English Teaching, Vol. 79, No. 3, Fall 2024, pp. 31-59
DOI: https://doi.org/10.15858/engtea.79.3.202409.31
http://journal.kate.or.kr

Enhancing L2 Listening Proficiency and Metacognitive Awareness Through Self-Annotation in High School EFL

Hyeyoung Jung*

Jung, Hyeyoung. (2024). Enhancing L2 listening proficiency and metacognitive awareness through self-annotation in high school EFL. *English Teaching*, *79*(3), 31-59.
Metacognition in L2 listening has garnered attention, but there is limited research on concrete methods to practice it in high school EFL contexts. This study aims to investigate the effectiveness of self-annotation interventions in a high school EFL listening classroom. Through the analysis of pre- and post-listening tests, students' annotations over 10 sessions, and a post-questionnaire, this study reveals that self-annotation interventions improve students' comprehension, confidence, and motivation. Both higher- and lower-level students demonstrated improved L2 listening test scores compared to the control group, with the former generating more analytical self-annotations and achieving more statistically significant progress. The thematic analysis of students' self-annotations identified recurring patterns that can inform effective listening instruction, while also enabling students to recognize and correct their mistakes, as well as receive cognitive and affective support. This study emphasizes the importance and feasibility of incorporating self-annotation in high school EFL listening classrooms, making students' reflection processes observable and actionable, and providing recommendations for effective metacognitive interventions.

Key words: self-annotation, metacognition, listening proficiency, listening instruction, EFL, high school

*Author: Hyeyoung Jung, Teacher, Gyeonggido Office of Education; 18 Jowon-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16279, Korea; Email: hyjung@hawaii.edu

Received 30 June 2024; Reviewed 17 July 2024; Accepted 10 September 2024

1. INTRODUCTION

> "Teachers who understand the ways in which their students think and learn are better prepared to motivate their students and encourage their development in appropriate directions." (Paris & Winograd, 1990, p. 26)

Listening comprehension in a target language poses complex challenges for many foreign or second language (L2) learners, making targeted instruction essential. Despite its importance, research on effective listening instruction remains limited due to the inherently unobservable nature of listening processes (Vandergrift & Goh, 2012). To document and make visible students' listening processes, many studies have utilized writing interventions in listening classrooms, such as listening diaries (Flowerdew & Miller, 1992; Goh, 1997; Matsumoto, 1996; Milliner & Dimoski, 2024), listening logs (Cotterall, 2000; Gilliland, 2015; Kemp, 2010; Lee & Cha, 2020), online comments (Chang & Chang, 2014), self-report questionnaires (Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006), and self-annotation (Yang & Kang, 2020). Commonalities across these studies indicate that retrospective writings following L2 listening practices positively influence learners' *metacognition*, which refers to "cognition about cognitive phenomena" or "thinking about thinking" (Flavell, 1979, p. 906). Metacognitive approaches encourage learners to monitor, evaluate, and improve their listening processes by actively engaging in their own listening development and understanding the mental and emotional processes involved in their learning (Vandergrift & Goh, 2012). This, in turn, enhances their listening proficiency, self-efficacy, strategy use, and motivation (Cross, 2009; Goh & Taib, 2006; Graham & Macaro, 2008; Vandergrift & Tafaghodtari, 2010).

In an EFL context, studies have utilized writing reflections to enhance students' metacognitive awareness and active engagement in listening practices. Chang and Chang (2014), for example, conducted an online videotext self-dictation-generation learning activity with undergraduate students in Taiwan. In this study, students were first introduced to bottom-up and top-down listening strategies and were tasked with creating cloze tests for YouTube video clips. The students documented the strategies they employed and the comprehension difficulties they encountered in the comments section. This activity reportedly enhanced students' metacognitive listening awareness and improved their listening test scores. Similarly, using YouTube as their primary source material, Lee and Cha (2020) implemented a structured approach involving listening logs among university students in Korea. Students wrote weekly logs that included summaries, content responses, reflections on their listening abilities, and the strategies they used. The study found significant improvements in listening proficiency and metacognitive awareness, particularly among high-proficiency students who utilized a variety of strategies simultaneously.

While these empirical studies on EFL listening metacognition through reflective writing provide invaluable options and insights for EFL listening classrooms, much of the current research targets university EFL students, leaving high school contexts largely unexplored. Many high school EFL classrooms in Korea face challenges distinct from university settings, including limited access to authentic listening materials, a wide range of student proficiency levels within a single class, time constraints due to packed curricula, and institutional pressures focused on standardized testing (Song & Kim, 2017). In the context of this study, high school EFL teachers primarily focus on test preparation, particularly for the College Scholastic Ability Test (CSAT), known as *Suneung*, with minimal use of authentic listening materials. Consequently, listening skills often receive limited instructional attention, leaving students to tackle listening tasks on their own or focus solely on solving test questions.

To address the research gap and cater to the specific needs of this context, this study is grounded in the theoretical framework of metacognitive approaches to L2 listening (Flavell, 1979; Goh, 2002; Vandergrift & Goh, 2012) and builds upon previous L2 reflective listening studies utilizing retrospective writings (Chang & Chang, 2014; Lee & Cha, 2020; Yang & Kang, 2020). The study incorporates self-annotation techniques in a high school EFL classroom as a simplified means of documenting and reflecting on students' listening processes. Following Yang and Kang (2020), self-annotation in this study refers to learners making personal notes or markings on listening materials at the end of a listening practice session to analyze their errors through planning, monitoring, and evaluation. Students utilize this space to analyze their frequent errors, express their feelings, write down resolutions, or pose questions for which they seek teacher feedback. In turn, the author (as a teacher in this classroom) provides feedback, including brief comments, questions, highlighting important parts, or offering praise stamps with phrases such as 'Excellent' or 'Good job' to acknowledge students' efforts. Understanding the extent to which self-annotation affects students' listening competence and how it unfolds in the classroom is crucial for teachers seeking effective interventions in test-oriented classrooms and similar contexts. By employing an experimental and control group design[1], with each group divided into higher- and lower-level students, the study aims to evaluate the feasibility and effectiveness of self-annotation techniques in this under-researched yet influential educational context for many EFL learners. The study is guided by the following research questions:

1) What patterns are observed in self-annotations across different proficiency levels?

