ORIGINAL RESEARCH ARTICLE

Real-time speech-to-text translation in Spanish secondary classrooms: a mixed-methods study on refugee student inclusion

Ricardo Scott[a]*, Clara Vila[b], Daniel Pérez-Alcaraz[b], Olga Vaello[b], Jose Manuel Pérez-Torres[b], Ricardo Ibanco-Cañete[a], Jorge Brotons-Mas[c], Cristina de-la-Peña[d], María José Álvarez-Alonso[e] and Teresa Pozo-Rico[a]

[a]*Department of Developmental Psychology and Didactics, University of Alicante, Alicante, Spain;* [b]*Secondary Education Teachers of Public Schools of the Valencian Community, Alicante, Spain;* [c]*Institute of Neurosciences UMH-CSIC, Alicante, Spain / Cardenal Herrera Oria University, Elche, Spain;* [d]*International University of La Rioja, Madrid, Spain;* [e]*Alfonso X El Sabio University, Madrid, Spain*

Following the 2022 invasion of Ukraine, thousands of Ukrainian children enrolled in schools across Europe. In Spain, most lacked prior knowledge of Spanish. This study examines whether real-time speech-to-text translation technology (STTT) can reduce classroom language barriers. Two activities – a fable reading and a neuroscience lecture – were conducted with 12–15-year-old Spanish-speaking students ($n = 23$) and Ukrainian students unfamiliar with Spanish but bilingual in Ukrainian and Russian. Using PowerPoint 365, the teacher's speech was transcribed and translated into Russian – which at the time was far more reliably supported by automatic translation tools than Ukrainian – and projected onto a shared classroom display. Although this choice was based on technical and pedagogical criteria, it later drew some resistance, reflecting the sociopolitical sensitivities surrounding language use in wartime contexts. Comprehension was assessed using content-specific questionnaires. Ukrainian students scored lower than their Spanish peers but significantly higher than a control group ($n = 22$; $p < 0.001$; Cliff's delta indicated large effect sizes). Qualitative analysis of teacher interviews highlighted improvements in comprehension and inclusion, along with implementation challenges. Taken together, these findings indicate that STTT has the potential to support newly arrived refugee students and help address multilingual education challenges.

Keywords: AI transcription processing; refugee education; language barriers; educational inclusion

Introduction

A few months after the Russian invasion of Ukraine, over 6 million Ukrainians, including 665 000 students, fled their country and were welcomed by neighbouring nations (World Bank, 2022). Between 2014 and 2023, more than 3300 Ukrainian

---

*Corresponding author. Email: ricardo.scott@ua.es

1

educational institutions were attacked, with 1888 occurring after the full-scale invasion (Save the Children, 2023).

In Spain, UNESCO reported that by March 2022, over 7000 Ukrainian students had joined educational institutions, 25% in secondary education. By February 2024, this number exceeded 38 000, with the Valencian Community hosting the largest share (12 730; INE, 2024). Initial rapid growth stabilised from March 2023 onwards.

Ukraine's linguistic history reflects its complex past: Ukrainian predominates in rural areas, while Russian has been more common in urban centres, particularly during the Soviet era (Besters-Dilger, 2023). Although Ukrainian is the state language, much of the population remains bilingual (Ivanova, 2013).

Most Ukrainian students arriving in Spain had no prior knowledge of Spanish, as English is the primary foreign language taught in Ukraine (84%; Smotrova, 2009). Their integration has posed challenges, with language barriers being among the most critical. Addressing these barriers, especially in education, is an urgent priority in such conflict scenarios.

Speech Recognition Technology (SRT) has been widely studied for its potential in supporting learning, particularly language acquisition (Dai & Wu, 2023; Jeon et al., 2023; Nickolai et al., 2024; Shadiev & Liu, 2023; Sun, 2023). Early studies highlighted its role in providing immediate feedback and improving pronunciation skills (Bernstein et al., 1990; Ehsani & Knodt, 1998; Warschauer & Healey, 1998). Recent advancements, including neural network-based systems, have enhanced SRT's accuracy and applicability in educational contexts (Gao et al., 2022; Shadiev et al., 2014). However, concerns remain regarding the accuracy and consistency of these tools, especially when used by non-native speakers or in classroom environments with background noise (Huang et al., 2016; Scott et al., 2022; Shadiev et al., 2020).

Despite these limitations, tools such as Google Speech Recognition and Dragon Naturally Speaking assist with listening, writing, and vocabulary acquisition (McKechnie et al., 2018; Shadiev & Liu, 2023). Research also shows SRT facilitates text comprehension by synchronising speech input with text output, benefiting learners with limited language proficiency (Chen, 2022; Shadiev et al., 2017a). These benefits extend beyond language learning. For example, Alvarez-Alonso et al., (2021) reported significant improvements in text comprehension when combining auditory and visual presentations, especially among boys and students with lower academic performance. This aligns with Gernsbacher's (2015) findings that captions enhance comprehension, memory, and attention across diverse populations, including non-native speakers and those with varying literacy levels.

An important application of SRT is Speech-to-Text Technology (STT), which provides accurate transcriptions of verbal content in natural contexts, improving access to spoken information (Furui et al., 2004; Reddy et al., 2023). Scott et al. (2022) highlighted STT's effectiveness in classroom settings, particularly for addressing auditory discrimination challenges, enhancing comprehension in noisy environments, and supporting diverse learning needs.

STT also addresses specific learning challenges, improving accessibility, engagement, and independence for students with learning difficulties (Berner & Alves, 2023; Matre & Cameron, 2022). Additionally, STT interventions have significantly enhanced text production in students with intellectual disabilities, increasing productivity and text quality (Sand et al., 2024). However, it is important to acknowledge that many such technologies were not originally designed with disability in mind, and claims of accessibility may reflect technoableist assumptions rather than actual inclusive design

(Shew, 2020). Despite these concerns, the reported improvements in learner outcomes suggest that STT – when thoughtfully implemented – holds promise for supporting a wider range of educational needs in diverse classroom contexts.

