

University of Tartu
Faculty of Social Sciences
Institute of Education
Educational Technology

Tran Van Khanh Linh

**BRIDGING THE BLACK BOX: A QUALITATIVE EXPLORATORY
STUDY ON ENHANCING TEACHER VISIBILITY INTO AI-ASSISTED
EFL SPEAKING PRACTICE THROUGH CHATBOT-MEDIATED
LEARNER REFLECTIONS**

Master's thesis

Supervisor: Professor Emanuele Bardone, PhD.
Programme Director for Educational Technology

Tartu 2026

Kokkuvõte/ Abstract

Käesolev kvalitatiivne eksploratiivne uuring keskendub sellele, kuidas juturoboti vahendatud õppijate refleksioonid võivad aidata inglise keele kui võõrkeele õpetajatel saada paremat ülevaadet õpilaste iseseisvast kõnepraktikast. Uuringu raames intervjueriti viit Vietnami inglise keele kui võõrkeele õpetajat, et uurida probleeme, millega nad puutuvad kokku siis, kui õpilased harjutavad kõnelemist tehisintellektil põhinevate tööriistade abil väljaspool klassiruumi, millist refleksiooniandmestikku nad peavad oma õpetamise jaoks kasulikuks ning kuidas nad hindavad reflectEFL prototüüpi. Tulemused näitavad, et õpetajatel puudub sageli teave õpilaste iseseisva praktika kohta, eriti suuremates klassides. Osalejad pidasid kasulikuks refleksiooniandmeid harjutatud teemade, harjutamise kestuse, keeleprobleemide, õpilaste enesekindluse taseme ja toetusvajaduste kohta. Üldiselt peeti prototüüpi tunniplaneerimise seisukohalt kasulikuks, kuigi esines muresid kontekstuaalse teostatavuse ja õpilaste aususe pärast.

Märksõnad: tehisintellekti toetatud keeleõpe, inglise keel võõrkeelena, kõnelemisoskus, õppija refleksioon, juturobot, õpetaja nähtavus, enesereguleeritud õppimine.

This qualitative exploratory study examines how chatbot-mediated learner reflections may help EFL teachers gain visibility into students' independent speaking practice. Five Vietnamese EFL teachers were interviewed to explore the problems they encounter when students practice speaking with AI tools outside class, the types of reflection data they find useful for teaching, and their evaluation of the reflectEFL prototype. The findings reveal that teachers often lack information about students' independent practice, especially in larger classes. Participants reported that reflection data regarding practice topics, duration, language challenges, students' confidence levels, and support needs was helpful. The prototype was generally regarded as beneficial for lesson planning; however, concerns persisted regarding its contextual feasibility and the reliability of students' self-reported information.

Keywords: AI-assisted language learning, EFL speaking, learner reflection, chatbot, teacher visibility, self-regulated learning.

Table of Contents

1 Introduction	4
2 Literature review.....	5
2.1 AI-powered tools and the invisibility of independent speaking practice.....	6
2.2 Learner reflection as a bridge between independent practice and classroom instruction ...	7
2.3 The teacher's role in AI-assisted EFL learning	8
3 Methodology	9
3.1 Research design.....	9
3.2 Research questions.....	10
3.3 The participants	10
3.4 The prototype	11
3.5 Data collection and analysis	14
3.6 Ethical considerations.....	15
4 Findings	15
4.1 RQ1: Challenges teachers face when students engage in unmonitored AI-assisted speaking practice.....	15
4.2 RQ2: What types of learner reflection data do teachers consider most actionable.....	16
4.3 RQ3: How teachers evaluate the usefulness and feasibility of the reflectEFL prototype...	18
5 Discussion	19
5.1 The visibility gap is real but context-dependent	19
5.2 Reflection data has clear diagnostic value for teachers.....	21
5.3 Feasibility depends on student honesty and teacher workload	22
5.4 The chatbot as a bridge, not a replacement.....	22
6 Conclusion	24
6.1 Summary of findings	24
6.2 Contributions of the study	24
6.3 Limitations.....	25
6.4 Implications for practice	25
6.5 Directions for future research	26
7 Author's declaration.....	26
References.....	28
Appendix A. Interview guide	

1 Introduction

In the last few years, the ways EFL learners practice English speaking outside the classroom have undergone a dramatic change. Previously, independent speaking practices relied on limited resources such as conversation partners, language labs, or self-study recordings. With the rise of technology, particularly AI-powered chatbots, this landscape has fundamentally changed. Learners can now have access to interactive, feedback-rich speaking practice at any time without relying on institutional support. Platforms like ChatGPT Voice, ELSA Speak, and Mondly allow students to get immediate feedback on their pronunciation, engage in real-time conversations, and practice a diverse array of subjects outside the traditional classroom confines.

Academic research already exists regarding the benefits that these tools may bring to students' speaking practice. Studies have been conducted in multiple academic settings with diverse levels of success regarding English fluency. The findings consistently show that AI-assisted speaking practice reduces the level of anxiety and increases learners' willingness to speak and improves their fluency, pronunciation, and grammatical structures (Nguyen, 2025; Nguyen et al., 2025; Ding & Yusof, 2025; Nuñez et al., 2025). These tools provide additional practice for those who lack sufficient chances for engaging in English conversations because of large class sizes or non-English-speaking environments.

Yet, this expansion has created a structural problem that research literature has not fully addressed. Learners engage in AI-assisted speaking practice independently outside institutional systems, resulting in a lack of accessible learning data for teachers. Teachers have little to no visibility into what students practiced, what they did well, and what they found difficult during their practice sessions. This issue is not simply a matter of missing information. Accumulated practice alone cannot result in speaking development; it also depends on cycles of performance, feedback, and uptake, which require teacher involvement to be pedagogically effective (Lyster & Ranta, 1997). When teacher visibility is limited in those independent speaking practice sessions, those cycles are interrupted. Research indicates that teachers' presence has a highly positive impact on the emotional and agentic engagement of students in language learning using GenAI chatbots, as compared to students learning without teachers' presence, who had significantly less of both types of engagement (Li, Wu & Chiu, 2025). The problem is not that students practice with AI-powered tools independently; it is that teachers lack the opportunities to capitalize on them.

This visibility gap is further magnified by a broader pattern that has been observed in the CALL literature. As noted by Lai et al. (2015), teachers tend to underestimate the mediating

roles they play between students' out-of-class technology-based activities and their instruction, failing to adopt a more active stance even when students expect and need greater involvement. With the emergence of those speaking applications that integrate AI, this passivity of teachers has become not just missed opportunities but an actual gap in the learning process. The more students practice with those tools outside the classroom, the wider the gap between independent practice and teacher-led instruction becomes, making it harder for teachers to plan lessons that are suitable for what students have already learned.