[1] A control group was included to more accurately measure the effect of the intervention by isolating the impact of the self-annotation intervention, as differences between pre- and post-test scores could result from test difficulty, familiarity with the test format, or other external variables.

2) How does the self-annotation activity impact high school EFL students' listening proficiency across different proficiency levels and between experimental and control groups?
3) How do students perceive the self-annotation activity in relation to their learning?

2. LITERATURE REVIEW

2.1. Metacognition in L2 Classrooms

Metacognition in language learning refers to an individual's awareness and regulation of cognitive and affective processes involved in understanding the target language (e.g., Goh, 1997). Unlike metalinguistic knowledge which helps the understanding of the structure and function of language itself, metacognition involves awareness and control over one's own cognitive processes. Flavell (1979), in his pioneering work, described metacognition as a cornerstone in cognitive development, comprising two primary dimensions: metacognitive knowledge and metacognitive experiences. Metacognitive knowledge entails understanding oneself as a learner (person), awareness of learning tasks (task), and knowledge of learning strategies (strategy). Metacognitive experiences encompass "any conscious cognitive or affective experience" such as feelings of familiarity, confidence, satisfaction, solution correctness, and effort expenditure (Flavell, 1979, p. 906). These dimensions collectively foster metacognitive strategies such as self-awareness and self-monitoring, which are essential for effective independent learning (Papaleontiou-Louca, 2000).

In the context of L2 listening, Goh (2002) identified three key metacognitive strategies: planning, monitoring, and evaluating, which guide effective listening instruction (Goh, 2010; Schraw, Crippen & Hartley, 2006; Vandergrift, 1997). Planning involves setting goals and selecting strategies, monitoring includes ongoing assessment of comprehension and focus, and evaluating entails reflecting on strategy effectiveness and comprehension outcomes. These strategies are crucial for enhancing listening proficiency, helping learners adapt and improve their listening skills (Goh, 2010; Vandergrift, 1998, 2003).

Empirical studies demonstrate that proficient listeners employ a range of metacognitive strategies more effectively than less skilled listeners (Goh, 2002; Lee & Cha, 2020; O'Malley & Chamot, 1990; Vandergrift, 1998, 2003). For instance, Vandergrift (2003) found that among seventh-grade students learning French in a Canadian junior high school, more skilled listeners used a range of metacognitive strategies, including planning, monitoring, selective attention, and self-evaluation, when engaging with authentic texts. This led them to employ cognitive strategies effectively, such as relating new information to prior knowledge and making context-based inferences. This combination of top-down and

bottom-up processing facilitated the creation of a coherent mental representation of the text and improved their ability to navigate comprehension challenges. These findings underscore the importance of metacognitive practices in L2 listening and their role in enhancing learners' capacity to manage complex listening tasks.

However, despite the evident benefits of integrating metacognitive interventions in L2 listening instruction (Al-Khresheh & Alruwaili, 2024; Goh, 2002; Rahimi & Katal, 2012), a noticeable gap persists in studies addressing its implementation within test-oriented secondary school EFL classroom environments, characterized by diverse proficiency levels and time constraints. Research that casts doubts on the effectiveness of metacognitive interventions (Renandya, 2012; Renandya & Farrell, 2011) and presents conflicting outcomes (Milliner & Dimoski, 2024) also highlights the necessity for further empirical evidence in various EFL contexts. In addition, the developmental trajectory of metacognitive skills during adolescence also necessitates tailored instructional approaches (Flavell, Miller, & Miller, 1993).

Addressing these issues, this study examines the impact of a self-annotation intervention, implemented alongside traditional listening instruction that emphasizes vocabulary and grammar, on high school EFL students' listening proficiency. This approach is particularly pertinent for understanding how increased awareness of language processing can influence both listening skills and attitudes in test-oriented language classrooms.

2.2. Self-annotation in L2 Listening Classrooms

Self-annotation, defined as the process of adding notes or comments to a text, has been widely researched in the field of L2 reading and writing education as a valuable tool for enhancing language comprehension, vocabulary acquisition, and metacognitive awareness (Charles, 1990; Reid, 1994; Storch & Tapper, 1996). Similarly, in listening, students' self-annotations, as part of listening logs or reflection papers, seem to serve as an excellent way to make the learning process observable and thus actionable for both students and teachers (Chang & Chang, 2014; Goh & Taib, 2006; Lee & Cha, 2020; Vandergrift, 2003).

While the aforementioned studies used self-annotation as part of larger listening activities, Yang and Kang (2020) focused on the use of self-annotation by employing a self-annotated transcription (SAT) task at a Korean university. In this task, students transcribed English sentences from TOEFL listening tests, identified difficulties, and provided self-annotations to express their challenges, which were complemented by teacher feedback. The study categorized L2 learners' errors into six types: segmentation, substitution, omission, spelling, insertion, and blank. Analysis of self-annotations showed that students actively engaged in identifying difficulties and reflecting on their listening processes, using both bottom-up and top-down information. A post-task questionnaire with students and an in-depth interview

with the teacher reported positive attitudes towards the self-annotation task, noting its effectiveness in helping students recognize weaknesses, understand error patterns, and develop metacognitive strategies for improvement. Moreover, the self-annotation task fostered a more supportive classroom environment, promoting active problem-solving among students. The study concluded that the SAT is viable and beneficial in L2 listening classes, enhancing student learning and teacher feedback.