On the other hand, Speech-to-Text *Translation* (STTT) technology enables the conversion of spoken language into text −STT− *and* its translation into another language, facilitating communication in multilingual contexts. Until recent years machine translation technologies have shown a growing but still limited application in critical fields such as healthcare, where they have improved multilingual communication and access to information (Dew at al., 2018). Recent advancements in translation models have achieved greater speed and accuracy by integrating advanced technologies such as neural networks and deep learning, significantly reducing processing errors, which ensure a promising future in many fields (Sethiya & Maurya, 2025).

The STTT has shown promising potential in multilingual education by reducing language barriers and enhancing comprehension. Tools such as Google Translate and Microsoft Translator support language learning, translation training, and multilingual lectures, helping learners better understand foreign-language content (Shadiev et al., 2024). In university lectures, STTT provides real-time translation and displays translated content, particularly benefiting non-native speakers with limited proficiency (Shadiev et al., 2017b; Shadiev & Sun, 2019). Research also highlights that STTT improves accessibility and comprehension, while its impact on cognitive load depends on students' language proficiency and the complexity of the content (Shadiev et al., 2020).

Despite these advancements, the application of STTT in natural classroom environments, particularly at the secondary school level, remains underexplored. Our study addresses this gap by exploring the feasibility and potential classroom utility of STTT to reduce language barriers and improve the inclusion of Ukrainian refugee students in Spain. By observing its impact on academic content comprehension and gathering teacher feedback, we aim to provide evidence of its effectiveness. These findings contribute to ongoing discussions about the role of technology in enhancing education in multicultural and multilingual contexts, fostering greater inclusivity and addressing diverse learning needs.

Methods

Participants

The study included 62 volunteer secondary school students, aged 12 to 15, 56% girls, from five secondary education and vocational training schools in the province of Alicante (Spain) selected through convenience sampling. Of all participants, 20 were Ukrainian students who had recently arrived in Spain following the Russian invasion of Ukraine, 23 were native Spanish-speaking students, and 22 formed a mixed control group that did not attend the sessions but completed the tests. A group of Ukrainian students assisting to the activities without STTT support was not considered due to ethical reasons (see Ethical considerations section). The control group included both Spanish- and Ukrainian-speaking students of similar age and grade level who did not attend the activities and completed the same questionnaires in their native language. This allowed us to estimate baseline performance without prior exposure to the content, independently of language, as both versions of the questionnaires were designed and scored equivalently. In other words, since none of the control group participants – whether Spanish or Ukrainian – had any knowledge of the class content, the group

served as a valid baseline, assuming comparable background knowledge across cultures. Notably, Spanish speakers in the control group tended to score slightly higher, introducing a conservative bias that reinforces the validity of the observed improvements in the Ukrainian STTT groups. As mentioned in the Ethical Considerations section, as the war progressed, it became increasingly difficult to recruit new Ukrainian participants, which limited the composition and size of the control group.

The Ukrainian students were selected by their teachers, who confirmed that they had no prior knowledge of Spanish, ensuring the presence of a language barrier for all participants in this group.

Participants were distributed across four conditions based on their language (Spanish or Ukrainian) and the educational activity they attended (a neuroscience lecture or the reading of a fable), with an additional control group for each activity. These groups were: Spanish-speaking students attending the neuroscience lecture (*Spanish Brain*), Spanish-speaking students attending the reading of the fable (*Spanish Fable*), Ukrainian students using the STTT system during the neuroscience lecture (*Ukrainian Brain*), Ukrainian students using the STTT system during the fable-reading activity (*Ukrainian Fable*), and control groups of students who answered the questionnaires without attending the activities (*Control Brain* and *Control Fable*).

Procedure

The study assessed STTT's effectiveness through two activities:

1. *Neuroscience Lecture.* A 15-min neuroscience lecture was delivered by a teacher using PowerPoint 365's STTT, transcribing and translating spoken Spanish into Russian in real time and projecting it onto a screen. Visual content (images) was included, but no additional text, and the presentation pace was slowed to approximately 90% of the usual speed to accommodate translation delays. Students completed a 10-question multiple-choice questionnaire in Spanish and Russian afterwards.
2. *Fable Reading.* The teacher read a fable aloud while the STTT system transcribed and translated it in real time onto a screen. This activity relied solely on auditory and textual input, excluding visual aids. Students then completed a similar comprehension questionnaire with 10 multiple-choice questions in Spanish and Russian.

The control group, which did not attend the sessions and lacked prior exposure to the content, answered the same questionnaires to establish baseline performance.

Questionnaire validation

Two bilingual experts in neuroscience and pedagogy validated the comprehension questionnaires to ensure their suitability for the participants' educational level. Translations from Spanish to Russian were reviewed for both functional and conceptual equivalence by a bilingual Russian–Spanish student, with additional support from DeepL.

Internal consistency was assessed using Cronbach's alpha to ensure the reliability of the questionnaires. Questions with high accuracy rates in the

control group (> 80%) were excluded to improve reliability. The final versions of the questionnaires consisted of six items for the Fable activity and seven items for the Brain activity. Analyses conducted using the *psych* package in R yielded standardised alpha scores of 0.72 for the Fable questionnaire and 0.75 for the Brain questionnaire, reflecting acceptable reliability across the evaluated groups.

Evaluation of STTT system

A functional evaluation of the STTT system was conducted prior to the study using a back-translation procedure involving a fully bilingual student and an independent evaluator. The process, repeated across texts of varying complexity and speech speeds, yielded an estimated average translation reliability of approximately 70%. While not a formal validation, this approach offered a practical and context-sensitive estimate of translation adequacy.

Quantitative analysis

Statistical analysis

Normality was tested for each group using the Shapiro-Wilk test, which revealed violations of normality in several groups ($p < 0.05$). Homogeneity of variances was assessed using Levene's test (centred on the median), confirming homogeneity across groups ($F(5, N) = 0.34$, $p = 0.887$). Due to the non-normal distribution of scores, non-parametric methods were employed. The Kruskal-Wallis test was used to assess overall differences between groups. The 'Group' variable combined participants' experimental conditions (Spanish, Ukrainian, or Control) and activity conditions (Brain or Fable) into six distinct groups ($\chi^2(5) = 68.47$, $p < 0.001$). Post hoc pairwise comparisons were conducted using the Mann-Whitney U test with Bonferroni-adjusted *p*-values to control for Type I error. Only comparisons with adjusted *p*-values below 0.05 were interpreted.