This study proposes to bridge this gap through a chatbot-mediated learner reflection approach. After students practice speaking independently, they receive a prompt from the chatbot to answer a list of guided reflective questions that ask them to think about what they learned, what they did well, what problems they encountered, how confident they were, and what help they need. The responses are then summarized by the chatbot and displayed on the instructor's dashboard together with some ideas for future lessons. This approach is backed up by findings from relevant research. First, structured reflections promote self-directed learning in an L2 context (Lee & Mori, 2021; Do, 2026) and provide teachers with actionable pedagogical insights (Do, 2026). Second, reflection can be viewed as the adaptive component of the process of self-regulated learning when learners reflect on their performance and modify their behavior (Winne & Hadwin, 1998). Finally, when students' reflective reports are collected, analyzed, and shared with instructors, there are notable achievements in teachers' adaptability and better academic performance by the learners (Menekse, 2020). What is particularly important about this approach is that it does not require any platform integration, proprietary AI data access, or any resources from the institutions, which makes it viable across a wide range of EFL teaching contexts.

Using an exploratory design study approach, this paper examines the difficulties arising from unobserved AI-driven speaking practice from the perspective of EFL teachers and the kinds of reflective data that they perceive as being valuable enough for them to act upon. In addition, the study also focuses on evaluating the effectiveness and usefulness of the chatbot-generated summary of reflection data as a tool for teachers preparing lessons before class. It seeks to build a practical, yet low-cost, model for bridging the gap between students' use of AI outside of the classroom and the lessons taught by the teacher in class.

2 Literature review

2.1 AI-powered tools and the invisibility of independent speaking practice

Past research has shown that the use of AI-based language learning tools, such as chatbots and other AI voice feedback tools, is effective in reducing foreign language speaking anxieties for EFL learners. Moreover, the use of such tools enhances the willingness to communicate for language learners. They provide learners with opportunities to practice their speaking skills without the social consequences that may take place during their interaction with the teacher or peers. Because many AI speaking tools provide private, repeatable, and low-pressure practice opportunities, learners may feel less exposed to the social pressure often associated with speaking in front of teachers or peers.

Nguyen (2025) carried out a 16-week quasi-experimental research among 240 Vietnamese university learners from four different institutions. The treatment group, who used AI-based tools to practice their speaking skills, performed better than the control group, who practiced speaking the traditional way, with an average effect size of $d = 1.11$. Moreover, the research showed that AI tools were more helpful in improving performance-based components of speaking, such as fluency and pronunciation, due to the provision of immediate feedback. Similar results have also been reported in other EFL contexts. Nguyen et al. (2025) conducted a study and found that chatbots and AI voice feedback tools reduced foreign language speaking anxiety and promoted willingness to communicate among primary school Vietnamese students. In a similar study, Ding and Yusof (2025) also showed that the use of the Mondly AI chatbot helped develop students' speaking abilities, at the same time lowering their speaking anxieties. According to Nuñez et al. (2025), ChatGPT Voice offered learners a non-judgmental speaking context where they enhanced their pronunciation, fluency, vocabulary, and grammar skills through consistent usage. They were more confident during the interactions due to the feedback obtained from the above-mentioned platform. Overall, findings of these studies confirm the existing literature on the benefits of AI-based language learning tools.

Most of the studies that have been conducted on the effectiveness of AI tools in teaching language have been from the learner's perspective. Compared with learner outcomes, the teacher's role in chatbot-assisted language learning has received less attention. The teacher's role remains underexamined in the existing literature. In the case of independent AI speaking practice, students access the tools using mobile applications and websites on their own terms. This happens outside the school system, so there is no record of how they actually practice. Therefore, the teacher cannot access any information about the practice sessions when students use the tools.

The above scenario has further implications regarding how AI is theorized within the educational research field. Li, Zhou & Chiu (2025) note that because most research on AI chatbots in language learning has concentrated mainly on learners' outcomes, there is still a lack of understanding of how the tool can be used by teachers to enhance their teaching skills. Thus, although some students might be making great strides through independent practice, their teacher will remain oblivious to this fact and will not have the opportunity to capitalize on it.

2.2 Learner reflection as a bridge between independent practice and classroom instruction

The lack of data on the use of these applications is not the only problem. Assuming that the teachers were able to see the students' log files of these applications, still, this information would not be enough for the teachers to understand the process of learning from the perspective of their students. It means that there is no knowledge about which tasks the students tried to accomplish, which challenges they faced, and how certain they felt while working with AI systems.

In this situation, learner reflection becomes relevant because it allows students to evaluate their own practice and identify learning needs. Reflection can support second language learners in becoming more aware of their learning experiences, identifying difficulties, and planning future improvement. This is consistent with Lee and Mori's (2021) finding that reflective practices are related to self-directed learning competencies in second language university classes, and with Do's (2026) study showing that learner reflection in Vietnamese EFL education can provide both learning value for students and pedagogical insights for teachers.

The importance of reflection is also supported by self-regulated learning theory. In Winne and Hadwin's (1998) model, adaptation is one phase of self-regulated learning. In this phase, learners reflect on previous studying and make adjustments for similar tasks in the future. Reflection is therefore not only an endpoint of learning, but part of the process through which learners prepare for future learning.

Reflective practice also plays an important part in learning beyond the learner themselves. Through reflective practice, the student gives their educator vital information regarding what they struggle with, their confidence in their knowledge, and what they believe is needed to assist them further. These are the kinds of information that the teachers would not get access

to by any other way. In a direct test of this idea, Menekse (2020) implemented a Reflection-Informed Learning and Instruction approach in which students, after each lesson, were prompted to reflect on what they found confusing in each lesson using a mobile application. All the responses provided were automatically analyzed and then made available for the instructor after 36 hours. Students in this experimental group demonstrated a significantly higher performance level than those in the control condition with a large effect size (Cohen's $d = .82$). For the present study, Menekse's model is useful not only because it reports improved academic performance, but because it explains how student reflections can become diagnostic input for instructors. Student reflections reveal learning gaps, enabling instructors to adjust pacing, tailor feedback, and direct learners toward appropriate support. Despite this potential, reflection remains under-supported in many chatbot-based learning environments. Guan et al. (2024) reviewed 27 studies on educational chatbots and self-regulated learning and found that most chatbots supported students mainly in identifying learning resources, enacting learning strategies, and monitoring studying. Much less support was provided for goal setting, planning, reflection on prior studying, and adaptation to future studying.

This is where a chatbot-mediated reflection approach becomes relevant for EFL contexts. Rather than trying to pull data out of AI-assisted speaking platforms, teachers can make use of a chatbot that can simply ask students what they experienced after independent practice and compile their responses into a summary that the teachers can actually use before class. The mechanism Menekse (2020) proposed still applied: student reflections as diagnostic input for teacher planning. What changes is the setup, which does not require any platform integration, data-sharing agreements, or institutional buy-in.