Building on Yang and Kang's (2020) study, this research made adaptations to implement self-annotation in a public high school classroom in Korea: 1) using cloze tests instead of transcribing entire sentences, 2) substituting TOEFL listening materials with CSAT test-preparation items, and 3) allocating 15 minutes of annotation time within class sessions instead of assigning it as homework. Since transcribing entire transcripts within a 50-minute class time is daunting for many high school students, cloze tests can be a manageable and time-efficient substitute for practicing self-annotation in this context. Although the CSAT does not include cloze tests, the inferencing skills required for fill-in-the-blank items—such as predicting missing words based on context, understanding implied meanings, and recognizing (key) words—are similar to those needed for the CSAT listening section, which makes cloze tests a typical routine in listening lessons at this school. In addition, integrating authentic materials or non-CSAT relevant content into listening resources can be challenging in this high school EFL classroom due to the pressing need for CSAT test preparation. To address these challenges, this study incorporates self-annotation activities using EBS CSAT mock listening tests[2]. Lastly, the extensive pressure on high school students to prepare for various subjects makes assigning extensive listening homework impractical.

The control group undertakes CSAT mock listening tests by responding to questions and completing cloze tests, while the experimental group follows the same routine but allocates approximately 15 minutes for self-annotation, as further elaborated in the subsequent section.

3. METHODS

3.1. Context

In such Korean high schools where students aim to go to college, preparing for the high-stakes CSAT influences English classrooms, as its scores play a significant role in college admissions. Hence, English teachers, especially those teaching senior high school students in those schools, often prioritize teaching English 'for the test,' focusing on form-focused

[2] In South Korea, the Educational Broadcasting System (EBS) CSAT mock test prep book is widely used in high schools because it mirrors the format, content, and difficulty of the actual CSAT exam, reflecting the latest trends and updates.

instruction and test-taking strategies rather than fostering genuine communicative skills (Song & Kim, 2017). Among the CSAT English sections, the Listening part constitutes one-third of the total score (37 points out of 100), making it a crucial aspect that should be taken seriously. However, limited attention has been given to English listening education compared to other aspects of English, such as grammar and reading comprehension which are often tested in in-school exams. Additionally, due to limited class time within a packed curriculum, teachers may struggle to allocate sufficient time to develop English listening skills. Thus, teaching listening is often left solely to teachers, who primarily focus on solving CSAT mock test questions, and to students for self-study without guidance. This prevailing educational context underscores the importance of investigating instructional interventions aimed at enhancing students' listening proficiency within the constraints of CSAT preparation and a rigid school curriculum.

3.2. Participants

Fifty-two senior-year students from two classrooms in a public high school in Gyeonggi, South Korea, participated in the study. The students included 24 females and 28 males. All participants had received at least ten years of English language instruction, starting from elementary school, and were preparing for the CSAT. None have lived in an English-speaking country for more than six months. Two classes were selected for the study based on their size (N = 26 for each class) and their similar level of listening proficiency based on a diagnostic listening test conducted at the beginning of the academic year. A diagnostic test was conducted using the Nationwide English Listening Proficiency Test for High School 3rd Grade[3], the results of which revealed no significant difference in mean scores between the Experimental Group (M = 11.88, SD = 5.35) and Control Group (M = 11.96, SD = 4.08), with a p-value of 0.95, ensuring a balanced baseline for the subsequent experimental interventions. Participants in each class were then divided into two groups based on whether their diagnostic score was above or below the overall mean score (M = 11.92): (1) Experimental Group, consisting of 12 students with higher-level listening proficiency and 14 students with lower-level listening proficiency, and (2) Control Group, comprising 11 students with higher-level listening proficiency and 15 students with lower-level listening proficiency. This division enables further analysis based on types of self-annotations and the extent to which their proficiency was affected, in relation to differing proficiency levels.

[3] It is a nationwide English listening proficiency test conducted twice a year by 15 provincial education offices nationwide, simultaneously broadcasted nationwide through EBS FM. It consists of a total of 20 questions and lasts approximately 25 minutes. The implementation varies depending on the school.

3.3. Self-Annotation Activity in L2 Listening Instruction

Before commencing the self-annotation task in the experimental group, the teacher allocated one hour to explain the rationale and technique of self-annotation, while also presenting samples of annotated notes from previous students and studies (Yang & Kang, 2020). The teacher began by explaining that self-annotation involves taking notes in a specific way that highlights important information, keywords, and personal thoughts. It was emphasized that this technique is useful not only for improving listening skills but also for preparing for exams by helping students concentrate better. The teacher also explained that during annotation, students could write down anything, including their strengths, areas for improvement, strategies for improvement, reasons for correct or incorrect answers, and their feelings. Students were informed that, at the end of class, their annotations would be collected and returned after review by the teacher, who would provide comments as necessary. Additionally, examples of self-annotations were shared, and an audio clip was played for the class to annotate together. Students then checked their annotations in pairs and shared insights, promoting peer learning and discussion of effective annotation techniques. This preliminary session aimed to help students become familiar with the self-annotation activity and understand its methodology and rationale.

