2000s (Liu et al., 2020). The introduction of neural machine translation (NMT) (Bahdanau et al., 2014)

```
Traditional              Neural Pipeline          End-to-End
Pipeline Approach        (Improved)               Speech Translation

  Speech Input            Speech Input             Speech Input
       |                       |                        |
  ASR: Automatic          Neural ASR              End-to-End
  Speech Recognition          |                   Speech
       |                  Neural MT               Translation Model
  MT: Machine                 |                        |
  Translation            Neural TTS               Translated Text
       |                 (optional)               or Speech Output
  TTS:                        |
  Text-to-Speech         Translated
       |                 Speech/Text Ouput
  Translated
  Speech/Text Ouput
```
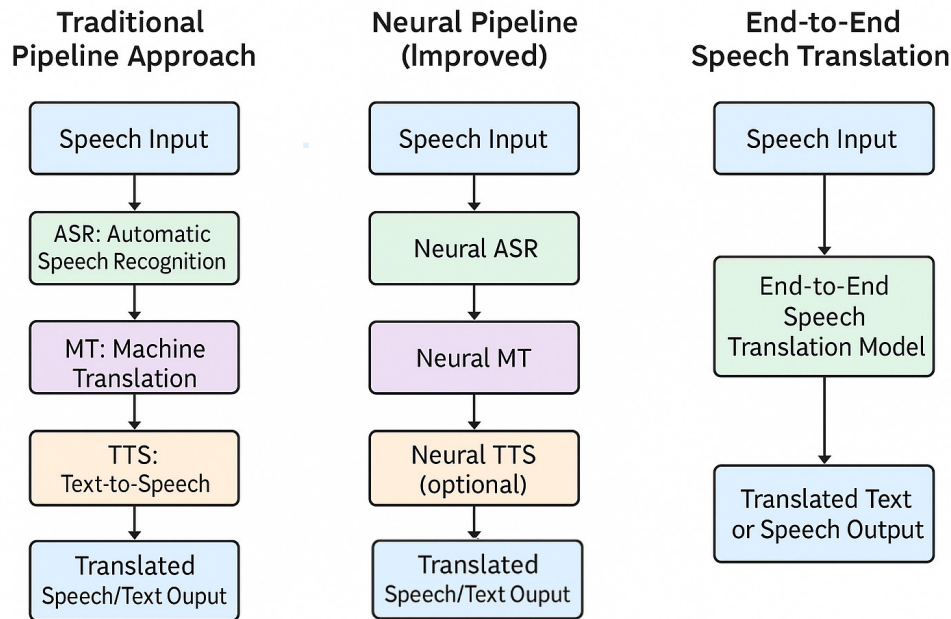
Figure 1. Progression of speech-to-text translation

and sequence-to-sequence (seq2seq) architectures (Sutskever et al., 2014) marked a paradigm shift, enabling fluent and context-sensitive translations (Pipatanakul et al., 2023). Modern end-to-end (E2E) systems, such as Google's Transformer (Vaswani et al., 2017), mBART (Liu et al., 2020), and SeamlessM4T (Barrault et al., 2023), integrate STT and MT into unified frameworks, eliminating intermediate representations (Bérard et al., 2018). This evolution, depicted in Figure 1, underscores key milestones in STT and MT, culminating in advanced end-to-end speech translation systems that facilitate seamless cross-lingual communication (Babatunde et al., 2024; Gao et al., 2023; Lavanya et al., 2024).

2.2. Theoretical frameworks for understanding STTT in language learning

There are several theoretical lenses that can help explain the role of STTT in language learning, offering unique insights into its cognitive, sociocultural, and pedagogical implications. Firstly, the cognitive load theory by Sweller (1988) posits that learning is optimized when extraneous cognitive load is minimized, allowing working memory to focus on schema construction. The STTT tools can reduce intrinsic cognitive load by providing real-time translations, thereby freeing learners to engage in higher-order tasks such as critical analysis and synthesis (Rivera-Trigueros, 2022; Shadiev et al., 2020). However, if the technology is unreliable or difficult to use, forcing learners to expend mental effort decoding errors rather than acquiring language, it may inadvertently increase cognitive load, leading to frustration, disengagement, and reduced learning outcomes (Bahari, 2022; Mayer & Moreno, 2003;

Sweller et al., 2011). This dilemma aligns with the findings that low-proficiency learners may struggle with error correction in STTT outputs, inadvertently increasing cognitive strain (Ivenz & Poláková, 2024).

Secondly, Vygotsky's (1978) sociocultural theory of learning posits that language acquisition is fundamentally a social process that is shaped by interaction and collaboration (Alkhudiry, 2022). Tools like STTT can create more inclusive learning environments by accommodating students with diverse linguistic backgrounds, thus increasing participation and reducing anxiety, which is a crucial factor in language acquisition (Oleksiienko et al., 2020). STTT tools can also act as a mediator, helping learners to understand new concepts and engage more effectively in discussions by removing the linguistic barriers in communication (Hara & Iqbal, 2015; Shadiev & Yang, 2020). Furthermore, STTT can serve as a scaffolding tool that enables learners to perform tasks independently beyond their current capabilities as a key concept of Vygotsky's zone of proximal development (Chancharoen et al., 1999; Lantolf et al., 2014; Samana, 2013), ultimately fostering self-initiated use of learning resources that allows students to proactively engage with learning opportunities, independently seek out support when needed, and develop the critical reflection skills necessary for continued autonomous lifelong learning (Benson, 2013). Self-directed learning using technology could foster learner independence, promote self-reflection, and facilitate personalized language learning experiences that cater to individual needs and learning styles (An et al., 2021; Indriani, 2020; Wei, 2022).

2.3. Pedagogical implications and arguments of STTT in EFL classrooms

The rapid advancements in artificial intelligence (AI) and natural language processing have significantly democratized access to STTT tools, enabling real-time, high-accuracy translation and transcription (Pabba et al., 2024). These transformative technologies hold particular potential in language education, where they effectively bridge comprehension gaps for learners of English as a foreign language by providing instant access to both spoken and written content (Ayyaz, 2025; Yashina et al., 2022). Widely used STTT platforms, such as the speech-translation feature of Google Translate (though Google Translate is considered as MT), have become ubiquitous in classrooms (Ducar & Schocket, 2018), enabling students to decode complex language input in real-time, supporting their comprehension and participation in multilingual discussions.