2.3 The teacher's role in AI-assisted EFL learning

The remaining question is why teacher involvement matters and why this involvement can be difficult to maintain when students use technology outside class. Guan et al. (2025) show that AI integration in EFL education can be understood as a balanced interaction among teachers, students, and AI technologies, in which AI mediates rather than replaces teacher-student interaction. Their study also found that students who learned through AI-mediated teacher-student interaction achieved better proficiency than those in the comparison group. This supports the view that AI-assisted learning still depends on the teacher's pedagogical role. A similar issue appears in Lai et al.'s (2015) study of autonomous technology-based language learning outside the classroom. They found a mismatch between students' and teachers'

expectations of teacher involvement. Students expected teachers to play a greater role in recommending technological resources and sharing strategies for using them effectively, whereas teachers tended to expect a more minimal role. This helps explain why the gap between students' independent AI use and teacher-led instruction may remain unresolved without a practical communication channel.

These insights, combined, suggest a structural problem that can be resolved through chatbot-mediated reflection. When teachers are presented with a structured summary of what their students have practiced, what they found difficult, and what support they need before the next class, it becomes harder for teachers to adopt a passive position, not because they are coerced into action, but because the information makes the need visible. The reflectEFL prototype is designed around this logic: not to replace teacher judgment, but to give teachers the visibility they need to exercise it.

3 Methodology

3.1 Research design

A qualitative exploratory research design was employed in this study. Qualitative research is an approach used for exploring and understanding the meanings individuals ascribe to social and human problems. The characteristics of such an approach are emerging research questions, inductive data analysis, and interpretive interactions with the perspectives of research participants (Creswell, 2014). An exploratory research design is chosen for this study because the object of the analysis, which is the visibility gap between students' independent AI speaking practices and teachers' instructional planning, has not been sufficiently theorized and researched in CALL studies. The goal of the study was to develop an initial understanding of this problem and examine teachers' perspectives on a proposed conceptual response. This aligns with qualitative exploratory research, which is suitable for complex or little-known issues and for examining participants' experiences and perspectives (Lösch et al., 2023).

Semi-structured interviews were used to collect data from five Vietnamese EFL teachers. This method was appropriate because the study required participants to describe their teaching contexts, their awareness of students' independent AI use, and their reactions to a prototype that was still conceptual rather than fully implemented. The interviews served two purposes: first, to identify the participants' challenges in tracking and monitoring students' unsupervised AI-assisted speaking practice; and second, to obtain initial feedback on the chatbot prototype

as a possible response to the identified problem. Participants were informed that the prototype was a low-fidelity concept rather than a fully functioning system and were invited to comment on its relevance, utility, and feasibility in their own teaching contexts.

3.2 Research questions

This study addresses three research questions:

1. What challenges do EFL teachers face in planning speaking instruction when students engage in unmonitored AI-based speaking practice outside the classroom?
2. What types of learner reflection data do EFL teachers consider most actionable for informing follow-up speaking instruction after students' independent AI practice?
3. How do EFL teachers evaluate the usefulness and feasibility of a chatbot-generated reflection summary as a pre-class teaching aid for speaking lessons?

3.3 The participants

Five EFL teachers from different teaching contexts were selected for this study through the researcher's network. The selection was done purposefully to ensure the inclusion of diverse instructional settings where AI-assisted independent practice is plausible taking place among students. In comparison to quantitative research that uses random sampling in an attempt to generalize the results, qualitative research utilizes purposive sampling to choose those individuals who have experienced firsthand the core phenomenon being studied (Creswell, 2014). Table 1 summarizes their profiles.

The sample size was considered appropriate for the exploratory purpose of the study. The aim was not to produce statistically generalizable findings, but to obtain initial insights into how teachers from different EFL teaching contexts perceive the problem of limited visibility into students' independent AI-assisted speaking practice. The five participants represented different instructional settings, including university teaching, private language centers, and home tutoring. This variation made it possible to examine whether the perceived relevance of the reflectEFL approach differed according to class size, student level, and teaching context.

Table 1. *Participant Profiles*

	Teaching context	Class size	Student level	AI-related observation
T1	University	25-40	A1–A2	Learners rarely mention their use of AI; few questions regarding unintended encounters with AI; possible use of AI for writing assignments outside class
T2	Private language centre	1-6	A1–B1	Focusing on Cambridge exams; no reports of AI use in speaking tasks; possible use of AI for writing
T3	Private language centre	1-6	A1–B1	Majority adults; one student brought Duolingo recordings to class to seek feedback on self-study effectiveness
T4	Home tutoring	1-4	A1–A2	Focused on remediation of grammar and vocabulary; learners rarely undertake independent AI learning
T5	Home tutoring	1-8	A1–B1	Focused on grammar and writing; one adult student used Duolingo and the Coach app

3.4 The prototype

The prototype is named *reflectEFL*, emphasizing the pedagogical value of reflective learning in EFL contexts. It served as a stimulus tool rather than as a deployed intervention. The prototype comprised two components, namely, a reflection chatbot for the students and a dashboard summarizing students' reflections through AI for the teachers. The prototype was introduced to the participants during the interview in order to elicit feedback on RQ3.

Participants were clearly informed that the prototype was low-fidelity and asked to provide feedback on the concept and the idea rather than the technical performance.

As illustrated in Figure 1, the student-facing component contains a chatbot that guides students through post-practice reflection in a bilingual format (English and Vietnamese). The chatbot asks students questions about the topic practiced, platform used, difficulties encountered, confidence level, and any particular support needed from the teacher. In order to keep the answers brief and save time for the students, the chatbot provides students with pre-set quick-reply options alongside open text input, allowing them to choose from suggested responses where appropriate rather than typing freely throughout. The bilingual setup is used to help students overcome language problems and complete the reflection. One additional feature of the chatbot is a “skip” option for selected questions (e.g., optional comments on

confidence or additional feedback to the teacher). This allows students to skip responses that are not critical to the core reflection data.

Figure 1. Student-facing chatbot interface of reflectEFL (bilingual English/Vietnamese)

The figure displays the student-facing chatbot interface of reflectEFL, which is bilingual (English/Vietnamese). The interface is divided into five panels showing the chatbot's interaction with a student.

Panel 1: The chatbot greets the student and asks, "What did you practice today?". The student responds, "My hobbies / free time".

Panel 2: The chatbot asks, "Which hobby did you talk about?". The student responds, "reading books".

Panel 3: The chatbot asks, "Got it! Which app or platform did you use?". The student responds, "ChatGPT / ChatGPT Voice".