To address the engagement difficulties experienced by some lower-level students during self-annotation activities, the teacher implemented three supportive measures from the start of the intervention. While these measures were primarily designed to assist lower-level students, they were made available to all students, as they can enhance engagement and support the effectiveness of the self-annotation intervention across different proficiency levels. These included 1) breaking tasks into manageable segments by allowing students to focus on one particularly challenging question rather than covering all questions; 2) sharing student samples for modeling during class: in this study, the teacher allocated 3 to 5 minutes during the next session to introduce and discuss some students' self-annotations from the previous class, highlighting common errors and exemplary annotations; and 3) permitting the use of L1 (Korean) when uncertain of spelling to annotate auditory information. English learners often struggle with English spelling proficiency while identifying key points amidst rapid speech or complex discourse structures (Goh, 2018). For instance, without the added concern of correct spelling, some students phonetically spelled out the words 'definitely' and 'absolutely' using Korean characters, omitting the /t/ sound and approximating the pronunciations as [dɛfənəli] and [æbsəluli]. This indicates the students' ability to focus on pronunciation even when unfamiliar with the spelling or the meaning, and it also provided the teacher with opportunities not only to explain those words by acknowledging students' efforts in 'noticing' English sounds but also to elucidate concepts such as the glottal stop [ʔ] using examples provided by the students (Schmidt, 1990). By incorporating instructions

based on students' experiences rather than solely relying on teacher-generated materials, the lesson's relevance and interest appeared to be enhanced.

Upon conclusion of each class, the completed handouts were collected, and all student self-annotations in the experimental group were categorized and quantified according to adapted subcategories for analysis. In the subsequent class, the annotated handouts were returned to the students for their reference, along with brief feedback from the teacher, which aimed to foster student engagement and provide guidance in the self-annotation process.

3.4. Data Collection Procedure

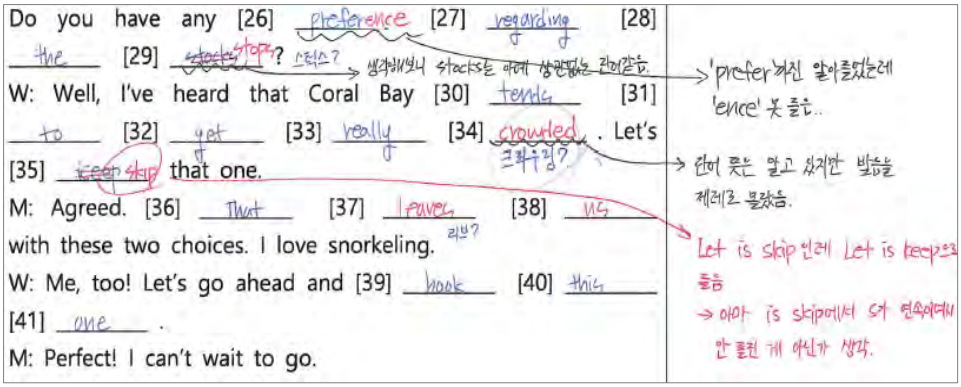
Listening instruction was integrated as part of an elective English course for senior-year students, consisting of 10 sessions over two months, with each session lasting 50 minutes. The primary objective of this instruction was to enhance students' performance in the CSAT listening component. Both the control and experimental classrooms utilized the same CSAT test-preparation materials published by EBS, widely regarded as the standard resource due to their alignment with the actual CSAT exam.

Typically, each class session followed a structured routine: it began with an explanation of CSAT listening question types and test-taking strategies (5-10 minutes), followed by a listening test (8-10 minutes), a cloze test (10-15 minutes), and vocabulary and grammar instruction led by a teacher (10-15 minutes). The control group concluded with whole-class repeated listening (10-15 minutes). In the experimental group, this final segment was replaced with self-annotation time.

During self-annotation, students freely wrote down anything that came to mind in the margins of their cloze test handouts (see Figure 1). In this session, students could listen individually through their mobile devices for tested items, document instances where they did not listen to the audio, specify the reasons for their oversight, reflect on how it impacted their answer choices, and ask questions. They were also encouraged to highlight instances in their annotations where they answered correctly or where their listening strategies were effective after previous reflection.

FIGURE 1

A Student's Self-annotation Following a Cloze Test Practice (04/25/2024)



To measure changes in students' listening test proficiency, both experimental and control group students took mock CSATs conducted in March (T1) as a pretest and in May (T2) as a posttest of the same year. These mock tests are designed to replicate the atmosphere and timetable of the actual CSAT, with the majority of high school students nationwide taking them on the same day and at the same time. These mock CSATs are administered by regional educational offices and share commonalities in adhering to the same level of difficulty, types of exam questions, and administration methods as the CSAT. To gain further insights into students' perceptions of the self-annotation activity, a qualitative questionnaire was administered at the end of the semester.

3.5. Data Analysis

This study employed a convergent mixed methods design to examine the effectiveness of self-annotation on high school EFL students' listening competence (Creswell & Creswell, 2018). This design involves collecting qualitative and quantitative data in parallel, analyzing them separately, and then merging the results. In this study, quantitative data will be used to test the differences in listening test scores between the experimental and control groups, as well as among different listening proficiency levels (i.e., higher and lower). The qualitative data will explore patterns of annotation and students' perceptions and experiences regarding self-annotation in high school EFL listening classrooms. The reason for collecting both quantitative and qualitative data is to triangulate findings and provide a comprehensive understanding of the impact of self-annotation on listening skills.

For Research Questions 1 and 3, thematic analysis was conducted following Braun and Clarke's (2006) theoretical framework. This method identifies, analyzes, and reports patterns

within the data, facilitating a comprehensive analysis that addresses specific research questions. Thematic analysis was chosen to gain sufficient information from a data-driven perspective through inductive coding.