The integration of STT and MT functions within such platforms reflects the broader trend toward technology-enhanced, learner-directed language learning. A growing ecosystem of freely accessible digital tools, including language-learning applications and AI-powered assistants like ChatGPT, Gemini, Grok and Claude, as well as automated translation services like Google Translate STTT mode, Microsoft Translator, Whisper AI, and DeepL

Translator, offers learners opportunities for individualized learning experience and immediate feedback in relatively low-anxiety environments (Kohnke et al., 2023). Such tools are particularly valuable for providing scaffolded input, a cornerstone of second language acquisition (SLA) theory (Krashen, 1982).

Nevertheless, recent scholarship on the integration of AI-enhanced tools, particularly STTT, in language learning, has yielded mixed results, which raises critical questions in relation to academic integrity, over-reliance and instructional scaffolding and the development of genuine language proficiency (Godwin-Jones, 2023; Zhu, 2020). Some studies suggest that real-time translation can improve listening and reading comprehension by providing learners with immediate access to meaning (Golonka et al., 2012; Hwang & Fu, 2018). For instance, students in EFL classrooms have reported that STTT tools help them understand lectures and participate more actively in discussions (Iberahim et al., 2023). Moreover, the integration of technology into language teaching offers opportunities for authentic language learning experiences among learners (Syathroh et al., 2021).

While these tools offer potential benefits, the use of STTT in language learning is not without challenges. For instance, Ivenz and Poláková (2024) pointed out the risk that students may over-rely on them, hindering the development of crucial language skills, such as grammar, vocabulary, and critical reasoning. In addition, critics contend that excessive dependence on translation may limit opportunities for L2 immersion, consequently slowing interlanguage development and hindering authentic communicative practices (Novikov & Kiseleva, 2024; Reitmaier et al., 2022; J. Yang, 2025). Furthermore, the prevalence of error in translation remains the major issue in utilizing STTT that could pose substantial pedagogical challenges. Although sophisticated algorithms have reduced error rates, inaccuracies persist and can potentially mislead learners, particularly in nuanced or idiomatic expressions (Melkonyan & Matevosyan, 2020). As such, educators must carefully weigh the affordances and constraints of STTT, integrating these tools thoughtfully into pedagogical designs that promote critical thinking and foster genuine language acquisition (Schenck, 2024). Other studies also highlight the importance of using multiple machine translation tools to cross-validate the outputs and improve the reliability of translations (Chang, 2022; Chung & Ahn, 2021; Gritsay & Vodyanitskaya, 2021).

Generally, the utilization of these technologies is affirmed by researchers noting that STTT have great potential for helping students with English language acquisition (Dennis, 2024; Golonka et al., 2012; Wood et al., 2017; Zhang & Zou, 2020). This reinforces the pedagogical impact that STTT can have in enhancing language learning experiences. Consequently, the use of STTT is not merely a tool for language translation but a means of enhancing cognitive processing, promoting social inclusion, and embracing linguistic diversity (Mohammed & Khalid, 2025).

3. Methodology

3.1. The aim of the study

Although existing research has provided valuable insights into the use of STTT tools in language learning, important gaps remain. First, most studies have focused on high-resource languages such as English, Spanish, and Chinese, while low-resource languages like Thai have received limited attention (Lyons, 2016; Pipatanakul et al., 2023). Second, research on students' self-initiated use of STTT tools in real-time classroom settings, particularly in EFL contexts is scarce (Singh & Yunus, 2021). This study addresses these gaps by examining how Thai university students employ STTT tools in real time within English language classrooms. Specifically, it seeks to answer the following research questions:

1. To what extent does students' real-time, self-initiated use of STTT tools relate to their comprehension and engagement in English lessons?

2. What motivates university students to use STTT tools during English language classes?

3. How does the use of STTT tools influence students' engagement and participation in classroom activities?

4. What challenges do students encounter when using STTT tools to support their English language learning?

3.2. Participants and the context

Participants were second-year undergraduates enrolled in a General English course at a government university in Thailand. Out of the 305 registered students in the course, 243 (79.7%) responded and consented to the online survey. After removing 19 incomplete or ineligible responses, only 224 complete responses were retained for analysis, yielding 73.4% of the total cohort, and a completion rate of 92.2% among those who started the survey. English proficiency was determined via institutional placement tests aligned with the Common European Framework of Reference for Languages (CEFR), ranging from A1 to B1. Participants were aged 18–21 and represented diverse academic disciplines across multiple faculties. Gender distribution consisted of 143 females (63.8%), 75 males (33.5%), and 6 undeclared (2.7%). In terms of device usage, 89.9% accessed STTT tools via smartphones, 66.1% via tablets, and 5.8% via laptops. Most students used Google Translate (89.2%), followed by ChatGPT (22.3%) and other translation tools (8.9%). Participants rated their frequency of STTT tools use on a 5-point Likert scale, ranging from 1 ("Never") to 5 ("Always"). The mean frequency score was 3.98 (SD = 0.78), indicating that most students used STTT tools "Often" during English classes.

3.3. Research design and data collection procedure

This study adopted a mixed-methods research design to provide a comprehensive investigation of Thai university students' self-initiated use of STTT tools in English language classrooms in real-time. This design was utilized for a holistic examination of the phenomenon under investigation, providing both statistical insights and understanding of the experiences and perspectives of the participants in using STTT tools. This approach enhanced the validity and depth of the findings by enabling data triangulation quantitatively and qualitatively (Johnson et al., 2007; Weyers et al., 2014). The target population consisted of students who have experienced using STTT tools in real-time class instructions.

Data collection was conducted utilizing the purposive sampling approach through an online survey that integrates both quantitative and qualitative elements using Google Forms. The survey instrument was developed in English. The questions were reviewed and refined based on the feedback by two experts in the field of technology-assisted language teaching before they were translated into Thai by a professional bilingual translator. A second, independent bilingual translator, blind to the original English draft, back-translated the Thai version into English. Then, the research team conducted a line-by-line comparison of the back-translated items against the original English questionnaire, identifying and resolving discrepancies through consensus discussions to preserve conceptual equivalence and clarity. Consequently, the original English items and their Thai translations were merged into a single bilingual Google Form, allowing participants to view each question in both languages side by side.

The questionnaire components included the consent form with the agreement to voluntarily join the survey. The survey began with the demographics, followed by 18 Likert-scale items with 1 as "Strongly disagree" and 5 as "Strongly agree", designed to capture the perceived impact on comprehension, and class participation and engagement in utilizing the STTT tools in language learning. As the questionnaire was just developed for this particular study, to assess the consistency and reliability of the quantitative items, it was pilot-tested with 30 senior students who also experienced using STTT tools. Internal consistency was confirmed with a Cronbach's alpha of $\alpha=0.97$. The qualitative component consisted of open-ended questions embedded in the questionnaire aimed at eliciting detailed insights into students' motivations, experiences, and perceived challenges associated with using these tools.