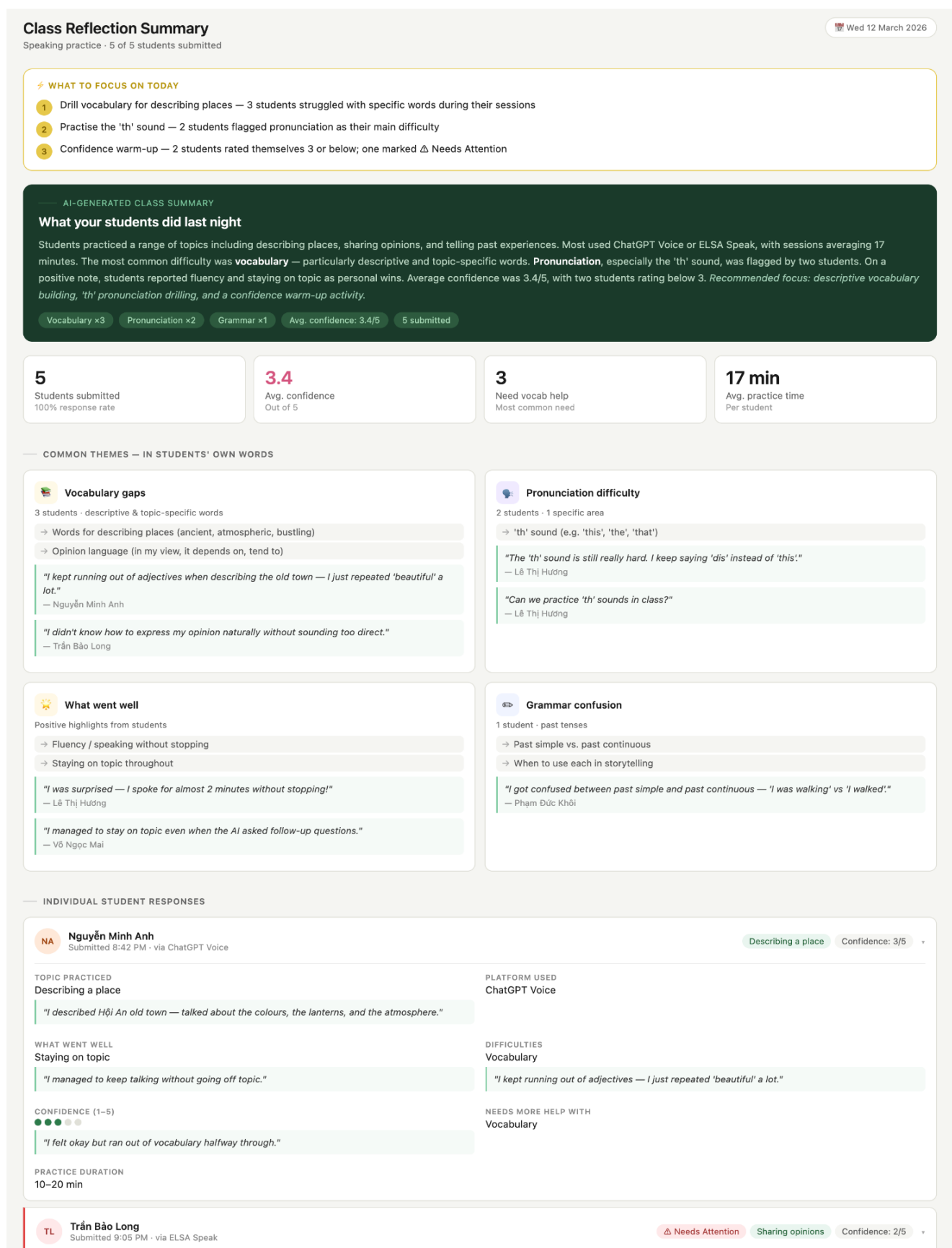
Panel 4: The chatbot asks, "Nice! About how long did you practice?". The student responds, "10-20 min".

Panel 5: The chatbot asks, "Good effort! What do you think went well today?". The student responds, "Fluency / speaking without stopping".

The interface includes a text input field at the bottom with the placeholder "Nhập câu trả lời..." and a "Skip" button. The chatbot also provides feedback and encouragement throughout the session.

Figure 2 represents the teacher dashboard, which allows teachers to view individual student responses and class-level summaries. It also shows the common themes that most students have mentioned, students' average confidence scores, and suggested instructional priorities for the next lesson.

Figure 2. Teacher dashboard displaying AI-generated reflection summary



3.5 Data collection and analysis

Data were collected through semi-structured interviews via Zoom, each of which lasted approximately 50 to 60 minutes. The interview guide consisted of 18 questions, which can be categorized into three thematic areas corresponding to three research questions. The first thematic area focused on teachers' current visibility into their students' independent speaking practice and how limited visibility affects their lesson planning and classroom activities, for example *"When students are practicing speaking on their own outside class, how do you currently find out what they have been working on?"* The second thematic area was about what kinds of learner reflection data the teachers would find useful and actionable, for example, *"If you could know one thing about what a student practiced in an AI speaking session before your next class, what would it be?"* The third thematic area examined teachers' impressions of the reflectEFL prototype concerning its utility and feasibility. The interviews were all conducted in Vietnamese in accordance with the participants' preferences and recorded with consent. All recordings were transcribed verbatim.

The interviews were analyzed using thematic analysis according to the framework of six phases proposed by Braun & Clarke (2006). The first phase involved reading the interview transcripts repeatedly until the researcher became familiar with the data. This was followed by coding of meaningful units related to the three research questions, such as useful types of reflection data, concerns about students' honesty, and evaluations of the prototype. Initial codes were then categorized into broader themes which were reviewed against the dataset. An inductive analysis process was employed to derive themes from the data but the three research questions served as the overarching framework for the analysis. As the interviews were conducted in Vietnamese, the analysis was also carried out in Vietnamese to preserve the original meaning of the data. Relevant excerpts were further translated into English by the researcher for inclusion in the findings.

Several steps were taken to improve the quality and the reliability of the qualitative analysis. Firstly, all interviews used the same questions to collect consistent data while still allowing follow-up questions when participants wanted to add more insight into a particular topic. Secondly, each interview was transcribed verbatim before the coding process began, ensuring that the coding reflected the teachers' accounts as closely as possible. Thirdly, coding and theme generation were conducted in Vietnamese to avoid loss of meaning in translation. Finally, selected excerpts were translated into English only after the themes had been

developed, and the translations were double-checked against the original language to preserve the meaning as close as possible.

3.6 Ethical considerations

All ethical guidelines were followed throughout the research. Before the interviews, participants were informed of the study's purpose, the nature of their participation, and how the data would be used. Participation was voluntary, and participants provided consent prior to interviews. They were informed of their right to withdraw from the study at any time without adverse consequences. Names were pseudonymized, and identifying information was removed from transcribed data. The research was carried out based on the principles of ethics set by the University of Tartu.

4 Findings

4.1 RQ1: Challenges teachers face when students engage in unmonitored AI-assisted speaking practice

All five participants reported limited or no visibility into what their students practiced outside of class. What they knew was that their students had access to AI tools, but they did not have insight into how students used AI, whether it was for speaking or writing, what they practiced, and how effective the practice sessions were.

T1, a university lecturer teaching classes of 25 to 40 students, noticed that her students rarely used AI for independent practice, and when they did, they tended to use it for written tasks rather than speaking practice. Her school provided students with an AI chatbot for practicing English, but the number of students using that chatbot was also low:

“I wish they would use AI more. Most students only use it when doing homework.

Occasionally, one or two students use ChatGPT for conversations, but usually not with the intention of practicing what they learned in class. They tend to use it casually, as part of a trend among young people, rather than for English practice. Overall, most of them do not use it at all.”