Cliff's Delta was calculated for all pairwise comparisons to measure the magnitude of the differences between groups. Effect sizes were interpreted as follows: small ($|d| \leq 0.147$), medium ($0.147 < |d| \leq 0.33$), and large ($|d| > 0.33$).

The analyses were conducted using R packages, including *dplyr* for data manipulation, *ggplot2* for visualisation, *car* for Levene's test, *rstatix* for Mann-Whitney U tests, *effsize* for effect size calculations, and *flextable* for formatting and exporting results.

Qualitative design and analysis

This study employed an exploratory qualitative design to examine the experiences, perceptions, and recommendations of secondary-level teachers who used STTT in their classrooms. Audio responses of participants were transcribed using the Online Automatic Transcription software (TAL), developed by the University of Alicante's IT service, which utilises Google's voice recognition technology. The transcriptions were refined to correct punctuation errors, hesitations, and unnecessary pauses using GPT-4o, and then cross-checked with the original recordings to ensure accuracy and reliability.

Interviewed participants

Five secondary-level teachers who regularly used the STTT system in their classrooms were selected for the interviews. Their profiles were diverse in terms of subject area and role within the school. The group included: (1) a technology teacher; (2) a language teacher who, in addition to teaching, supported foreign students (not only Ukrainian); (3) a computer science teacher; (4) a vocational training teacher in graphic arts; and (5) a school counsellor. Convenience sampling was employed, prioritising teachers with substantial experience using the system. The interviews focused on five key questions addressing the following topics:

1. *How would you describe your experience using the real-time transcription and automatic translation system (STTT) in the classroom?*
2. *What impact have you observed on the learning and participation of Ukrainian students when using the system?*
3. *What have been the main technical or pedagogical difficulties you faced when implementing the system?*
4. *From your perspective, what aspects of the STTT system should be improved to make it more effective in the classroom*
5. *What recommendations would you give to other teachers interested in implementing this technology in similar contexts?*

Qualitative analysis

The qualitative analysis was conducted collaboratively by two authors, with support from ChatGPT-4 as a triangulation tool (Morgan, 2023), to help identify and refine meaningful themes in the interview transcripts. Human and AI analysis were mainly coincident, except for the identification of notable individual comments. Initial codes were generated based on the five main interview questions and expanded inductively to capture emergent insights. The final coding structure was organised into five thematic categories: (1) Teachers' experiences with the STTT system. (2) Its impact on students' learning and participation. (3) Technical and pedagogical challenges. (4) Areas for improvement. (5) Recommendations for other teachers.

A thematic diagram was created using Python to visually represent the findings. This diagram mapped the relationships between themes and subthemes, providing a clear overview of the qualitative results.

Use of artificial intelligence

To support the writing process, the authors used ChatGPT (OpenAI, 2024) in a dynamic and interactive manner, primarily to improve grammar and clarity, as English is not their first language. The entire manuscript was manually reviewed and revised several times by the authors to ensure accuracy, coherence, and academic rigour. All conceptual development, experimental design, data analysis, and interpretation were carried out using traditional scholarly methods. ChatGPT was also used as a support tool during the qualitative analysis, complementing the manual coding and interpretation process described in the corresponding section.

Ethical considerations

This study was reviewed and approved by the Ethics Committee of the University of Alicante, ensuring compliance with the ethical and legal principles established for research involving human participants. The project was evaluated under file number UA-2024-07-22_1, and its design adhered to the Committee's guidelines, including the protection of participants' rights and the confidential handling of their data.

All participating teachers and students were previously informed about the study's objectives and the conditions of their participation. The voluntary nature of their collaboration was ensured, and data anonymisation was guaranteed, both for interviews and comprehension tests. Furthermore, teachers signed an informed consent form detailing the use of the technological tools involved and the research purpose of the data collection.

We did not include a group of Ukrainian students with no knowledge of Spanish and no access to STTT support in the activities, as it was considered unethical to involve them in a research setting where they could not understand the content. Such a condition would have placed them in a situation of exclusion and academic vulnerability, conflicting with both ethical research standards and the inclusive objectives of the intervention. The students were newly arrived refugees, unfamiliar with the local language and educational system, and were exposing them to learning activities without any linguistic support could have caused distress or a sense of failure. For this reason, we deliberately chose not to establish an experimental group without access to STTT, accepting the methodological trade-off in favour of student well-being.

In addition, while Russian was selected due to better system performance and participants' familiarity, it is important to note the growing demand for Ukrainian-language support, especially given the sociopolitical context and community preferences. This language choice, although practical at the time, may have also limited participation as more families and organisations began to express a preference for Ukrainian-only materials, complicating further recruitment efforts during the study.

Results

This study aimed to evaluate the effectiveness of a real-time transcription and translation system in the classroom. The goal was to improve the comprehension of educational content by displaced Ukrainian students in Spain who lacked knowledge of Spanish.

Quantitative analysis of student performance

Descriptive statistics

Table 1 and Figure 1 summarise the descriptive statistics for the percentage of correct responses across six groups, categorised by native language and activity. The Control Brain group exhibited the lowest mean score ($M = 44.09$, standard deviation [SD] = 16.81), while the Spanish Brain group displayed the highest performance ($M = 88.26$, SD = 16.69). Similarly, among the fable-reading groups, Spanish Fable achieved a higher mean score ($M = 87.83$, SD = 16.44) compared to Ukrainian Fable ($M = 67.46$, SD = 15.39) and Control Fable ($M = 50.88$, SD = 15.41). These results highlight a consistent advantage for Spanish-speaking participants across both activities. However,

Table 1. Descriptive statistics for correct responses by group and activity.