For Research Question 1, a total of 905 self-annotations from 10 sessions involving 26 students in the experimental group were examined to identify common types of self-annotations and their frequency across two proficiency levels (i.e., higher-level vs. lower-level). In this study, self-annotations were only counted when students provided specific focus and further elaboration, engaging in the enhancement of their listening skills or incorporating the mental or emotional aspects of their learning processes (Vandergrift & Goh, 2012). The coding process was driven by the students' own analysis of their errors, prioritizing their perspectives over the researcher's interpretation. That is, the emphasis was placed on understanding how students themselves perceived and categorized their learning experiences, which offered authentic insights into their metacognitive processes. To address Research Question 2, a two-way repeated measures analysis of variance (ANOVA) was conducted to investigate within-subjects (pre-test vs. post-test) and between-subjects (group and proficiency) differences before and after the self-annotation intervention. For Research Question 3, a post-questionnaire was analyzed to understand students' perspectives regarding self-annotation interventions in test-oriented listening classrooms.

4. RESULTS

4.1. Types of Students' Self-annotations

In the experimental group, consisting of 26 students, a total of 905 self-annotations were recorded, averaging 34.80 self-annotations per student over 10 self-annotation sessions (*SD* = 14.33, Range = 8-63). Subsequently, thematic analysis was performed on the collated self-annotations to uncover recurring themes and patterns. As shown in Table 1, students' self-annotations are categorizable into five overarching types: phonetic features, lexical features, resolution, affect, and prior knowledge. Overall, higher-level students demonstrated a notably higher frequency and diversity of self-annotations (*M* = 44.08, *SD* = 13.16, Range = 18-63) across all categories than the lower-level group (*M* = 26.86, *SD* = 10.04, Range = 8-47), except for the negative emotions domain, suggesting a more robust capacity for reflection.

TABLE 1

Thematic Frequency Distribution of Self-annotations

Themes/ subthemes	Higher-level (%) (*n* = 12)	Lower-level (%) (*n* = 14)	Total (%) (*N* = 26)
Phonetic Features	***188(35.54)***	***133(35.37)***	***321(35.47)***
Pronunciation	116(61.70)	87(65.41)	203(22.43)
Speed	24(12.77)	25(18.80)	49(5.41)
Connected speech	18(9.57)	10(7.52)	28(3.09)
Unstressed syllable	19(10.11)	8(6.02)	27(2.98)
Contraction	11(5.85)	3(2.26)	14(1.55)
Lexical Features	***148(27.98)***	***104(27.66)***	***252(27.85)***
Meaning	83(56.08)	79(75.96)	162(17.90)
Spelling	59(39.86)	24(23.08)	83(9.17)
Collocation	6(4.05)	1(0.96)	7(0.77)
Resolution	***91(17.20)***	***77(20.48)***	***168(18.56)***
Self-study plan	70(76.92)	63(81.83)	133(14.70)
Test-taking strategy	21(23.08)	14(18.18)	35(3.87)
Affect	***85(16.07)***	***58(15.43)***	***143(15.91)***
Positive emotions	51(60.00)	12(20.69)	63(6.96)
Negative emotions	34(40.00)	46(79.31)	80(8.84)
Prior knowledge	***17(3.21)***	***4(1.06)***	***21(2.21)***
Lexical knowledge	11(64.71)	3(75.00)	14(1.55)
Grammatical knowledge	6(35.29)	1(25.00)	7(0.77)
Total	***529(58.45)***	***376(41.55)***	***905(100)***

The most prevalent theme was annotating on phonetic features in English that require caution. For instance, students frequently annotated instances where the pronunciation of English words deviated from their expected phonetic representations based on spelling. This phenomenon, characterized by a relatively low grapheme-phoneme correspondence, posed considerable challenges for students, often leading to misinterpretation or difficulty in comprehension (Bassetti, Cerni, & Masterson, 2022). In Excerpt 1, for instance, S1 highlighted a discrepancy between the spelling and pronunciation of the word '*determine*,' revealing how incorrect knowledge of word pronunciation can hinder accurate interpretation of spoken English. The student recognized that merely knowing a word's meaning is insufficient if its pronunciation is unclear, as this gap in pronunciation knowledge can impede comprehension during listening tasks.

Excerpt 1. Pronunciation (S1, 04/16/2024)

- Correct sentence - May I explore the building now to determine the best angles for photography?
- Student's answer- _____*, angles, photograpy*
- Student's self-annotation - Words that I don't pronounce accurately are not heard well. I got this question correct, but even on the second hearing, I could not fill in the blank. It's because I pronounced it as [*dɪtɜrmaɪn*], not as [*dɪtɜrmɪn*], so I

couldn't recognize it as that word.

In addition, some students annotated instances of connected speech (e.g., assimilation, consonant clusters, liaison) in rapid speech, highlighting the influence of fast-paced speech on their perception of spoken English. For example, in Excerpt 2, S2 noted instances where words were run together or linked through assimilation, which posed challenges for comprehension.

Excerpt 2. Connected speech (S2, 04/25/2024)
- Correct sentence: Let's skip that one.
- Student's answer: keep*, that, one
- Sudent's self-annotation: Since I don't know the pronunciation of words, listening feels really difficult, and I'm a bit scared that not hearing sounds like /s/ in phrases like "Let's skip" might cause mistakes in later exams.