T2, who teaches classes of up to six students preparing for Cambridge English Qualifications, stated that she had no clear picture of what students practiced at home:

“I’m not sure. However, I think they definitely use AI because I’ve seen some students apply it to their writing homework. I’m often instructed to flag AI-generated content in

students' writing, so that's how I know. I think students know how to use AI to support their studies at home. Beyond writing, I'm not sure what else they do."

T3 gave an example in her teaching practice that illustrated the visibility gap most clearly. One of her adult students had been practicing independently with Duolingo and other tools but came to class uncertain whether his approach was effective: "He recorded the sessions he did with Duolingo and showed them to me so I could check whether his learning strategies were appropriate or not. He used Duolingo every day, but he was unsure whether it was actually effective." This example is important because the student was not asking whether Duolingo was good in general. He was asking whether his personal way of using it was effective. This shows that the teacher's role is not removed by AI practice; rather, students may need teachers even more to help them interpret the value of that practice. Moreover, T3 explained that when students asked such questions unexpectedly during class, she often had to provide rather general advice because she came to class unprepared for those discussions: "I also had to think about how to respond. Normally, I would put it off until the next time and just say something quite vague, such as that you need to encounter a word seven or eight times in order to remember it."

T5, whose classes focused on grammar, confirmed that students rarely disclosed their independent learning activities: "Almost never. Only one adult student, before studying with me, had used Duolingo and an app called Coach. He recommended it highly."

T4, however, had an alternative viewpoint. Teaching one-on-one and in groups of up to four students, she already had sufficient visibility through direct observation and found the gap less relevant to her teaching context. As she mentioned in the interview, she could identify each student's specific weaknesses through in-class interaction and did not feel that additional data from independent practice would alter her approach significantly.

Apart from the missing data, teachers also pointed out another issue they faced in their interviews. Even when students did practice independently, teachers found no way to connect those practice sessions to classroom instruction. As T3 observed: "Students are using AI quite indiscriminately. And students don't even know whether they're using it correctly and effectively. And then they come to the class and ask me, and I sometimes don't know how to answer either."

4.2 RQ2: What types of learner reflection data do teachers consider most actionable

When the participants were asked about what kinds of information about students' independent speaking practice they found useful for lesson planning, all five identified broadly similar categories. Yet, each one had a slightly different focus related to their teaching context.

The most consistently mentioned category was students' difficulties with vocabulary and pronunciation. T2 said: "What vocabulary or pronunciation do they get wrong most often? Which sounds do they find hardest to pronounce? Because in English, Vietnamese people often confuse the sound /l/ and /n/, as some students still read /l/ as /n/ and /n/ as /l/." She also pointed out that grammar mistakes, fluency, and idea development were of comparable importance: "The next thing is grammar errors when speaking in sentences. The fluency when developing ideas. And whether they have enough ideas, to see if their vocabulary range is sufficient."

The two categories that were also widely valued were topic practiced and duration. T1 emphasized that the understanding of the practice session context was considered as important as the results:

"I might ask students questions such as: What topic are you talking about? Do you practice for a long time? How long does the conversation last? Are you prompting the AI for language practice, or is it simply a casual conversation without feedback on language use? Then I would follow up with questions like: Did you gain any new vocabulary, grammar, expressions, or pronunciation from that conversation?"

T2 associated duration with dedication to practicing speaking English: "In order to do well in speaking, you spend at least two or three hours a day practicing so that you get better at it." Meanwhile, T5 viewed duration as a way to assess the consistency of daily practice and to monitor the time students spent practicing.

The last two categories represented, which were confidence level and what students needed support, were recognized as valid measures for the majority of participants, but less frequently brought up spontaneously compared to the other three. T3 supported all five categories while adding another one not listed before:

"Probably the type of activity or some kind of summary activity. Like reflecting on what they have learned. I've noticed that students often finish studying, but if there is no wrap-up at the end, they forget everything by the next lesson. So after completing an activity or exercise, there should also be a wrap-up to help them reflect on how much they actually remember from it."

As for T5, the emphasis was on long-term development and the process rather than a single training period:

“I would prefer that the students use the chatbot after every practice so that I could trace the progress every day. It means I would be able to understand what they did well, where they were struggling, how they progressed day to day, and what problems they faced in different areas, so I could assist them more effectively. Instead of coming to me twice a week, they could just save the time. After all, during our class sessions, I will be able to use the time efficiently.”

T4, as seen from her responses to the first research question, found most of the categories less relevant due to her small group classes and face-to-face observations, and expressed her desire to have bullet points if a summary of any kind was included.

4.3 RQ3: How teachers evaluate the usefulness and feasibility of the reflectEFL prototype

The general reaction from participants was positive, with teachers expressing that the idea addressed an important gap in their teaching practice. T4 was the exception, finding the tool less applicable for her teaching context.

When shown the teacher dashboard, T1 was enthusiastic about its usage, highlighting that her classes would benefit from the tool: “I mean, if there are 30 students in a class, I might only have the chance to speak individually with each student once or twice throughout the entire course. This tool would therefore be especially helpful in larger classes.” In addition, T1 confirmed that she would use the tool if it were available: “Yes, definitely. It’s not just 30 students; I teach three classes. In some semesters, when students submit recordings of their speaking exercises, I have to listen to hundreds of students every day. After a while, I become completely exhausted and lose concentration.”

T2 reacted similarly, saying she would apply this tool to help her plan her lessons: “I would use this tool to spend less time thinking about additional activities for upcoming lesson plans and focus more on the areas my students actually need, rather than deviating from the main lesson content.” T2 suggested two additional improvements to the prototype, in addition to those already listed above. First, she would like more specific instructional suggestions:

“What I mean is that the tool could provide even more detailed support. For example, if a student is struggling with the /θ/sound, it could suggest ideas for teachers, such as specific activities that might help. It could also recommend assigning personalized exercises for that student to practice the /θ/ sound at home before the next lesson.”

Secondly, she pointed out the problem of prioritization when multiple students had different needs:

“Let’s say there are three students who require support. However, the teacher is unable to attend to all three students simultaneously due to time constraints. In such cases, the use of AI can be useful in deciding which student should be provided support first. If the teacher cannot teach all three students simultaneously, they can teach one student this week and then move on to the other two students next week and so forth.”

T3’s first impression of the tool was positive, and she noted the tool’s novelty: The idea seems useful. It will help collect information. I’ve thought before that students use AI, but I didn’t know of any way to help. This is one of the tools that could solve that problem for teachers. It is quite new. I’ve never heard of anything like this before.” One of her recommendations was that the chatbot for students should ask which application the student uses, as this will help teachers better prepare and give appropriate advice.

T5 found the prototype practical because, in their view, it shifted some responsibility onto the student. Rather than leaving the teachers to discover the learning gaps on their own, the tool required students to actively report on their practice: “It helps share responsibility between both sides, and the content is realistic. I think it will be useful for both teachers and students.” Honesty in student responses was a common concern among participants. According to T2, the responses would be roughly half honest: “If students suspect that the chatbot is being used by the teacher to monitor their understanding, some may become dishonest out of fear.”