Group	Activity	Mean	SD	Sample_Size
Control	Brain	44.09	16.81	22
Spanish	Brain	88.26	16.69	23
Spanish	Fable	87.83	16.44	21
Ukrainian	Brain	77.37	18.21	19
Ukrainian	Fable	67.46	15.39	14
Control	Fable	50.88	15.41	19

Note: Each row represents a group defined by students' native language and the activity attended (Brain = neuroscience lecture; Fable = fable reading). Ukrainian groups used the STTT system. Control group did not attend the activity and serve as a baseline for comprehension scores.
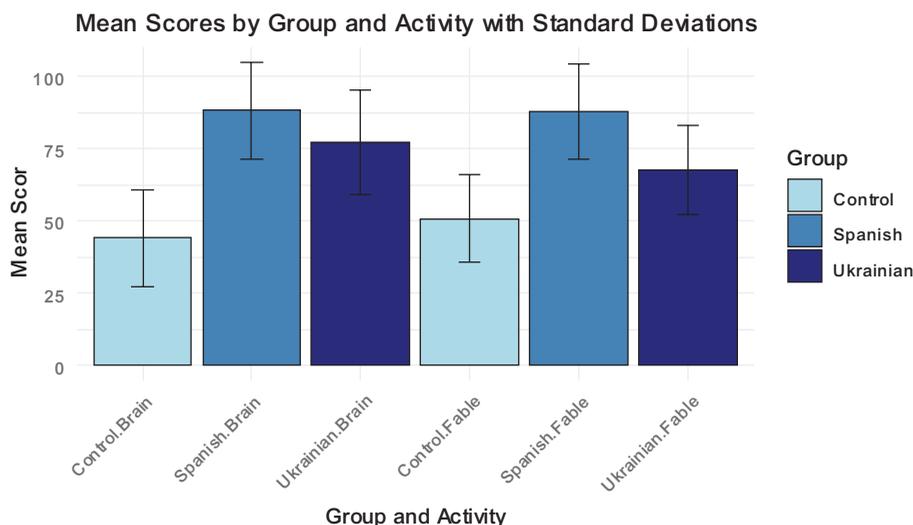

Figure 1. Mean scores by group and activity with standard errors. Error bars represent standard deviations. Group names indicate the students' native language and the activity completed (Brain = neuroscience lecture; Fable = fable reading). Ukrainian groups used real-time speech-to-text translation (STTT). Control groups did not attend the activities and completed the comprehension questionnaires without prior exposure to the content.

Ukrainian participants using the STTT showed marked improvement compared to the Control group in both tasks.

Normality was assessed for each group using the Shapiro–Wilk test, revealing significant deviations in the Spanish Brain ($p < 0.001$), Spanish Fable ($p < 0.001$), and Ukrainian Brain ($p = 0.005$) groups, while other groups did not deviate significantly ($p > 0.05$). Levene's test confirmed homogeneity of variances across groups ($F(5, N) = 0.34, p = 0.887$). These results justified the use of non-parametric methods for subsequent analyses.

The Kruskal–Wallis test revealed statistically significant differences among the six groups ($\chi^2(5) = 68.47, p < 0.001$). Post hoc pairwise comparisons were conducted using the Mann-Whitney U test with Bonferroni-adjusted p-values to control for Type I error. Significant differences were identified in both activities (Table 2). For the Brain activity, significant differences were observed between the Control Brain and Spanish

Brain groups ($p < 0.001$), as well as between Control Brain and Ukrainian Brain ($p < 0.001$). For the Fable activity, significant differences were found between Control Fable and Spanish Fable ($p < 0.001$), Control Fable and Ukrainian Fable ($p < 0.05$), and Spanish Fable and Ukrainian Fable ($p < 0.001$).

These results suggest that STTT improved comprehension among Ukrainian students, enabling them to outperform control groups in both activities and effectively reduce language barriers.

The effect size analysis (Cliff's Delta) suggests significant and large effects for several group comparisons, highlighting the potential impact of the intervention on the Ukrainian children (Table 3). The use of the STTT with Ukrainian children revealed a large effect size when comparing Ukrainian Brain and Ukrainian Fable groups to the Control Brain and Control Fable groups. Specifically, the comparison between Control Brain and Ukrainian Brain yielded a Cliff's Delta of -0.828 ($p < 0.001$), while the comparison between Control Fable and Ukrainian Fable produced a Cliff's Delta of 0.940 ($p < 0.001$). These findings indicate a substantial improvement in performance in the Ukrainian group for both activities.

Comparisons between the Ukrainian Brain and Ukrainian Fable groups themselves revealed only small effect sizes ($|d| = 0.083$, $p = 0.706$), indicating that performance differences within the Ukrainian group were minimal. Similarly, comparisons within the Spanish groups also yielded small to medium effect sizes, such as for Spanish Brain versus Spanish Fable ($|d| = 0.085$, $p = 0.642$).

Table 2. Results of Kruskal-Wallis analysis and post hoc comparisons.

Group comparison	Group 2	P	Activity
Control	Spanish	< 0.001	Brain
Control	Ukrainian	< 0.001	Brain
Control	Spanish	< 0.001	Fable
Control	Ukrainian	< 0.05	Fable
Spanish	Ukrainian	< 0.001	Fable

Note: Significant comparisons ($p < 0.05$, Bonferroni-adjusted) with activities specified. Each comparison involves two groups defined by language and activity type (Brain = neuroscience lecture; Fable = fable reading). All Ukrainian groups used the STTT system; Control groups did not attend the activity.

Table 3. Effect sizes and statistical significance for group comparisons.

Group 1	Group 2	Cliff's delta	Effect size interpretation	P	Adjusted P-value
Control Brain	Spanish Brain	-0.95	Large	0.000	0.000
Control Brain	Ukrainian Brain	-0.83	Large	0.000	0.000
Control Brain	Control Fable	-0.05	Small	0.786	1.000
Spanish Brain	Spanish Fable	0.08	Small	0.642	1.000
Spanish Brain	Ukrainian Brain	0.38	Large	0.035	0.524
Spanish Fable	Ukrainian Fable	0.40	Large	0.048	0.718
Spanish Fable	Control Fable	0.99	Large	0.000	0.000
Ukrainian Brain	Ukrainian Fable	0.08	Small	0.706	1.000
Ukrainian Fable	Control Fable	0.94	Large	0.000	0.000

Note: P-values adjusted using Bonferroni correction. Cliff's Delta interpretation: Small ($|d| \leq 0.147$), Medium ($0.147 < |d| \leq 0.33$), Large ($|d| > 0.33$). Group names indicate students' native language, and the activity completed (Brain = neuroscience lecture; Fable = fable reading). Ukrainian groups used the STTT system; Control groups did not attend the activity.