All quantitative analyses were conducted using Jamovi Version 2.7, an open-source statistical platform built on the R environment. The R syntax exported from Jamovi for all analyses is available in the supplementary materials to ensure transparency and reproducibility (see *Data Code Availability*). Exploratory factor analysis (EFA), confirmatory factor analysis

(CFA), and reliability testing were performed to validate the data before conducting descriptive analysis to summarize central tendencies and variability across responses. Open-ended responses were subjected to thematic analysis, following Braun and Clarke's six-phase approach (Terry et al., 2017), which ensured a systematic and transparent approach to identifying meaningful patterns in students' narratives. Initial familiarization involved repeated reading of all responses, followed by the generation of preliminary codes grounded in sociocultural and cognitive load theory constructs. Themes were then iteratively reviewed and refined, with attention to how STTT tools functioned in specific classroom tasks. To ensure rigor and reliability, a draft codebook was developed and piloted on 30 randomly selected responses, holding calibration sessions to align code definitions. Two independent coders applied the finalized codebook across the full dataset, yielding Cohen's κ values between 0.74 and 0.78 for the high-level themes, indicating substantial agreement (Sim & Wright, 2005). Subsequent phases involved grouping related codes into overarching themes, defining and naming each theme, and producing a narrative report. Ethical approval for this study was obtained and is detailed in the *Ethics Declaration* at the end of the paper.

4. Analyses, findings and discussion

Following established guidelines for scale validation (Hurley et al., 1997; Kyriazos, 2018), a two-phased sequential analytical approach was employed to validate the factorial structure and psychometric properties of the scale. The process consisted of EFA to identify the underlying dimensional structure, followed by CFA to verify the stability and validity of extracted factors.

Prior to conducting the EFA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was computed to assess the appropriateness of the data for factor extraction. The overall KMO value was .952, substantially exceeding the recommended threshold of .60 and indicating excellent sampling adequacy. Individual item-level measures of sampling adequacy (MSA) ranged from .936 to .973, confirming that all variables contributed meaningfully to the correlation matrix. Bartlett's test of sphericity yielded a statistically significant result, $\chi^2(153) = 2793$, $p < .001$, verifying the presence of sufficient intercorrelations among items to warrant factor analysis. Principal axis factoring with Oblimin rotation was employed to extract latent factors from the correlation matrix. The analysis revealed a two-factor structure with eigenvalues exceeding 1.0. Factor 1, designated as Comprehension Dimensions, accounted for 31.6% of the total variance (eigenvalue = 5.68), comprised ten items with standardized factor loadings ranging from .562 to .846, capturing students' cognitive understanding, linguistic processing capabilities, and cognitive efficiency in comprehending instructional content. Factor 2, designated as Engagement and Participation, explained 27.8% of the variance (eigenvalue = 5.00), encompassed eight items with loadings ranging from .511 to .903, reflecting students' affective and behavioral involvement

in the learning process. Collectively, these two factors accounted for 59.3% of the cumulative and shared variance among items. The Oblimin rotation revealed a substantial inter-factor correlation of r = .731 between Comprehension Dimensions and Engagement and Participation, suggesting that while the two constructs are theoretically distinct, they share considerable common variance. Model fit indices demonstrated adequate to good fit for the two-factor solution: RMSEA = .0634 (90% CI [.0509, .0763]), TLI = .947, χ^2 (118) = 225, p < .001. These results provided preliminary empirical support for a two-dimensional measurement structure distinguishing between cognitive comprehension processes and affective-behavioral engagement outcomes.

Following the EFA, a confirmatory factor analysis was conducted to validate the hypothesized two-factor measurement model. Maximum likelihood estimation was employed to assess the factor loadings and overall model fit. All ten indicators measuring the Comprehension Dimensions factor demonstrated statistically significant standardized factor loadings ranging from .495 to .623 (all p < .001), with critical ratio values (Z) ranging from 10.9 to 13.9. The eight indicators comprising the Engagement and Participation factor similarly exhibited significant loadings ranging from .590 to .650 (all p < .001), with critical ratios between 12.7 and 15.7. The consistently high critical ratios across all 18 indicators confirmed that each item contributed meaningfully to its designated construct.

The covariance between Comprehension Dimensions and Engagement and Participation was estimated at 0.788 (SE = .0308, Z = 25.6, p < .001), corroborating the strong positive association observed in the EFA and supporting the theoretical premise that enhanced comprehension facilitates increased engagement and participation. Multiple fit indices were examined to evaluate model adequacy: CFI = .949, TLI = .942, RMSEA = .0680 (90% CI [.0563, .0795]), SRMR = .0410, χ^2 (134) = 273, p < .001. The CFI and TLI values exceeded the recommended threshold of .90, the SRMR fell well below the cutoff of .08, and the RMSEA remained within acceptable bounds, collectively indicating adequate to good model fit (Hurley et al., 1997; Kyriazos, 2018).

Descriptive statistics were computed to examine students' self-reported perceptions of the STTT tool utility across comprehension and engagement dimensions, as shown in Table 1. Among the comprehension-related items, two areas demonstrated particularly high student-reported benefits. Understanding Vocabulary yielded the highest mean score (M = 4.34, SD = 0.747, skewness = −0.786), indicating strong student agreement that STTT facilitated their understanding of unfamiliar words. The negative skewness value suggests that responses were concentrated toward the upper end of the scale, reflecting consistently positive perceptions. Teacher Clarity exhibited the second-highest mean (M = 4.29, SD = 0.733, skewness = −0.612), with

students reporting enhanced comprehension of instructors' speech when using STTT, a finding particularly relevant in multilingual classroom contexts. In contrast, Identify Grammar Patterns received the lowest mean score among comprehension items (M = 3.92, SD = 0.847, skewness = −0.377). The relatively higher standard deviation and less pronounced negative skewness suggest greater variability in student perceptions, indicating more diverse individual experiences regarding STTT's utility for grammar comprehension. While responses remained above the scale midpoint, this finding suggests that STTT may provide differential support across comprehension subdomains, with vocabulary and lecture comprehension being more consistently perceived as beneficial than grammatical pattern recognition.