However, T1 and T3 disagreed, arguing that students who have truly done the practice would genuinely seek support from teachers, thereby responding honestly. T5 echoed this, putting emphasis on framing: “If students understand that filling it in is to help them get better support, they will fill it in. But if it just feels like a report, they might fill it in carelessly.”

T4’s feedback focused on format rather than concept, as she stated that she did not need the tool in her teaching context, given the small number of students in her classes and her individualized approach. She suggested using bullets rather than summaries to make reading easier.

5 Discussion

5.1 The visibility gap is real but context-dependent

One common pattern observed in the interviews was that most teachers lacked insight into what students practice with AI tools outside class. Although some participants suspected that their students used applications such as ChatGPT or Duolingo at home, they usually learned about it indirectly, for example, through writing assignments, occasional classroom conversations, or individual student questions. None of the participants described having any way to keep track of what they did or discuss these independent speaking activities before class.

This lack of visibility seemed especially noticeable in teaching contexts with larger classes. T1, for example, explained that with classes of 30 or more students, opportunities for individual interaction were already limited in class. Under those conditions, independent AI practice became almost completely invisible unless students brought it up themselves. A similar issue appeared in T3's account, where one adult learner recorded his Duolingo practice sessions and asked whether his learning approach was actually effective. What stood out in this example was not only the student's uncertainty, but also the teacher's difficulty responding immediately because she had not expected questions about AI-supported self-study during class time.

These examples align with Lai et al.'s (2015)'s observation that teachers usually adopt a passive stance toward their students' out-of-class technology use, often expecting students to come forward with the problems they face rather than proactively seeking to understand their independent learning. However, the interviews in this study suggested that the problem may be less about teachers intentionally choosing not to engage, and more about the absence of a practical mechanism to do so. In other words, teachers cannot easily respond to learning activities for which they have no information.

At the same time, the finding also shows that the visibility gap is not equally relevant across all teaching contexts. T4, who worked mainly in one-to-one and very small group tutoring contexts, did not consider the proposed tool particularly necessary. Because she came into direct contact with her learners during the lesson, she argued that she had sufficient information about their needs without gathering any additional information. T4's case limits the scope of the argument. In small tutoring classes, the teacher may already know enough about each learner through regular interaction. The visibility gap becomes more serious when the teacher has too many students to follow individually. For this reason, reflectEFL should not be treated as useful for every EFL classroom. It is most relevant in settings where teachers cannot easily follow what individual students do outside class.

5.2 Reflection data has clear diagnostic value for teachers

The findings from the second research question indicate that teachers shared broadly similar views regarding the kinds of learner reflection data they considered useful for instructional planning. Language difficulties, the practiced topic, and practice duration emerged as the most consistently valued categories among participants, with confidence level and perceived support needs also found useful when presented. The teachers were not simply asking for more data. They wanted information to help them prepare for the next lesson. This use of reflection data is broadly consistent with Menekse's (2020) reflection-informed learning and instruction model, where student reflections provide instructors with information about learners' difficulties and can support teaching adjustments.

It is important to note that the teachers did not pose their information requirements in general terms. T2 mentioned specific phonological errors common among Vietnamese learners, T1 wanted to know whether students used AI for language practice or only for casual conversation, and T3 wanted information about the types of activities students had completed. These responses show that teachers wanted information that could help them evaluate the quality and focus of students' practice, rather than simply whether practice had taken place. This finding can be linked to Winne and Hadwin's (1998) adaptation phase, in which learners reflect on previous studying and make adjustments for future learning. Guan et al. (2024) similarly found that educational chatbots have provided limited support for learners' reflection on prior studying and adaptation to future studying. In the present study, reflection data also had value for teachers. When students described their difficulties and support needs, teachers received information that could guide follow-up instruction. This is consistent with Menekse's (2020) reflection-informed learning and instruction model, where student reflections provide diagnostic information that instructors can use to adapt teaching, feedback, pacing, and support.

T5's framing of the tool as something that "shares responsibility between both sides" captures this dynamic well. Currently, independent AI speaking practice places the full burden of learning on the student, leaving the teacher's guidance out of the process entirely. Structured reflection, however, helps redistribute responsibilities – students provide information based on their learning experience, while teachers can use this information to make their instruction more responsive and relevant to students. Although such redistribution is limited in scope, it is crucial to re-establish the teacher's role in a learning process that has effectively excluded them.

5.3 Feasibility depends on student honesty and teacher workload

While participants reacted positively to the reflectEFL concept, two main concerns emerged in the findings regarding feasibility: whether students provide truthful feedback and whether the information collected through this form would add to, rather than reduce, the teacher's workload.

Regarding student honesty, the participants had different views. T2 estimated that about 50% of the students would give honest responses, reasoning that some might lie if they suspected the tool was a monitoring mechanism rather than a support one. T3 was more optimistic, believing that students who actually practiced at home would be motivated to respond truthfully because they needed the teacher's support. Similarly, T5 believed that framing the issue is crucial, saying, “If students understand that filling it in is to help them get better support, they will fill it in. But if it just feels like a report, they might fill it in carelessly.” These perspectives on student honesty suggest that this is not a fixed characteristic but rather a communication and design challenge. The chatbot would therefore need to be introduced as a supportive tool rather than as an assessment or monitoring mechanism. In this sense, features such as bilingual prompts, concrete questions, and optional skip buttons are not merely usability features; they may also help reduce pressure on students and encourage more meaningful reflection.

Regarding workload, the findings were generally positive. Listening to hundreds of student audio files from each class was clearly too much work for T1, and summarizing them into a single dashboard would help alleviate this burden. Similarly, T2 and T5 thought the tool could save them time, and T3 reported that, even with small classes, it was hard for her to monitor students' work on her own. The only voice to speak against the chatbot's potential was T4, who argued that the tool was solving a problem she did not have in her classroom.

An important implication of these findings relates to design. For the reflectEFL approach to work across multiple teaching contexts, it needs to be lightweight enough to be implemented with minimal additional effort on the part of teachers, while those in smaller or more intensive contexts can simply opt out. Several participants identified the lack of institutional requirements — such as platform integration or data-sharing agreements — as a practical advantage, suggesting that this feature should be maintained in future development of the tool.

5.4 The chatbot as a bridge, not a replacement

A consistent pattern in teachers' perspectives on the reflectEFL tool can be observed across all three research questions: they did not consider it a substitute for teacher judgment but

rather an instrument to inform their judgment. None of the participants suggested replacing the pre-class preparation or direct interaction with students with the chatbot. Instead, they envisioned themselves getting additional input from the summary to be more responsive, in other words, arriving at class already having insights into what students need.