In summary, the analysis suggests large and statistically significant effects of the STTT intervention on Ukrainian children, although limitations in control group composition and the low number of participants should be taken into account.

Qualitative analysis of teachers' testimonials

The analysis of the semi-structured interviews was structured around the five pre-defined themes corresponding to the interview questions: (1) teachers' experiences with the STTT, (2) its impact on students' learning and participation, (3) technical and pedagogical challenges, (4) areas for improvement, and (5) recommendations for other teachers. These themes, along with their interrelations, are visually represented in Figure 2, which provides an overview of the relationships between the main categories and subcategories derived from the coding process.
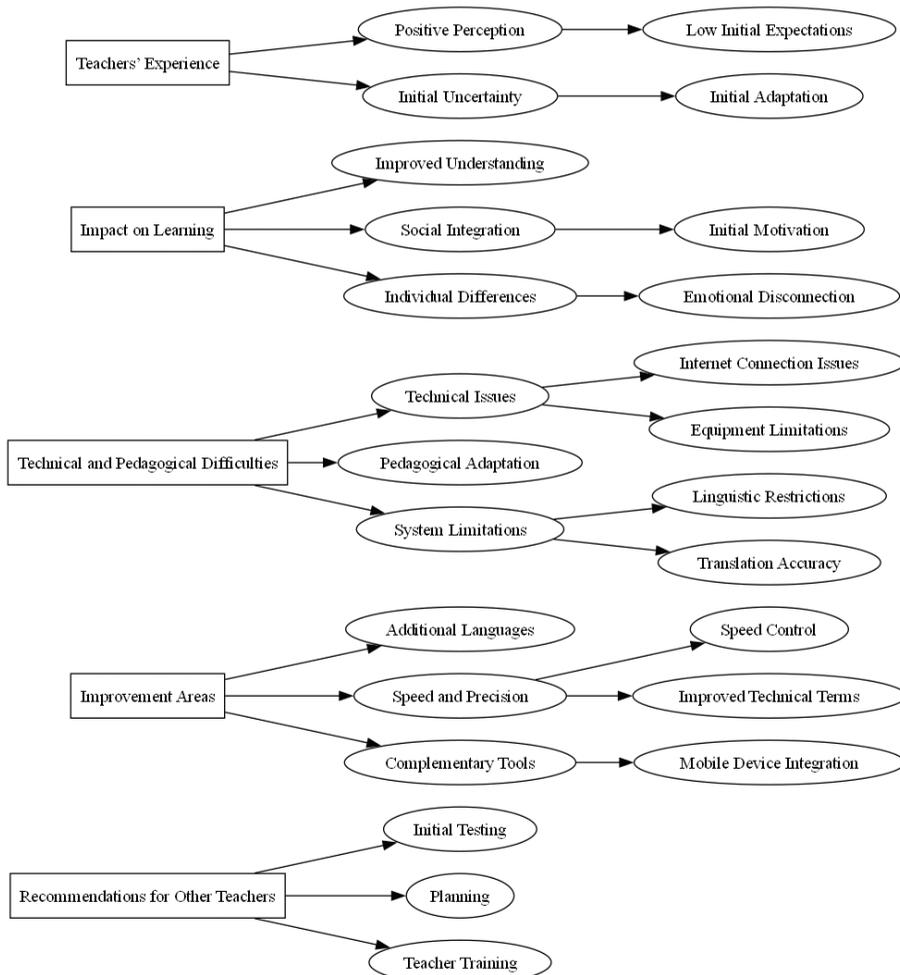
Figure 2. Thematic map of teachers' experiences, challenges, and recommendations related to the STTT system. The diagram represents the results of the qualitative analysis of transcriptions from the responses of five teachers after using the STTT system.

Teachers' experiences with the STTT system were predominantly positive. Participants highlighted the tool's potential to enhance classroom communication in multilingual contexts, although initial uncertainty was common. For example, one teacher stated, '*My initial expectations were not very high because I didn't know how it worked*' Despite these reservations, the system quickly suggested its utility, with teacher participants reporting positive surprises, such as improved student engagement. As one teacher remarked, '*From the very first moment, I noticed they connected somewhat with the class*'

The perceived impact on students' learning and participation was significant. Teachers reported improvements in comprehension, social integration, and group motivation. One participant noted, '*Students adapted effectively to this system for learning', while another emphasized that it 'reduced students' initial anxiety*' However, individual differences were apparent, as some students engaged enthusiastically while others quickly lost interest. Motivated students appeared to benefit the most from the system, according to the teachers.

Technical and pedagogical challenges were among the barriers to the effective use of the STTT system. Technical issues such as unreliable internet connections and the dependency on classroom projectors were frequently cited. One teacher noticed, '*Internet connection was a recurring problem*'. System limitations, including the inability to support multiple languages simultaneously and the occasional inaccuracy of translations, further complicated its implementation. On the pedagogical side, participants needed to adapt their teaching styles, often slowing down their speech and articulating more clearly, to improve the system's effectiveness.

Teachers identified areas for improvement that could enhance the STTT system's functionality and accessibility. Expanding language options, particularly for regional languages such as Valencian, was a recurring suggestion. Improving transcription speed and accuracy, particularly for specialised terminology, was identified as a key area for enhancement. Additional features, such as integration with mobile devices or the ability to send subtitles directly to students' phones, were proposed to improve accessibility and usability.

Finally, the participating teachers offered recommendations for colleagues planning to implement the STTT system. Many emphasised the importance of conducting initial trials with small groups to build familiarity and confidence. One teacher explained, '*Testing the system several times before regular use builds confidence*'. Effective planning was also crucial, particularly in managing pauses and adjusting the rhythm of explanations. Training was another recurring recommendation, with one participant asserting, '*Training on this tool should be mandatory for teachers*' and another highlighting the need to develop technical skills with complementary tools such as Office 365.