Within the engagement domain, two items demonstrated particularly elevated mean scores. Less Hesitation to Participate yielded a mean of 4.28 (SD = 0.790), indicating that students perceived reduced reluctance to engage in classroom activities when using STTT. Meanwhile, Reduced Anxiety in Learning produced a mean of 4.25 (SD = 0.729), reflecting students' reports of decreased anxiety levels in English language learning contexts when STTT tools were available. These findings suggest that students perceived notable affective benefits associated with STTT use, particularly in reducing psychological barriers to classroom participation and alleviating language-learning-related anxiety.

The data demonstrates consistent and reliable positive perceptions among participants. The narrow range of variation across all measured dimensions indicates strong agreement regarding the technology's perceived effectiveness, suggesting that reported benefits are experienced broadly rather than being concentrated within specific learning domains or student subgroups. Most remarkably, vocabulary comprehension emerged as the area of highest reported benefit and the most pronounced negative skewness. This finding holds particular importance given the central role of vocabulary in language comprehension and the documented challenges Thai learners face with English lexical acquisition (Nirattisai & Chiramanee, 2014). Teacher clarity and following along with the lecture also received exceptionally high ratings, suggesting that these tools effectively address two fundamental barriers in English-medium instruction: understanding instructor speech and maintaining comprehension throughout extended lessons, particularly for less proficient learners (Shadiev & Yang, 2020). The consistently higher ratings for vocabulary and teacher clarity can be interpreted through CLT as evidence that STTT may reduce intrinsic load, potentially freeing working memory resources to process meaning and complex schema construction, thereby supporting students' participation. Concurrently, SCT provides a framework for understanding how real-time translation and transcription function as mediational tools that scaffold participation and extend learners' zone of proximal development. These findings address RQ1 regarding the extent

Table 1. Summary of item responses: Comprehension and engagement dimensions

Dimensions	Items	N	Mean	SD	Min	Max	Skewness
Comprehension	Understand teacher clearly	224	4.27	0.733	3	5	-0.467
	Follow along with the lecture	224	4.25	0.744	3	5	-0.431
	Comprehend sentences & expression	224	4.21	0.786	2	5	-0.510
	Understand unfamiliar vocabulary	224	4.34	0.747	2	5	-0.786
	Save time understanding the lesson	224	4.15	0.793	2	5	-0.542
	Identify grammar patterns	224	3.92	0.849	1	5	-0.379
	Feeling less lost in the discussion	224	4.23	0.744	3	5	-0.395
	Focus better in the lesson	224	4.13	0.788	2	5	-0.392
	Easy to recall lesson content	224	4.18	0.754	2	5	-0.435
	Complete assigned task accurately	224	4.2	0.767	1	5	-0.652
Engagement and participation	Engaged in class discussion	224	4.17	0.802	1	5	-0.737
	Confident to speak	224	4.16	0.810	1	5	-0.864
	Encouraged to ask question	224	4.15	0.789	1	5	-0.718
	Focused on class discussion	224	4.18	0.779	1	5	-0.733
	Reduced anxiety in learning	224	4.25	0.729	1	5	-0.714
	Less hesitation to participate	224	4.28	0.790	1	5	-1.040
	Encouraged to contribute & collaborate	224	4.21	0.751	2	5	-0.503
	Increased interest in learning	224	4.23	0.762	1	5	-0.778

to which self-initiated use of speech-to-text translation tools is associated with Thai university students' self-reported real-time comprehension and engagement in English lessons, demonstrating overall positive perceptions across multiple comprehension domains. The comparatively varied responses regarding grammar pattern identification suggest that, while STTT may support lower-order processing tasks, students perceive a continued need for targeted instructional scaffolds to facilitate schema development for morphosyntactic patterns.

Open-ended responses identified six primary themes regarding students' reported motivation for using STTT tools in the classroom, as summarized in Table 2.

Comprehension enhancement emerged as the most frequently cited theme, with students indicating that STTT supported their ability to follow lessons more effectively, particularly when instructors spoke rapidly or with unfamiliar accents. These reported motivations reflect the immediate practical concerns of Thai students in English-medium instruction environments, where perceived comprehension gaps have been associated with academic disengagement and diminished learning outcomes (Tipprachaban, 2022). Vocabulary acquisition was closely associated with comprehension, as students reported utilizing the tool to learn new words, translate unfamiliar terms, and identify lexical patterns. This finding indicates that students

Table 2. Motivation of using STTT tools

Motivation themes	Quotes from responses	Frequency
Enhanced comprehension	"The tool makes it easier to follow the lesson, especially when the teacher speaks quickly." (S6) "It helps me understand the content that the teacher wants to communicate clearly." (S17) "I can go back and review what the teacher said after class." (S42)	95
Understanding vocabulary	"It helps translate unfamiliar words so I can learn their meaning." (S8) "I learned many new words and their patterns in sentences." (S26) "The translation lets me recognize and remember difficult vocabulary." (S95)	58
Boost confidence	"I feel more confident answering questions because I understand better." (S31) "Knowing the meaning of words makes me less nervous to talk in class." (S45) "It gives me courage to ask questions in English." (S82)	34
Pronunciation and listening aid	"It helps me listen carefully and pronounce new words correctly." (S5) "I can check how words are pronounced while following the lesson." (S33) "I hear and understand sounds that I usually miss." (S67)	21
Efficiency and accessibility	"It saves time when translating long sentences." (S9) "It makes learning faster and smoother." (S14)	18
Communication clarity	"It helps me understand the teacher's instructions correctly." (S73)	15

Note. S = Student. Quoted excerpts are anonymized and representative of typical responses within each theme.

perceive these tools not as passive translation devices but as active learning aids that support their vocabulary development objectives. The alignment between this reported motivation and the quantitative finding that vocabulary comprehension received the highest benefit rating suggests a coherent relationship wherein perceived vocabulary support is associated with continued tool adoption.

Students also described increased confidence in classroom communication when using the tools, which they perceived as encouraging greater participation. This reported motivation holds particular relevance in the Thai educational context, where traditional hierarchical classroom structures and cultural factors have been documented as potential barriers to student participation (Pattapong, 2015). Additional motivational factors revealed that students identified multiple pathways through which STTT tools could support their English language development. These findings address RQ2 concerning the motivations underlying university students' use of speech-to-text translation tools during English language classes, revealing a complex interplay of immediate comprehension requirements, vocabulary acquisition objectives, confidence-building aspirations, and broader learning enhancement goals.