S2 mistakenly interpreted the speaker as saying 'Let's keep that one' instead of 'Let's skip that one.' S2 attributed this mishearing to perceiving the /s/ at the end of 'let's' linking with the following word 'keep,' leading to a mispronunciation that sounded like 'let's keep' in connected speech. However, this error can be also traced to S2's unawareness of the allophones of /k/—where the [k^h] sound in 'keep' is pronounced differently from the [k] in 'skip'. This insensitivity to allophonic variations contributed to S2's misunderstanding of the listening test, causing them to think they should 'keep' the option rather than 'skip' it. This provided the teacher with a real-life teaching moment—arising naturally from a student's actual experiences or misunderstandings—to explain phonetic distinctions.

Though less frequent, some students made annotations regarding the occurrence of unstressed syllables and contractions. Excerpt 3 shows how the reduced emphasis on certain syllables can affect students' ability to follow spoken discourse accurately. S3 noted difficulties in discerning between '*won't*,' '*won*', and '*want*,' which led to an incorrect answer. Based on S3's annotation, it appears that S3 was uncertain about whether they heard 'won' or 'want.'

Excerpt 3. Unstressed syllables and contractions (S3, 04/09/2024)
- Correct sentence: This tour won't work for me.
- Student's answer: won*, work
- Sudent's self-annotation: I heard 'won' or 'want', so I thought she likes this tour. I often miss 't' at the end.

This uncertainty indicates a perceptual challenge with the /t/ sound, which is often reduced

or unexploded in natural speech, especially in contractions like ‘won’t.’ S3’s transcription as ‘won’ and the comment, ‘I often miss ‘t’ at the end,’ highlight a recurring issue with perceiving this sound, likely due to unfamiliarity with how /t/ sounds may be pronounced less distinctly in English. Moreover, S3’s confusion between ‘won’ and ‘want’ in the annotation suggests a gap in understanding how contractions function and sound in English. Therefore, while the immediate issue may be described as a failure to perceive an unexploded /t/, S3’s annotation shows it is also intertwined with the broader context of contraction recognition. This underscores the importance of teaching both the phonetic subtleties of unstressed syllables and the specific characteristics of contractions, ensuring that students are aware of how these linguistic features manifest in natural speech. By documenting these phonetic features, students demonstrated their awareness of the nuances of English pronunciation (e.g., intonation, stress patterns, vowel and consonant sounds) and their efforts to navigate these challenges in their listening comprehension endeavors. For the teacher, these students’ annotations offered valuable insights into students’ individual learning needs and provided student-produced examples for listening instruction, which in turn illuminate the multifaceted nature of pronunciation challenges students experience and tailor instructional strategies accordingly.

The second most common hindrance is vocabulary recognition, particularly when encountering unfamiliar words or phrases. When faced with unknown vocabulary, students often experience confusion, relying solely on contextual clues to grasp the essence of the listening material. This reliance on partial understanding can sometimes hinder overall comprehension, as illustrated by the experience of S4.

Excerpt 4. Collocation (S4, 04/23/2024)

- Correct sentence - (1) She read an article on eye health risks.
  (2) A screen filter that blocks blue light emission from the monitor.
- Student’s answer- (1) eye, help*, risks
  (2) blue, right*, mission*
- Student’s self-annotation - I knew every word in these sentences, but I was confused because I had never seen them used like that before. So I was thinking, ‘What is ‘eye help risk’ and ‘blue right mission’?’ When I listened again, I realized that these were words I had heard a lot in Korean, but I didn’t know their English expression, and it puzzled me. I need to study everyday English expressions more!

Despite understanding the meanings of individual words, S4, like many others in this class, encountered difficulty deciphering phrases such as ‘eye health risks’ and ‘blue light

emission', highlighting the challenges associated with unfamiliar collocations[4]. The struggle with collocation may stem from a lack of exposure to spoken English in authentic contexts, as well as a limited awareness of the significance of lexical chunks in L2 listening comprehension. This study corroborates the importance of teaching noun phrases, or "lexical bundles," to help learners of English in the areas of "vocabulary development and sentence formation in an acceptable way" for better grasping the gist of spoken language (Dahunsi & Ewata, 2022, p.16). Indeed, since S4 knew each word separately, S4 would understand their meaning with ease if they were written. In spoken language, however, teaching the collocation of noun phrases can aid students in easily grasping the intended meaning, thereby facilitating smoother comprehension and bridging the gap between what they actually hear and the correct interpretation to construct "a coherent mental representation of the text" (Vandergrift, 2003, p. 488). This exploration through student self-annotation not only seems to broaden students' understanding of collocation usage but also offers teachers insights into effective vocabulary teaching strategies tailored to this specific context.

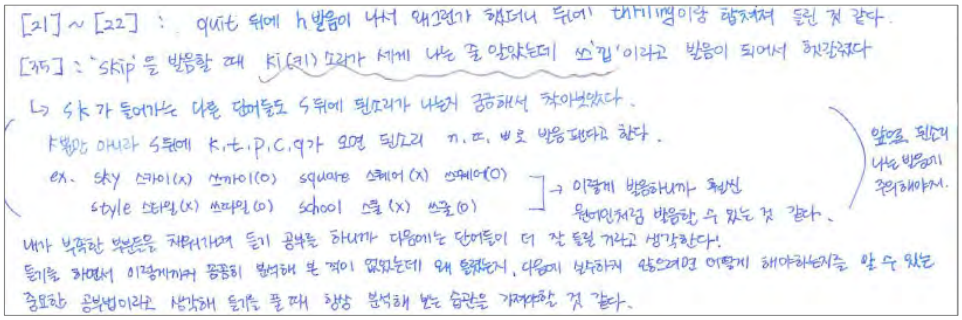
Notably, many students, like S4 who stated, 'I need to study everyday English expressions more!', articulated their future direction of L2 self-study, suggesting that reflection through self-annotation may have facilitated their motivation. Figure 2 presents a student's brief note from further self-study on English phonetic features. S5 specifically noted hearing an [h] sound between 'quite' and 'thrilling.' This could be due to the prevalence of intrusive sounds or epenthesis in connected or fast speech (Uffmann, 2007). Although a detailed academic explanation of epenthesis may be beyond the classroom level, this annotation opened opportunities for relevant whole-class pronunciation practice. By acknowledging S5's meticulous observation, the teacher introduced simple yet frequent examples of epenthesis, such as the schwa [ə] in 'athlete' (/æθəliːt/) and the extra [p] sound in 'hamster' (/hæm*p*stər/), to demonstrate how these processes make transitions between sounds smoother. This approach seemed to enhance students' 'noticing' (Schmidt, 1990), which is crucial for transforming input into intake. Additionally, S5's annotation shows the student's active engagement by repeatedly and voluntarily listening to identify different qualities of aspirated voiceless stops (/p/, /t/, /k/) depending on their position within a word. This indicates a deeper understanding of phonetic variation, which likely helped S5 better grasp the nuances of