Discussion

Although voice recognition and speech-to-text technologies have existed for decades (Klatt, 1987; McTear, 2002), their use in classroom settings – particularly beyond second-language learning or learning disabilities – has been explored to a limited extent. Recent studies focusing on the university level have shown that speech recognition and translation tools significantly enhance comprehension in multilingual environments, particularly when visual and auditory modalities are combined to reinforce learning

(Shadiev et al., 2020; Shadiev et al., 2024). However, little to no research has specifically focused on secondary school students, and even less on vulnerable populations such as refugee participants.

This study evaluated the effectiveness STTT in improving the comprehension of educational content among Ukrainian refugee secondary school students with no knowledge of Spanish. The results indicate that, although Ukrainian students scored lower than their Spanish peers in academic-style tasks, their performance was significantly higher than that of the control group. While this difference suggests that STTT may contribute to reducing language barriers and improving comprehension, the design limitations – particularly the absence of a pretest and the lack of an equivalent control group – mean these findings should be interpreted as preliminary. Nonetheless, they provide promising indications of the tool's potential value, especially when considered alongside the qualitative insights provided by teachers. Importantly, it remains unclear whether the observed benefits would persist under conditions of regular classroom use. Future research should therefore examine the effects of sustained STTT use on students' day-to-day comprehension and academic progress, rather than focusing solely on one-off interventions.

Despite these limitations, the findings align with previous studies that highlight the role of speech recognition technologies in facilitating learning in multilingual contexts and second-language instruction (Shadiev & Liu, 2023), as well as improving literacy among students with learning difficulties (Matre & Cameron, 2022). Similar insights were reported by Ulum et al., (2025), who found that automatic translation tools such as Google Translate helped bridge communication gaps between Turkish police and refugee populations. Likewise, it has been reported that refugees generally perceived automated translation as helpful for navigating daily life, but found that these tools often failed in emotionally complex or context-dependent situations (Agrawal et al., 2023). In contrast, the STTT system in our study was embedded within structured classroom activities and supported by trained teachers, a setting that may help reduce some of these limitations by providing clearer communicative goals and pedagogical scaffolding.

Moreover, our results must be understood within the broader mental health challenges faced by refugee populations. Recent analyses have shown that Ukrainian refugees, particularly children and adolescents, experience elevated risks of post-traumatic stress disorder, depression, and anxiety due to cumulative war-related stress and displacement (Kapel Lev-Ari et al., 2024; Osokina et al., 2023). Factors such as separation from family members, exposure to violence, and adapting to a new cultural and linguistic environment contribute to these outcomes. However, protective elements such as supportive school environments and structured educational interventions such as STTT could contribute to bolster resilience and improve psychological well-being. These findings highlight the importance of integrating educational technologies into a framework that simultaneously supports learning and emotional well-being, aligning with recommendations for refugee mental health services (Osokina et al., 2023).

This perspective is further supported by broader policy-level insights, such as those presented by the European Union Institute for Security Studies (AbuJarour, 2022), which emphasised the importance of using information and communication technologies (ICT) in education to promote the integration of refugee students. Their report highlights how early access to digital educational tools – particularly in linguistically challenging environments – can improve both learning outcomes and social inclusion. The structured use of STTT in our study reflects this perspective, offering

a practical example of how educational technologies can contribute to both cognitive and emotional aspects of integration.

In our study, the comparison of Ukrainian and Spanish students revealed significant differences in both activities, with the native group outperforming the Ukrainian group. This difference, expected due to the language barrier and the imperfections of the STTT, highlights the need to continue optimising technological tools to further narrow the performance gap between non-native and native students (Shadiev & Liu, 2023).

The development of STTT systems has seen significant advancements in recent years, particularly with the emergence of end-to-end (E2E) models, which aim to overcome the cascading errors prevalent in traditional systems such as those used in PowerPoint 365 by unifying speech recognition and translation into a single framework (Sethiya & Maurya, 2024).

Despite the rapid advancements and effectiveness of STTT, it cannot replace the necessity of complementary pedagogical strategies or broader programmes focused on linguistic immersion and emotional support. In this context, the emotional impact on refugee students is critical, as highlighted by research on vulnerable populations (Save the Children, 2023).

From a qualitative perspective, teacher testimonies provide relevant complementary insights. Teachers agreed on the system's usefulness in fostering inclusion and student participation, which aligns with previous studies on assistive technologies applied to linguistically diverse contexts (Matre & Cameron, 2022). At the same time, it is important to acknowledge that teachers were aware of their participation in a study, which may have influenced their responses – a possibility consistent with the Hawthorne effect (Sedgwick & Greenwood, 2016).

Most teachers also pointed out technical challenges, such as the dependence on a stable internet connection and the need to adjust the teaching pace to facilitate translation. In particular, teachers noticed that latency and segmentation issues could momentarily hinder comprehension, especially during fast speech or when complex terminology was used. To reduce these effects, teachers reported adjusting their delivery, introducing pauses, or repeating key information to ensure clarity. These adaptations reflect a flexible pedagogical response to technological constraints, and highlight the importance of teacher training for effective implementation. These technical issues are consistent with limitations reported in the implementation of similar tools in other studies (Li, 2022; Mehrish et al., 2023). Suggestions such as offline functionality or the integration of more dynamic visual tools underscore the potential for improving these technologies in practical applications.

From a pedagogical perspective, and to enhance teacher motivation to use STTT when necessary, the qualitative information obtained in this study highlights the importance of preliminary practice for successful implementation. In addition, self-pacing the rhythm of speech and incorporating a dynamic delivery style have been shown to further enhance the effectiveness of these tools. While the integration of PowerPoint's STTT features and wireless microphones is technically straightforward – requiring only a simple plug-and-play setup and a few clicks in the software – teachers have often been hesitant to experiment with these tools. This suggests the importance of offering targeted training programmes that not only build educators' technical skills but also support the integration of digital tools into their personal and professional identities. Such initiatives can help teachers navigate the challenges of digital transformation in a way that aligns with their individual teaching practices (Clark, 2020).