A substantial proportion of students (N=75) reported a greater willingness to contribute verbally in class and to initiate interactions with both teachers and peers when using STTT tools (see Table 3). This pattern is consistent with the principles of CLT, suggesting that the availability of real-time textual support was perceived to alleviate extraneous processing demands, thereby allowing learners to allocate more cognitive resources to language production and interaction. Additionally, the relatively high frequency of responses highlighting enhanced classroom participation (N = 45) resonates

Table 3. Responses on students' engagement in class

Impact to students	Description	Common quotes	Frequency
Willingness to speak/interact	Encourages students to initiate verbal participation and interaction	"I dare to answer more." (S1) "I am able to talk more." (S18)	75
Enhanced participation	More active and frequent engagement in classroom activities	"It helps me to participate more." (S8) "I am become more involved." (S14)	45
Increased confidence	Tools boost students' confidence to speak, answer, and present	"I am more confident to speak." (S4) "I am confident in expressing myself." (S12)	38
Improved communication	Helps students better understand and respond to teacher interactions	"Helps me communicate with teachers." (S21) "I understand teachers more." (S34)	30
Motivation and interest	Increases students' interest and enjoyment, leading to greater engagement	"The translation tools make the class more fun and interesting." (S25) "It makes the lesson more engaging." (S68)	28

Note. S = Student. Quoted excerpts are anonymized and representative of typical responses within each theme.

with SCT, which underscores the importance of mediated tools and collaborative interaction in language development. Within this framework, STTT tools appeared to act as cultural mediators, enabling a clearer understanding of instructional input, fostering peer collaboration, and contributing to students' reported confidence and willingness to engage actively in classroom discourse.

This behavioral shift is particularly salient within the Thai educational context, where students have traditionally been described as more reserved in classroom interaction due to face-saving concerns and deference authority figures (Pattapong, 2015). Reports of greater willingness to speak and collaborate, therefore, indicate a meaningful transition from predominantly passive reception to more active participation, a crucial pedagogical outcome given that interaction is widely recognized as central to developing communicative competence in second language acquisition. Improved communication with teachers emerged as another frequently mentioned benefit. Participants perceived that STTT tools helped reduce interactional barriers, enabling clearer exchanges that may otherwise be constrained by hierarchical classroom norms (Bratchuk & Smith, 2023). In addition, several students expressed that STTT tools made learning more enjoyable and less intimidating, which in turn appeared to support their motivation to engage more fully in classroom tasks. This affective dimension suggests that the technology was perceived to address not only cognitive barriers to comprehension but also emotional factors that influence participation and language learning success (AlTwijri & Alghizzi, 2024). These qualitative accounts complement the quantitative finding, indicating that self-reported reduced anxiety was associated with greater willingness to participate, illustrating how students linked improved comprehension to greater confidence and, ultimately, to more active involvement in class discourse.

Collectively, these findings respond to RQ3 by elucidating how STTT tools were perceived to shape engagement and participation through interconnected cognitive, affective, and sociocultural processes.

While participants acknowledged the pedagogical value of STTT tools, their real-time classroom use was reported to present several linguistic, technical, and cognitive barriers. The most frequently cited concern, reported by 44 students, was translation inaccuracy. This limitation is particularly prominent in the English-Thai context, where sentence structures, idiomatic usage, and grammatical patterns hinder accurate real-time rendering. For instance, the English construction "This book is good" translates to Thai as "หนังสือเล่ม นี้ดี" (book-เล่ม-this-good), where the copula "is" is omitted and a classifier (เล่ม) appears between the noun and the demonstrative. Such structural divergences may contribute to students' reported difficulties with certain grammatical patterns when relying on STTT tools. Several participants noted that STTT tools struggled with natural conversational English, particularly with culturally embedded or context-specific expressions, reflecting a lack of localization and domain adaptation in current translation models (Naveen & Trojovský, 2024). Thai students, many of whom have limited exposure to idiomatic English, were further disadvantaged when the tool produced confusing or misleading interpretations of figurative language.

The second most notable barrier, cited by 15 students, related to internet connectivity and wi-fi stability issues. Pronunciation and accent-recognition issues, reported by 12 students, also hindered the tool's effectiveness: participants expressed frustration when the system failed to recognize regional or non-native English accents, a limitation that reflects the broader inclusivity gap in ASR models, which are often trained primarily on native-speaker corpora. Timing and speed constraints, mentioned by 10 students, highlighted the difficulty of keeping pace with fast moving discussions or lecturers, leading to delays that some participants felt interfered with communication flow and undermined confidence in class participation. Additional challenges involved device-related constraints such as outdated smartphones, limited data plans, and unreliable hardware performance. This report reflects the digital divide within the Thai student population, particularly those from rural areas or lower socioeconomic backgrounds who might arguably benefit most from assistive STTT support, but who often experience the least reliable access to technology. It is noteworthy that several participants either reported no significant issues or did not provide a response to this question. These findings correspond to RQ4 regarding the challenges students encounter when using STTT tools during class.

5. Conclusion

This study investigated the association of STTT tools in real-time use and Thai university students' English language learning experiences, focusing on comprehension perceptions, student motivations, engagement patterns,

and implementation challenges. The findings reveal a complex landscape wherein these emerging technologies demonstrate educational potential while simultaneously presenting barriers to effective implementation. This research contributes to the growing body of literature on technology-enhanced language learning, particularly in the context of low-resource languages and developing educational environments. The study provides evidence that students perceive STTT tools as supporting real-time comprehension of English lessons. The consistently positive perceptions across diverse comprehension dimensions establish a foundation for understanding the educational potential of these technologies. The identification of distinct motivational themes underlying STTT tool adoption contributes to the theoretical understanding of students' agency in technology-mediated language learning. The documentation of the relationship between improved comprehension perceptions and enhanced self-reported classroom participation represents a contribution to understanding how assistive technologies may influence traditional dynamics in hierarchical educational settings. The observed progression from increased confidence to willingness to interact provides a pedagogical framework for understanding how real-time language support tools may facilitate more inclusive and participatory learning environments. Most importantly, this study contributes to the limited research on STTT implementation challenges in non-Western educational contexts. The comprehensive catalogue of technical, linguistic, and cognitive barriers provides insights for researchers and developers working on language learning technologies for diverse global populations. The finding that translation inaccuracy was the most frequently cited obstacle, particularly for English-Thai language pairs, highlighting the ongoing limitations of current machine translation models when applied to low-resource languages with complex structural differences.