[4] S4's errors in the answer can also be attributed to L1 transfer, such as the difficulty distinguishing the /θ/ and /p/ sounds in 'health' and /r/ and /l/ sounds in 'light'—issues commonly encountered by Korean learners due to the absence or indistinguishability of these phonemes in the Korean language. Additionally, an omission error was also observed, as in transcribing 'mission' instead of 'emission.' However, if the student's self-annotation emphasized the challenge of using appropriate collocations rather than pronunciation or omission issues, the error was classified under the theme of 'collocation'. As mentioned in Section 3.5, the coding process was guided by the students' self-analyses of the errors in their annotations, prioritizing their perspectives over the researcher's interpretation.

English pronunciation, ultimately improving their listening comprehension (Kissling, 2018). S5 concluded that an analysis of why mistakes occurred and how to address them would help improve listening proficiency. For S5, self-annotation in listening class seems to encourage motivation and autonomy through increased study and repeated listening.

FIGURE 2

A Student's Further Study Note (S5, 04/25/2024)



In terms of affect, both proficiency levels exhibited comparable overall proportions (Higher-level: 16.07%, Lower-level: 15.43% of the whole self-annotation types for each level); however, upon closer examination of the data, it was revealed that lower-level students expressed significantly more negative sentiments (n = 46) than positive ones (n = 12), whereas the opposite was observed among higher-level students. These adverse emotions encompassed anxiety (e.g., 'I'm afraid I'll make the same mistake next time'), feeling overwhelmed (e.g., 'I couldn't write anything because they talk too fast'), confusion (e.g., 'I don't understand at all in this part'), and frustration (e.g., 'Why is English so difficult? I feel like I have such a long way to go'). Although higher-level students also experienced negative emotions, their expressions tended to be more specific (e.g., 'Again, a calculation problem! I tend to lose focus when they mention numbers') and often included resolutions or study plans (e.g., 'I will both listen and speak when memorizing words'). This highlights the significance of understanding students' emotional states, which play critical roles in L2 learning motivation and performance (For a detailed discussion, see Saito, Dewaele, Abe, & In'nami, 2018).

In small numbers, when students did not discern the words in the blank of cloze tests, some filled the gap actively using prior lexical ('When I heard the word 'budget', it seemed like 'exceed' would fit, so I wrote it and it was correct.') and grammatical knowledge ('Only 'prefer' was heard, but 'any prefer' didn't make sense, so I wrote 'any preference').

Overall, the multifaceted self-annotations provided by Korean EFL students in L2 listening classrooms offer insights into the intricacies of EFL learners' comprehension

challenges and learning processes. These self-annotations helped students engage in learning as active participants and problem-solvers, while the use of student-produced examples as teaching materials ensured authentic and relevant learning experiences. By incorporating these self-annotations, the teacher could tailor her instructional approaches to provide targeted support, positively affecting students' listening proficiency, which will be discussed in the subsequent sections.

4.2. The Changes in Students' Scores in Listening Comprehension Tests

The descriptive statistics provided in Table 2 show the mean score improvements between the pre- and post-test (T2-T1[5]) for both the experimental and control groups, segmented by proficiency level. While both the experimental and control groups improved their scores, affirming the benefit of listening instruction on students' performance, higher proficiency students in the experimental group had a mean improvement of 2.33 (*SD* = 1.30) compared to 1.27 (*SD* = 0.90) in the control group. Similarly, lower proficiency students in the experimental group improved by 1.43 (*SD* = 1.09), whereas their counterparts in the control group showed a mean improvement of 0.67 (*SD* = 1.23). Overall, the experimental group saw a mean improvement of 1.84 (*SD* = 1.26), compared to 0.92 (*SD* = 1.13) in the control group, demonstrating the potential effectiveness of the self-annotation activity in enhancing listening proficiency.