A similar view appears in Guan et al.'s (2025) study, where AI integration in EFL education is presented as collaborative rather than replacement-based. In that model, AI mediates teacher-student interaction and supports learning while preserving teachers' pedagogical roles. This aligns with how teachers in the present study described reflectEFL. T1, T2, and T3 all described using the chatbot as a tool to support their teaching rather than replace their decision-making. The chatbot would provide useful information, but teachers still decide how to respond in class. For example, T2 suggested that the tool could be used to identify which students needed support most urgently, while T3 wanted to know which applications students use so they could give better guidance. These examples show that teachers viewed the tool as a means of supporting and strengthening their professional judgment rather than replacing it. This also connects with Lai et al.'s (2015) study of autonomous technology-based language learning outside the classroom. Their students expected teachers to play a greater role by recommending technological resources and sharing strategies for using them effectively. In the present study, reflectEFL addresses a similar issue indirectly: it provides teachers with more information about students' independent practice, making it easier for them to offer guidance rather than remain disconnected from what students do outside class.

T4's case is important because her rejection of the tool did not signify that she rejected the teacher's role. In fact, her approach, which relied on close, direct observation of students, placed even greater emphasis on teacher involvement in students' learning. It should be noted that she had refused the chatbot but had not refused the idea that teachers should stay informed about students' learning. This reveals that many EFL teachers share the belief that teachers need a way to stay connected to students' learning progress, even when they do not need a specific tool like the reflectEFL chatbot to do so.

Taken together, the findings suggest a modest role for AI. The chatbot can collect and organize information that teachers usually do not see, but the teacher still decides what to do with that information. It should be noted that the approach proposed in this study is much more restrained than many other visions of AI in education currently under discussion. However, it might offer a more practical and sustainable approach, especially in diverse teaching contexts with limited resources.

6 Conclusion

6.1 Summary of findings

The purpose of this study was to examine the visibility gap between classroom instruction and students' independent speaking practice outside the classroom. It is also set out to evaluate whether the chatbot-mediated learner reflection approach could bridge this gap. Three research questions guided the inquiry: what challenges EFL teachers face when students practice with AI-powered tools independently, what kinds of reflection data teachers find most useful and actionable, and how they evaluate the reflectEFL tool prototype as a pre-class teaching aid.

First, the findings show that the visibility gap appears to be a genuine problem in EFL teaching contexts, especially ones with larger classes. Teachers have no systematic way of getting information on how students practiced outside class, which could limit their ability to connect independent practice with classroom instruction. Second, teachers consistently identified language difficulties, topic practiced, and duration of practice as the most actionable items when asked about what learner data would be most useful. Students' confidence level and what support they need from their teachers were also valued. Third, responses to the prototype were largely positive, with most of the participants expressing that the concept would bridge the gap in their teaching practice. Most of them also expressed their concerns about student honesty, which could affect the feasibility of the tool, and highlighted the importance of framing the tool as supportive rather than evaluative. There was a boundary case with T4, who found the tool unnecessary in her small classroom teaching context, suggesting that the approach is more relevant for larger class settings.

6.2 Contributions of the study

This study makes three main contributions to the field. First, it contributes to the limited research on chatbot-assisted learner reflection as a way of enhancing teacher visibility into students' independent EFL practice. Although chatbot-supported language learning has received growing scholarly attention, as shown by Huang et al.'s (2022) systematic review of empirical studies in this area, the connection between AI-supported language learning and self-regulated learning remains less developed. Dizon et al. (2025) note that "AI tools have received scant attention when it comes to L2 learning and SRL" and describe ChatGPT for self-regulated language learning as "an under-explored topic in AI-assisted language learning

literature.” Therefore, the specific use of learner reflection within chatbot-assisted EFL learning remains relatively understudied.

Secondly, the study reconceptualizes the visibility gap as a structural problem rather than a motivational one. In this study, teachers are passive not because they are uninterested in how students practice independently, but because they lack a communication channel to access that information. This re-evaluation suggests the solution lies in creating communication channels, rather than changing teachers’ attitudes.

Finally, the study provides a teacher-validated set of design criteria for the reflectEFL prototype. Information about the difficulties students encounter, the topics they discuss with AI-assisted platforms, the time required to complete practice sessions, their confidence levels, and the areas where they need assistance can be used to develop chatbots. Teachers’ suggestions regarding AI-powered teaching recommendations, prioritizing student needs, and presenting the tool in a low-pressure way also provide practical guidance for future improvements.

6.3 Limitations

There are three limitations that are worth acknowledging. First, the prototype in this study was low-fidelity. It was rather a conceptual demonstration that was shown in the interview, not a fully functioning system that could be used in real teaching contexts. The participants’ responses reflect perceived usefulness rather than the tool’s effectiveness. It remains to be seen if the tool will perform as expected in real-world teaching contexts.

Second, the study was limited to the teacher perspective. Although it was explicitly clear in the research questions, the voices of students in their willingness to engage with the chatbot honestly and consistently still remain untested. Students are the main users of the reflection interface; therefore, their experience with the tool is important to assess its usefulness and viability.

Third, the proposed system relies on self-reported reflection data. As Winne (2020) argues, self-report data may not always accurately capture learners’ actual behaviors because they depend on how respondents interpret prompts, recall events, and express their experiences through language. This limitation is relevant to reflectEFL because the chatbot depends on students’ descriptions of their own AI-assisted speaking practice.

6.4 Implications for practice

The reflectEFL approach provides a low-cost mechanism for reconnecting students' independent practice with classroom instruction, particularly for teachers working in large classes where individual monitoring is limited. The independence of the tool from any platform and reliance on no institutional or AI data could make it valuable especially in low-resource environments typical of EFL instruction. However, teachers have to be careful about how they present the tool to the students, framing it as a means to support learning rather than a monitoring tool.

For educational technology designers, the findings suggest that teachers may not need dashboards with highly complex visualizations. Instead, they value information that is relevant, easy to understand, and immediately actionable within their limited preparation time. T2 suggested that the tool should highlight common patterns across the class before focusing on individual differences. This suggestion offers a practical design principle: address the shared need first, then the individual.

6.5 Directions for future research

Future research can follow from the limitations of this study. The most direct direction would be to test reflectEFL with actual student users to examine whether engagement and honest reflection can be sustained over time. A longitudinal evaluation would also be useful for examining whether teacher access to students' reflection data affects lesson planning, feedback, or students' speaking development. This direction is consistent with Guan et al. (2024), who noted that many educational chatbots have been evaluated only in small-scale or one-time studies and that the lack of longitudinal data limits understanding of the long-term benefits of SRL chatbots.

Another viable direction is developing a fully functioning prototype and testing it in real classroom conditions, moving beyond the conceptual demonstration in the current study. Finally, one important question left unanswered by this study is how teachers actually use reflection data once they receive it. Future research should also examine why teachers may or may not use reflection data in their actual lesson planning.

7 Author's declaration

I hereby declare that I have written this thesis independently and that all contributions of other authors and supporters have been referenced. The thesis has been written in accordance with the requirements for graduation theses of the Institute of Education of the University of Tartu and is in compliance with good academic practices.