Several limitations should be acknowledged in interpreting the findings of this study. The research was conducted within a specific educational and cultural context, Thai university students learning English, which may constrain the generalizability of findings to other linguistic and cultural settings. Additionally, the reliance on self-reported data for both quantitative scales and qualitative themes introduces potential bias related to social desirability, memory limitations, and individual differences in self-awareness and articulation ability. While student perceptions provide valuable insights into user experience, they may not fully correspond to the actual learning gains or skill development. Consequently, the sampling approach may overrepresent the perspectives of technologically proficient or highly engaged learners, potentially skewing favorable perceptions of STTT tools. Furthermore, the rapid pace of advancement in machine translation and speech recognition technologies indicates that some identified challenges may be mitigated by newer systems, while novel challenges may emerge. Future research could employ longitudinal designs to examine sustained patterns of STTT tool use and associated perceptions of learners' comprehension

and classroom engagement over time. Such approaches would offer deeper insights into developmental trajectories and long-term pedagogical value. Additional studies incorporating objective performance measures alongside self-reported perceptions would provide a more comprehensive understanding of STTT tools' educational impact.

Despite these limitations, this study provides valuable insights into the potential and challenges of STTT tools in language education contexts, contributing important evidence to inform both theoretical understanding and practical applications of these emerging technologies. The findings establish a foundation for future research that can address these limitations while building upon the insights generated by this investigation.

Conflict of interest

The authors declare that they have no financial or non-financial conflicts of interest in relation to this work.

Ethics Declaration

Ethical approval for this study was obtained from the Walailak University Ethics Committee (Approval No. WUEC-25-187-01; May 20, 2025). All procedures, including participant recruitment, data collection, and data handling were conducted in accordance with the committee's ethical guidelines. Informed consent was built into the online questionnaire: before proceeding to any survey items, participants reviewed a consent statement embedded in the Google Form and indicated their voluntary agreement to participate. To protect confidentiality, no personally identifiable information was collected, and all responses were stored on a secure university server accessible only to the research team.

AI Disclosure

Participants utilized speech-to-text translation tools across operating systems such as Android, iOS, and Microsoft Windows, using versions of Google Translate, ChatGPT, Microsoft Translator, Azure, Notta, and Soniox. These services were accessed primarily via personal smartphones and tablets, with a minority of participants using laptops. While specific app versions were not recorded, data collection occurred between May to July 2025. All audio processing and text translation took place within the respective applications or via encrypted cloud connections, in accordance with each vendor's privacy policy. The research team did not collect or retain any raw audio files, transcription logs, translated texts, or personally identifiable information. Input audio was in English, with Thai as the target output language. For transparency, reported English-to-Thai translation challenges include idiomatic expressions, classifier mismatches, syntactic differences such as word-order shifts, word-for-word translation mismatches, and semantic distortion or unnatural phrasing.

Data and Code Availability

Anonymized survey instruments (English and Thai versions) and analysis scripts have been deposited in a public repository. The files are openly available at https://doi.org/10.5281/zenodo.17289425 without the need for special permission. Researchers may access, cite and adapt these materials in accordance with the repository's licensing terms.

Author contributions

Conceptualization, N.S, J.B. and R.L..; Methodology, J.B...; Literature review, N.S, R.L. and J.B; Ethical approval submission, J.B. and N.S.; Data planning and participants selection, N.S..; Task and consent form design, J.B..; Recruitment of Participants, N.S. and J.B; Communication with external evaluator, J.B.; Data collection, N.S and R.L.. Data entry and curation, J.B and R.L; Analysis, and visualization, J.B.; Proofreading, N.S and R.L..; Writing – original draft, J.B. and N.S..; Writing – review and editing,N.S,J.B. and R.L; Project administration, N.S.. All authors contributed equally to the development and finalization of the manuscript.

REFERENCES

Alkhudiry, R. (2022). The contribution of Vygotsky's sociocultural theory in mediating L2 knowledge co-construction. *Theory and Practice in Language Studies*, *12*(10), 2117. https://doi.org/10.17507/tpls.1210.19

AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon*, *10*(10). https://doi.org/10.1016/j.heliyon.2024.e31053

An, Z., Wang, C., Li, S., Gan, Z., & Li, H. (2021). Technology-assisted self-regulated English language learning: Associations with English language self-efficacy, English enjoyment, and learning outcomes. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.558466

Ayyaz, S. (2025). AI translation tools in language learning: Bridging gaps and enhancing skills. In M. M. Ali, M. N. Anwar, S. Fazal, & S. Ayyaz (Eds.), *AI Applications for English Language Learning* (pp. 43–72). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-9077-1.ch003

Babatunde, A. N., Babatunde, R. S., Balogun, B. F., Umar, E., Mohammed, S. B., Oke, A. A., & Obiwusi, K. Y. (2024). Speech-to-text hybrid English to Yoruba SMS translator. *Innovative Computing Review*, *4*(1), 15–36.

Bahari, A. (2022). Challenges and affordances of cognitive load management in technology-assisted language learning: A systematic review. *International Journal of Human–Computer Interaction*, *39*(1), 85–100. https://doi.org/10.1080/10447318.2021.2019957

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv Preprint arXiv:1409.0473*.

Barrault, L., Chung, Y.-A., Cora Meglioli, M., Dale, D., Dong, N., Duquenne, P.-A., ... Elsahar, H. (2023). *SeamlessM4T: Massively multilingual & multimodal machine translation*. https://doi.org/10.48550/arXiv.2308.11596

Bensalah, H. (2024). Exploring the potential synergy of Quillbot as a natural language processing tool in demystifying academic writing. *أطراس*, *5*(3), 80–96. https://doi.org/10.70091/atras/ai.5

Benson, P. (2013). *Teaching and Researching Autonomy in Language Learning* (2nd ed.). Routledge.