Limitations and future work

A key strength of this study lies in its focus on a vulnerable and hard-to-reach population, complemented by the integration of quantitative and qualitative analyses to provide a comprehensive understanding of the system's benefits and limitations. While the moderate sample size poses a limitation (Creswell, 2015), the qualitative insights from five teachers, who reported positive experiences with the efficacy of the STTT system under routine classroom conditions, provide strong support for the findings.

In addition, several other methodological limitations must be acknowledged. The absence of a pretest prevented the establishment of baseline equivalence between groups. Moreover, the control and experimental groups differed in more than just their use of the STTT system, as the control group did not attend the classroom activities. In addition, the control groups included students from both Spanish- and Ukrainian-speaking backgrounds, each of whom completed the comprehension questionnaires in their native language. While this design allowed us to estimate baseline performance without prior exposure to the content, it may have introduced a source of variability. Notably, Spanish-speaking students tended to perform slightly better within the control groups, suggesting that any language-related bias would act against, rather than in favour of, the main hypothesis – thereby reinforcing the observed effect of the STTT system. A further limitation is the short exposure time to STTT, which was limited to approximately 15 min during both the fable and the neuroscience lecture, thus constraining the scope of the observed effects.

Furthermore, the comprehension measures used in the study were not formally validated using standardised psychometric methods. While we employed expert review and conducted an informal back-translation reliability assessment prior to data collection, future studies should incorporate formal validation procedures and collaborate with language education specialists to ensure the robustness of assessment tools. Additionally, the estimated 70% reliability of the STTT output highlights the current limitations of these technologies, as discussed by other authors (e.g. Chun & Lewis, 2022; Li, 2022; Mehrish et al., 2023; Shadiev et al., 2024), and points to the need for ongoing refinement.

Nevertheless, as the initial influx of refugee students has decreased and many have since acquired proficiency in the local language, recruiting additional participants for similar studies has become increasingly challenging. Future research should address these challenges by employing larger and more diverse samples across multiple educational contexts. It should also consider the dynamic nature of language acquisition, particularly in contexts where learners start with no prior knowledge of the target language. The use of STTT systems is especially valuable during the initial stages, offering crucial support in challenging situations, such as those faced by refugee students. However, as learners reach a certain threshold of proficiency in the local language, a gradual transition from STTT to STT could serve as an effective tool to reinforce comprehension and support continued language development until reaching a certain average level (Matthew, 2020). This phased approach could help balance the immediate need for translation with the long-term goal of fostering independent language skills.

Conclusion

Taken together, the quantitative outcomes and qualitative teacher feedback suggest that STTT can support academic content comprehension and promote inclusion among

refugee secondary school students with limited proficiency in the local language. Maximising its impact requires enhancing technical robustness, adding complementary pedagogical resources, and implementing targeted teacher training. These findings underscore the role of technology in reducing educational barriers and highlight the need for further research on its application in multilingual and vulnerable contexts.

The importance of such solutions is amplified by the challenges faced by children in war-affected regions, like those from Ukraine, where the conflict has severely impacted physical and mental health. These children face disrupted access to basic services and significant mental health challenges, worsened by the destruction of schools and healthcare facilities. Addressing these barriers demands both immediate humanitarian efforts and long-term strategies incorporating innovative tools such as STTT to foster inclusion, resilience, and well-being in multicultural settings.

Contributions

RS conducted the project and analysed data. RS / CV / DPA designed experiments. CV / DPA / RI / OV / JMPT provided teaching experience with STTT and coordinated with secondary schools. JB / CdlP / TPR / MJAA contributed to methodological review. RS wrote the manuscript, and TPR / JB / CdlP / MJAA contributed to manuscript revision.

References

AbuJarour, S. (2022). Integration through education: Using ICT in education to promote the social inclusion of refugees in Germany. *Journal of Information Systems Education, 33*(1), 51–60. Retrieved from https://aisel.aisnet.org/jise/vol33/iss1/7

Agrawal, A. et al. (2023, October 9–13). All translation tools are not equal: Investigating the quality of language translation for forced migration. In *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1–10). IEEE. https://doi.org/10.1109/DSAA60987.2023.10302481

Alvarez-Alonso, M. J. et al. (2021). Boys-specific text-comprehension enhancement with dual visual-auditory text presentation among 12–14 years-old students. *Frontiers Psychology, 12*, 574685. https://doi.org/10.3389/fpsyg.2021.574685

Berner, K., & Alves, A. N. (2023). A scoping review of literature using speech recognition technologies by individuals with disabilities in multiple contexts. *Disability and Rehabilitation: Assistive Technology*, *18*(7), 1139–1145. https://doi.org/10.1080/17483107.2021.1986583

Bernstein, J. et al. (1990). Automatic evaluation and training in English pronunciation. *ICSLP, 90*, 1185–1188. https://doi.org/10.21437/ICSLP.1990-313

Besters-Dilger, J. (2023). Language policy in Ukraine-overview and analysis. *Ukrainian Analytical Digest, 1*, 2–6. https://doi.org/10.3929/ethz-b-000623475

Chen, K. T. C. (2022). Speech-to-text recognition in University English as a Foreign Language Learning. *Education and Information Technologies, 27*(7), 9857–9875. https://doi.org/10.1007/s10639-022-11016-5

Clark, D. (2020). Tech and me: An autoethnographic account of digital literacy as an identity performance. *Research in Learning Technology, 28*, 2389. https://doi.org/10.25304/rlt.v28.2389

Creswell, J. W. (2015). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.

Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning, 36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

Dew, K. N. et al. (2018). Development of machine translation technology for assisting health communication: A systematic review. *Journal of Biomedical Informatics, 85*, 56–67. https://doi.org/10.1016/j.jbi.2018.07.018

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology, 2*(1), 54–73. https://doi.org/10.64152/10125/25032

Furui, S. et al. (2004). Speech-to-text and speech-to-speech summarization of spontaneous speech. *IEEE Transactions on Speech and Audio Processing, 12*(4), 401–408. https://doi.org/10.1109/TSA.2004.828699

Gao, J. et al. (2022). Review of the application of intelligent speech technology in education. *Journal of China Computer-Assisted Language Learning, 2*(1), 165–178. https://doi.org/10.1515/jccall-2022-0004

Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 195–202. https://doi.org/10.1177/2372732215602130

Huang, Y. M., Shadiev, R., & Hwang, W. Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education, 101*, 15–28. https://doi.org/10.1016/j.compedu.2016.05.011

Instituto Nacional de Estadística (INE). (2024). *Flujos de estudiantes ucranianos por meses*. Retrieved October 3, 2025, from https://public.tableau.com/app/profile/instituto.nacional.de.estad.stica/viz/FlujosUcranianosMeses/Dashboard4

Ivanova, O. (2013). Bilingualism in Ukraine: Defining attitudes to Ukrainian and Russian through geographical and generational variations in language practices. *Sociolinguistic Studies, 7*(3), 249–272. https://doi.org/10.1558/sols.v7i3.249

Jeon, J., Lee, S., & Choi, S. (2023). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments, 32*(8), 4613–4631. https://doi.org/10.1080/10494820.2023.2204343

Kapel Lev-ari, R., Aloni, R., & Ben-ari, A. (2024). Understanding the dyadic mental health of refugee parents and children after fleeing the 2022 Ukraine war. *Psychological Trauma: Theory, Research, Practice, and Policy*. https://doi.org/10.1037/tra0001715

Klatt, D. H. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America, 82*(3), 737–793. https://doi.org/10.1121/1.395275

Li, J. (2022). Recent advances in end-to-end automatic speech recognition. *APSIPA Transactions on Signal and Information Processing, 11*(1), e8. https://doi.org/10.1561/116.00000050

Matthew, G. (2020). The effect of adding same-language subtitles to recorded lectures for non-native, English speakers in e-learning environments. *Research in Learning Technology, 28*, 2340. https://doi.org/10.25304/rlt.v28.2340

Matre, M. E., & Cameron, D. L. (2022). A scoping review on the use of speech-to-text technology for adolescents with learning difficulties in secondary education. *Disability and Rehabilitation: Assistive Technology, 19*(3), 1103–1116. https://doi.org/10.1080/17483107.2022.2149865

McKechnie, J. et al. (2018). Automated speech analysis tools for children's speech production: A systematic literature review. *International Journal of Speech-Language Pathology*, *20*(6), 583–598. https://doi.org/10.1080/17549507.2018.1477991

McTear, M. F. (2002). Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, *34*(1), 90–169. https://doi.org/10.1145/505282.505285

Mehrish, A. et al. (2023). A review of deep learning techniques for speech processing. *Information Fusion, 99*(19), 101869. https://doi.org/10.1016/j.inffus.2023.101869

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of ChatGPT. *International Journal of Qualitative Methods, 22*, 1–10. https://doi.org/10.1177/16094069231211248

Nickolai, D., Schaefer, E., & Figueroa, P. (2024). Aggregating the evidence of automatic speech recognition research claims in CALL. *System*, *121*, 103250. https://doi.org/10.1016/j.system.2024.103250

OpenAI. (2024). *ChatGPT* [Large language model]. Retrieved from https://chat.openai.com/

Osokina, O. et al. (2023). Impact of the Russian invasion on mental health of adolescents in Ukraine. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(3), 335–343. https://doi.org/10.1016/j.jaac.2022.07.845

Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023, July 19–21). Speech-to-text and text-to-speech recognition using deep learning. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 657–666). IEEE.

Save the Children. (2023). *Back to School 2023–2024: Report on Education for Children Displaced by the Conflict in Ukraine at the Start of the Second School Year*. Save the Children.

Scott, R. et al. (2022). Transcripción simultánea de voz a texto en el aula como medio de inclusión lingüística. En: Satorre Cuerda, R. (Ed.). *El Profesorado, Eje Fundamental de la Transformación de la Docencia Universitaria.* Octaedro, 416–427. ISBN 978-84-19506-52-8

Sedgwick, P., & Greenwood, N. (2016). Understanding the Hawthorne effect. *BMJ, 2015,* 351. https://doi.org/10.1136/bmj.h4672

Sethiya, N., & Maurya, C. K. (2025). End-to-end speech-to-text translation: A survey. *Computer Speech & Language*, *90*, 101751. https://doi.org/10.1016/j.csl.2024.101751

Shadiev, R., Chen, X., & Altinay, F. (2024). A review of research on computer-aided translation technologies and their applications to assist learning and instruction. *Journal of Computer Assisted Learning*, *40*(6), 3290–3323. https://doi.org/10.1111/jcal.13072

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), 215824402095317. https://doi.org/10.1177/2158244020953177

Shadiev, R., Huang, Y. M., & Hwang, J. P. (2017a). Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation. *Educational Technology Research and Development, 65*, 1239–1261. https://doi.org/10.1007/s11423-017-9516-3

Shadiev, R. et al. (2017b, August 1–4). Are STR & CAT-generated texts useful for comprehension of lecturing content in a foreign language? In *2017 10th International Conference on Ubi-Media Computing and Workshops (Ubi-Media)* (pp. 1–6). IEEE. https://doi.org/10.1109/umedia.2017.8074121

Shadiev, R. et al. (2014). Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society*, *17*(4), 65–84.

Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL. 35*(1), 74–88. https://doi.org/10.1017/S095834402200012X

Shadiev, R., & Sun, A. (2019). Using texts generated by STR and CAT to facilitate student comprehension of lecture content in a foreign language. *Journal of Computing in Higher Education, 32*(3), 561–581. https://doi.org/10.1007/s12528-019-09246-7

Shew, A. (2020). Ableism, technoableism, and future AI. *IEEE Technology and Society Magazine, 39*(1), 40–85. https://doi.org/10.1109/MTS.2020.2967492

Smotrova, T. (2009). Globalization and English language teaching in Ukraine. *TESOL Quarterly, 43*(4), 727–732. https://doi.org/10.1002/j.1545-7249.2009.tb00200.x

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, *14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

The World Bank. (2022). *Displaced Education in Ukraine: Impact and Responses*. The World Bank.

Ulum, Ö. G. (2025). Refugee voices unheard: Bridging the communication divide between Turkish police and refugees. *Journal of Immigrant & Refugee Studies*, *22*, 1–22. https://doi.org/10.1080/15562948.2025.2529482

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/S0261444800012970