Date: 25 May, 2026

Name: Tran Van Khanh Linh

Author's signature:

A handwritten signature in black ink, appearing to be 'Linh', with a stylized flourish at the end.

Supervisor's signature:

References

- Braun, V., & Clarke, V. (2006).** Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Carrera Núñez, A. A., Carrera Núñez, M. S., Zambrano Pachay, J. F., & Carrera Bosquez, A. M. (2025).** Using ChatGPT voice to improve speaking skills in English language learners. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 7143–7161. https://doi.org/10.37811/cl_rcm.v9i1.16390
- Creswell, J. W. (2014).** *Research design: qualitative, quantitative, and mixed methods approaches*. 4th ed. SAGE Publications.
- Ding, D., & Yusof, A. M. B. (2025).** Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, 12, 1223. <https://doi.org/10.1057/s41599-025-05550-z>
- Dizon, G., Gold, J., & Barnes, R. (2025).** ChatGPT for self-regulated language learning: University English as a foreign language students' practices and perceptions. *Digital Applied Linguistics*, 3, 102510. <https://doi.org/10.29140/dal.v3.102510>
- Do, H. (2026).** Boosting students' reflection on English learning in the Vietnamese context. *Frontiers in Education*, 11, 1738960. <https://doi.org/10.3389/feduc.2026.1738960>
- Guan, L., Lee, J. C.-K., Zhang, Y., & Gu, M. M. (2025).** Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study. *Computers and Education: Artificial Intelligence*, 8, Article 100384. <https://doi.org/10.1016/j.caeai.2025.100384>
- Guan, R., Raković, M., Chen, G., & Gašević, D. (2024).** How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, 30, 4493–4518. <https://doi.org/10.1007/s10639-024-12881-y>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022).** Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Lai, C., Yeung, Y. and Hu, J. (2015).** University Student and Teacher Perceptions of Teacher Roles in Promoting Autonomous Language Learning with Technology outside the Classroom. *Computer Assisted Language Learning*, 29, 703-723. <https://doi.org/10.1080/09588221.2015.1016441>

- Lee, H., & Mori, C. (2021).** Reflective practices and self-directed learning competencies in second language university classes. *Asia Pacific Journal of Education*, 41(1), 130–151. <https://doi.org/10.1080/02188791.2020.1772196>
- Li, Y., Wu, Y., & Chiu, T. K. F. (2025).** How teacher presence affects student engagement with a generative artificial intelligence chatbot in learning designed with first principles of instruction. *Journal of Research on Technology in Education*. <https://doi.org/10.1080/15391523.2025.2493942>
- Li, Y., Zhou, X., & Chiu, T. K. F. (2025).** Systematics review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory (SDT): what are the roles of teachers? *Interactive Learning Environments*, 33(3), 1850–1864. <https://doi.org/10.1080/10494820.2024.2400090>
- Li, Y., Zhou, X., Yin, H., & Chiu, T. K. F. (2025).** Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review. *Smart Learning Environments*, 12, 24. <https://doi.org/10.1186/s40561-025-00379-0>
- Lösche, S., Rambo, C. A., & Ferreira, J. de L. (2023).** Exploratory research in the qualitative approach in education. *Revista Ibero-Americana de Estudos em Educação*, 18, e023141. <https://doi.org/10.21723/riaee.v18i00.17958>
- Lyster, R., & Ranta, L. (1997).** Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19(1), 37–66. <https://doi.org/10.1017/S0272263197001034>
- Menekse, M. (2020).** The reflection-informed learning and instruction to improve students' academic success in undergraduate classrooms. *The Journal of Experimental Education*, 88(2), 183–199. <https://doi.org/10.1080/00220973.2019.1620159>
- Nguyen, H. H. (2025).** AI for English speaking practice: A study of effectiveness and engagement among Vietnamese university learners. *Australasian Journal of Educational Technology*, 41(3), 45–61. <https://doi.org/10.14742/ajet.10050>
- Nguyen, Q. N., Le, H. V., & Nguyen, T. T. T. (2025).** The impact of AI-supported speaking practice on EFL learners' confidence in Vietnam. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(5), 384–393. <https://doi.org/10.62225/2583049X.2025.5.5.4924>
- Winne, P. H. (2020).** A proposed remedy for grievances about self-report methodologies. *Frontline Learning Research*, 8(3), 164–173. <https://doi.org/10.14786/flr.v8i3.625>

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum.

Appendix A. Interview guide

The semi-structured interview consisted of 18 questions divided into three thematic areas corresponding to the three research questions.

Section A: Visibility in teaching/ Black box problem

1. Can you tell me a little about your teaching context, i.e., how long you've been teaching EFL and what level of students you work with?
2. Are you aware of any students who are using AI speaking tools like ChatGPT Voice, ELSA Speak, and so on to practice speaking on their own?
3. When students are practicing speaking on their own outside class, how do you currently find out what they have been working on?
4. How does that, or the lack of it, affect how you plan your speaking lessons?
5. Have you ever felt that a student's in-class performance did not match what you expected, and later realized it might be connected to what they were doing outside class?

Section B: What information would be useful

6. If you could know one thing about what a student practiced in an AI speaking session before your next class, what would it be?
7. I will read you a list of possible types of information a student reflection might include. For each one, can you tell me how useful it would be for your lesson planning?
 - The topic or task the student practiced
 - Vocabulary or grammar the student found difficult
 - How confident the student felt during the session
 - What the student thinks they need more help with
 - How long the student practiced
8. Are there any types of information not on that list that you think would be more useful?
9. How much information would be too much? Is there a point where a reflection summary becomes overwhelming rather than helpful?
10. How much time realistically do you have before a lesson to review student reflection summaries?
11. What concerns, if any, would you have about receiving this kind of information from students?/ Do you think students would engage honestly with a reflection prompt after an AI speaking session, or might they write what they think the teacher wants to hear?

Section C: Prototype evaluation

12. What is your overall first impression of this tool?
13. Looking at the summary the teacher receives, does it include the types of information you said would be useful to you earlier in our conversation?
14. Is there anything important missing from the summary that you would want to see?
15. Looking at the questions the student is asked, do you think your students would be willing to answer them? Would the questions make sense to them?
16. Can you imagine using something like this before a speaking lesson? What would need to be true for you to actually use it?
17. What would make this tool difficult to use in your current teaching context?
18. If you could change one thing about this tool, what would it be?

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Tran Van Khanh Linh ,
(*author's name*)

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

BRIDGING THE BLACK BOX: A QUALITATIVE EXPLORATORY STUDY ,
ON ENHANCING TEACHER VISIBILITY INTO AI-ASSISTED EFL
SPEAKING PRACTICE THROUGH CHATBOT-MEDIATED LEARNER
REFLECTIONS

(*title of thesis*)

supervised by Professor Emanuele Bardone, PhD. ;
(*supervisor's name*)

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;
3. am aware of the fact that the author retains the rights specified in points 1 and 2;
4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.



Signature

25/05/2026

Date