Bérard, A., Besacier, L., Kocabiyikoglu, A. C., & Pietquin, O. (2018). End-to-end automatic speech translation of audiobooks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6224–6228). IEEE. https://doi.org/10.1109/ICASSP.2018.8461690

Bratchuk, H., & Smith, P. (2023). Overcoming of communication barriers in the classroom. *EIKI Journal of Effective Teaching Methods*, *1*(1). https://doi.org/10.59652/jetm.v1i1.6

Chancharoen, K., Tannin, N., & Sirinaovakul, B. (1999). Pattern-based machine translation for English-Thai. In J.-F. Wang & C.-H. Wu (Eds.), *Pacific Asia Conference on Language, Information, and Computation* (pp. 329–336). National Cheng Kung University. https://www.aclweb.org/anthology/Y99-1036

Chang, L.-C. (2022). Chinese language learners evaluating machine translation accuracy. *The JALT CALL Journal*, *18*(1), 110–136. https://doi.org/10.29140/jaltcall.v18n1.592

Choemue, S., & Muljani, R. (2021). Thai university students' attitudes towards ELF in classroom contexts: A mixed-methods study. *ELT Echo The Journal of English Language Teaching in Foreign Language Context*, *6*(1), 13. https://doi.org/10.24235/eltecho.v6i1.7317

Chung, E. S., & Ahn, S. (2021). The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, *35*(9), 2239. https://doi.org/10.1080/09588221.2020.1871029

Deng, L., & Yu, D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197–387. https://doi.org/10.1561/2000000039

Dennis, N. K. (2024). Using AI-powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education*, *12*(2), 107. https://doi.org/10.22492/ije.12.2.05

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. *Foreign Language Annals*, *51*(4), 779–795. https://doi.org/10.1111/flan.12366

Dwijayani, I., & Musigrungsi, S. (2022). Investigating speaking tasks in relation to communicative goals: Possibilities and obstacles. *Studies in English Language and Education*, *9*(2), 501–520. https://doi.org/10.24815/siele.v9i2.23566

Furui, S. (2010). History and development of speech recognition. In F. Chen & K. Jokinen (Eds.), *Speech Technology*. Springer. https://doi.org/10.1007/978-0-387-73819-2_1

Gao, Y., Wang, R., & Hou, F. (2023). *How to design translation prompts for ChatGPT: An empirical study*. arXiv. https://doi.org/10.48550/arxiv.2304.02182

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, *27*(2), 6–27. https://doi.org/10.64152/10125/73501

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2012). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Gritsay, I., & Vodyanitskaya, L. (2021). Pedagogical technologies of machine translation skills forming on the example of bachelor students specializing in mechatronics and robotics at Don State Technical University. *E3S Web of Conferences*, *273*, 12140. https://doi.org/10.1051/e3sconf/202127312140

Hara, K., & Iqbal, S. (2015). Effect of machine translation in interlingual conversation: Lessons from a formative study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 3473–3482). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702407

Hirai, A., & Kovalyova, A. (2024). Speech-to-text applications' accuracy in English language learners' speech transcription. *Language Learning & Technology*, *28*(1), 1–22. https://doi.org/10.64152/10125/73555

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, *18*(6), 667–683. https://doi.org/10.1002/(SICI)1099-1379(199711)18:6

Hutchins, W. J. (1995). Machine translation: A brief history. In E. F. K. Kroner & R. E. Asher (Eds.), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists* (pp. 431–445). Pergamon.

Hwang, G., & Fu, Q.-K. (2018). Trends in the research design and application of mobile language learning: A review of 2007–2016 publications in selected SSCI journals. *Interactive Learning Environments*, *27*(4), 567. https://doi.org/10.1080/10494820.2018.1486861

Iberahim, A., Yunus, M. M., & Sulaiman, N. A. (2023). A review on technology enhanced language learning (TELL). *International Journal of Academic Research in Business and Social Sciences*, *13*(2). https://doi.org/10.6007/ijarbss/v13-i2/16496

Indriani, S. (2020). Utilizing a recorded video to assess speaking and presentation skills of pre-service teachers. *Scholaria Jurnal Pendidikan Dan Kebudayaan*, *10*(1), 60. https://doi.org/10.24246/j.js.2020.v10.i1.p60-65

Ivenz, P., & Poláková, P. (2024). Gen Z students and their perceptions of technology in the process of second language acquisition based on the language proficiency level. *Arab World English Journal*, *15*(3), 3–17. https://doi.org/10.24093/awej/vol15no3.1

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Jolley, J. R., & Maimone, L. (2022). Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal: An Open Access Refereed Journal for World Language Educators*, *14*(1). https://doi.org/10.5070/L214151760

Kawinkoonlasate, P. (2020). Online language learning for Thai EFL learners: An analysis of effective alternative learning methods in response to the Covid-19 outbreak. *English Language Teaching*, *13*(12), 15. https://doi.org/10.5539/elt.v13n12p15

Kenny, D. (2018). Machine translation. In J. P. Rawling & P. Wilson (Eds.), *The Routledge Handbook of Translation and Philosophy* (pp. 428–445). Routledge. https://doi.org/10.4324/9781315678481-27

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.

Kyriazos, T. A. (2018). Applied psychometrics: Writing-up a factor analysis construct validation study with examples. *Psychology*, *9*(11), 2503–2530. https://doi.org/10.4236/psych.2018.911144

Lampou, R. (2023). The integration of artificial intelligence in education: Opportunities and challenges. *Review of Artificial Intelligence in Education*, *4*(00), e15. https://doi.org/10.37497/rev.artif.intell.educ.v4i00.15

Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2014). Sociocultural theory and second language development. In B. van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 207–226). Routledge.

Lavanya, R., Gautam, A. K., & Anand, A. (2024). Real time translator with added features for cross language communication. In *Proceedings of the 2024 10th International Conference on Communication and Signal Processing (ICCSP)* (pp. 854–857). IEEE. https://doi.org/10.1109/ICCSP60870.2024.10543567

Lim, F. V., & Toh, W. (2024). Apps for English language learning: A systematic review. *Teaching English with Technology*, *24*(1), 79–98. https://doi.org/10.56297/FSYB3031/GAQR3589

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., & Zettlemoyer, L. (2020). Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, *8*, 726–742. https://doi.org/10.1162/tacl_a_00343

Lyons, S. (2016). Quality of Thai to English machine translation. In H. Ohwada & K. Yoshida (Eds.), *Knowledge Management and Acquisition for Intelligent Systems. PKAW 2016. Phuket, Thailand, 22-23 August 2016*. Springer. https://doi.org/10.1007/978-3-319-42706-5_20

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Melkonyan, A., & Matevosyan, A. (2020). Technology-assisted foreign language learning (TALL) in the digital age. *SHS Web of Conferences*, *88*, 2005. https://doi.org/10.1051/shsconf/20208802005

Mohammed, S. J., & Khalid, M. (2025). Under the world of AI-generated feedback on writing: mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, *15*(1). https://doi.org/10.1186/s40468-025-00343-2

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *iScience*, *27*(10), 110878. https://doi.org/10.1016/j.isci.2024.110878

Nirattisai, S., & Chiramanee, T. (2014). Vocabulary learning strategies of Thai university students and its relationship to vocabulary size. *International Journal of English Language Education*, *2*(1), 273. https://doi.org/10.5296/ijele.v2i1.5366

Novikov, P., & Kiseleva, A. (2024). Over-reliance on technology in foreign language learning: case study of LLB undergraduates. *INTED2024 Proceedings, 18th International Technology, Education and Development Conference*, 2962–2966. https://doi.org/10.21125/inted.2024.0803

Oleksiienko, L., Oksana, B. V., Trubitsyna, O., Olena, M. I., & Tetiana, P. D. (2020). Interactive methods of teaching foreign languages in higher education institutions. *Revista Tempos e Espaços Em Educação*, *13*(32), 1. https://doi.org/10.20952/revtee.v13i32.14960

Pabba, M. P., Sai, C. Y., Manasa, Y. S., Nityadeep, V., & Chakridhar, P. (2024). A comprehensive study on live multimodal language translation System. *International Journal of Engineering Research and Science & Technology*, *20*(3), 10–15.

Pattapong, K. (2015). Complex interactions of factors underlying Thai EFL learners' willingness to communicate in English. *PASAA*, *49*(1), 105–136. https://doi.org/10.58837/CHULA.PASAA.49.1.4

Pipatanakul, K., Jirabovonvisut, P., Manakul, P., Sripaisarnmongkol, S., Patomwong, R., Chokchainant, P., & Tharnpipitchai, K. (2023). Typhoon: Thai large language models. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2312.13951

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning* (pp. 28492–28518). PMLR. https://doi.org/10.48550/arXiv.2212.04356

Ravanelli, M., Brakel, P., Omologo, M., & Bengio, Y. (2017). Improving speech recognition by revising gated recurrent units. *Interspeech*, 1308–1312. https://doi.org/10.21437/interspeech.2017-775

Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023). Speech-to-text and text-to-speech recognition using deep learning. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 657–666. https://doi.org/10.1109/ICECAA58104.2023.10212222

Reitmaier, T., Wallington, E., Kalarikalayil Raju, D., Klejch, O., Pearson, J., Jones, M., ... Robinson, S. (2022). Opportunities and challenges of automatic speech recognition systems for low-resource language speakers. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery, Inc. https://doi.org/10.1145/3491102.3517639

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Lang Resources & Evaluation*, *56*, 593–619. https://doi.org/10.1007/s10579-021-09537-5

Rodphotong, S. (2018). The effectiveness of collaborative learning to enhance English communicative competence: A case study of the first-year students at Thepsatri Rajabhat University. *IJPTE International Journal of Pedagogy and Teacher Education*, *2*, 15. https://doi.org/10.20961/ijpte.v2i0.25174

Samana, W. (2013). Teacher's and students' scaffolding in an EFL classroom. *Academic Journal of Interdisciplinary Studies*, *2*(8). https://doi.org/10.5901/ajis.2013.v2n8p338

Schenck, A. (2024). Examining relationships between technology and critical thinking: A study of South Korean EFL learners. *Education Sciences*, *14*(6), 652. https://doi.org/10.3390/educsci14060652

Seligman, M., & Waibel, A. (2019). Advances in Speech-to-Speech Translation Technologies. In M. Ji & M. Oakes (Eds.), *Advances in Empirical Translation Studies* (pp. 217–251). Cambridge University Press. https://doi.org/10.1017/9781108525695.012

Shadiev, R., Chien, Y.-C., & Huang, Y.-M. (2020). Enhancing comprehension of lecture content in a foreign language as the medium of instruction: Comparing speech-to-text recognition with speech-enabled language translation. *SAGE Open*, *10*(3). https://doi.org/10.1177/2158244020953177

Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), 524. https://doi.org/10.3390/su12020524

Sim, J., & Wright, C. (2005). The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

Singh, K. K. A. P. H., & Yunus, M. M. (2021). Technology-enhanced language learning (TELL) tools: A survey on attitude of a private secondary school ESL teachers. *International Journal of Academic Research in Progressive Education and Development*, *10*(1). https://doi.org/10.6007/ijarped/v10-i1/8431

Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, *69*, 343–418. https://doi.org/10.1613/jair.1.12007

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27).

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer Science & Business Media.

Syathroh, I. L., Kareviati, E., Lestari, A. P., & Fitria, N. (2021). Exploring the potentials of technology integration for teaching language skills: A literature review. *PROJECT (Professional Journal of English Education)*, *4*(3), 488. https://doi.org/10.22460/project.v4i3.p488-496

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. *The SAGE Handbook of Qualitative Research in Psychology*, *2*(17–37), 25. https://doi.org/10.4135/9781526405555.n2

Tipprachaban, B. (2022). Challenges of using English as a medium of instruction (EMI) in Thailand. *PUPIL: International Journal of Teaching, Education and Learning*, *6*(2), 01–12. https://doi.org/10.20319/pijtel.2022.62.0112

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30).

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. *Engineering*, *18*, 143–153. https://doi.org/10.1016/j.eng.2021.03.023

Wei, Y. (2022). Toward technology-based education and English as a foreign language motivation: A review of literature. *Frontiers in Psychology*, *13*, 870540. https://doi.org/10.3389/fpsyg.2022.870540

Weyers, M., Strydom, H., & Huisamen, A. (2014). Triangulation in social work research: The theory and examples of its practical application. *Social Work/Maatskaplike Werk*, *44*(2). https://doi.org/10.15270/44-2-251

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2017). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities*, *51*(1), 73–84. https://doi.org/10.1177/0022219416688170

Yang, J. (2025). A systematic review of online AI translation tools for English language learning and teaching. *International Journal of Internet, Broadcasting and Communication*, *17*(1), 129–139. https://doi.org/10.7236/IJIBC.2025.17.1.129

Yang, Z., Liu, F., Yu, Z., Keung, J., Li, J., Liu, S., Hong, Y., Ma, X., Jin, Z., & Li, G. (2024). Exploring and unleashing the power of large language models in automated code translation. *Proceedings of the ACM on Software Engineering*, *1*, 1585. https://doi.org/10.1145/3660778

Yashina, E., Rubanik, T., & Chukhray, A. (2022). Speech-To-Text software design for the high education learning. *2nd International Workshop of IT-Professionals on Artificial Intelligence (ProfIT AI 2022)*, 44–56.

Zhang, R., & Zou, D. (2020). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, *35*(4), 696. https://doi.org/10.1080/09588221.2020.1744666

Zhu, X. (2020). Machine translation in foreign language learning classroom-learners' indiscriminate use or instructors' discriminate stance. *English Linguistics Research*, *9*(4), 1. https://doi.org/10.5430/elr.v9n4p1