TABLE 2

Descriptive Statistics of Score Improvement (T2-T1) by Level and Group

Level	Group	*N*	*M*	*SD*
Higher	Experimental	12	2.33	1.30
	Control	11	1.27	0.90
Lower	Experimental	14	1.43	1.09
	Control	15	.67	1.23
Total	Experimental	26	1.84	1.26
	Control	26	.92	1.13

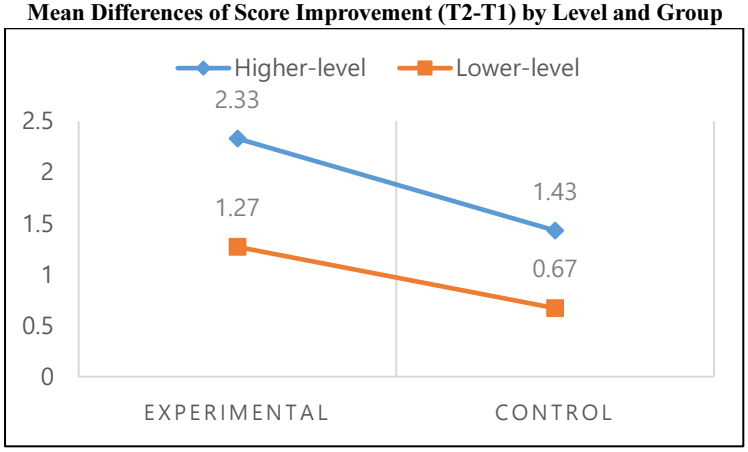
Figure 3 visually illustrates the outcomes, revealing a notable increase in proficiency among students in the experimental group in comparison to those in the control group, irrespective of their initial listening proficiency levels. Upon closer examination, it becomes apparent that higher-level students in the experimental group experienced a slightly more pronounced improvement (*M* = 2.33) than lower-level students (*M* = 1.27) after exposure to

[5] In the context of the CSAT, comprising 17 questions, the inherent scoring system assigns varying point values to each question, ranging between 2 and 3 points per item. However, for the methodological expediency of this study, each question has been uniformly allotted a value of 1 point.

the self-annotation intervention. This suggests that, while the self-annotation intervention appears to be beneficial across all proficiency levels, its effects may be more significant among higher-level listening proficiency students when applied in the short term. This contrasts with previous studies on self-annotation in writing, which demonstrate more significant benefits for lower-level learners (Lee, 2011; Yeh, Lo, & Chu, 2014). This phenomenon is potentially attributable to the greater cognitive readiness and autonomy of higher-level listeners, enabling them to derive more substantial benefits from self-annotation interventions within a limited timeframe.

FIGURE 3

Mean Differences of Score Improvement (T2-T1) by Level and Group



The results of the repeated measures ANOVA within-subjects contrasts, as shown in Table 3, highlight the effectiveness of the self-annotation activity. There was a significant main effect of time ($F = 78.53$, $p < .001$, $\eta^2_p = .62$), indicating substantial improvement in scores over the course of the intervention. The interaction between time and level was significant ($F = 5.52$, $p = .02$, $\eta^2_p = .10$), suggesting that the rate of improvement differed based on proficiency level. Most importantly, the significant time by group interaction ($F = 8.03$, $p = .01$, $\eta^2_p = .14$) demonstrates that the experimental group showed greater improvement over time compared to the control group. However, the three-way interaction between time, level, and group was not significant ($F = 0.22$, $p = .65$, $\eta^2_p = .00$), indicating that the combined effect of time, proficiency level, and group did not significantly influence the outcomes. These findings suggest that the self-annotation activity significantly improved L2 listening scores, with greater gains in the experimental group compared to the control group. Improvement varied by proficiency level, but the combined effects of time, proficiency level, and group did not significantly alter the outcomes.

TABLE 3

Two-Way Repeated Measures ANOVA Results: Within-Subjects Effects and Interactions

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Time	52.04	1	52.04	78.53	.00	.62
Time*Level	3.65	1	3.65	5.52	.02	.10
Time*Group	5.32	1	5.32	8.03	.01	.14
Time*Level*Group	0.14	1	0.14	.22	.65	.00

Note. *SS* = Sum of Squares; *MS* = Mean Square

Table 4 further provides insights into the effectiveness of the self-annotation activity intervention. The main effect of proficiency level was highly significant (F = 182.82, p < .001, η^2_p = .79), demonstrating that higher proficiency students outperformed their lower proficiency counterparts, with proficiency level accounting for a substantial portion of performance variance. However, the main effect of group (experimental vs. control) was not significant (F = 0.52, p = .47, η^2_p = .01), indicating that the self-annotation activity did not result in a statistically significant overall performance difference between the experimental and control groups. The interaction effect between proficiency level and group was also non-significant (F = 0.12, p = .91, η^2_p = .00), indicating that the effectiveness of the self-annotation activity did not significantly vary across different proficiency levels.

TABLE 4

Two-Way Repeated Measures ANOVA Results: Between-Subjects Variable

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Level	1583.37	1	1583.37	182.82	.00	.79
Group	4.54	1	4.54	.52	.47	.01
Level*Group	.11	1	.11	.12	.91	.00

Note. *SS* = Sum of Squares; *MS* = Mean Square

One possible reason for the lack of statistical significance could be the short duration of the intervention, which may not have been sufficient to reveal a significant difference between groups. Additionally, the limited number of listening questions (17 in total, each worth one point) might have constrained the ability to detect notable differences. Despite the lack of statistical significance, observable mean differences hint at the potential benefits of the self-annotation activity. A longer intervention period or a larger set of listening questions may provide more conclusive results.

The next section will analyze students' survey responses to better understand the impact of the self-annotation activity by examining the students' perceptions, experiences, and engagement with the intervention. This qualitative analysis aims to identify factors that could enhance the intervention's effectiveness in future applications.

4.3. The Students' Perceptions of Self-annotation in Listening

After the post-test, a questionnaire adapted from Yang and Kang (2020) was administered (see Table 5). This questionnaire aimed to investigate students' perceptions regarding the annotating technique, which was presented in Korean. The questionnaire was distributed using an online survey platform familiar to the students, i.e., Naver Form, and was conducted anonymously. 25 out of the 26 students in the experimental group participated in the survey.

TABLE 5

Questions on the Post-Questionnaire

1. Have you ever experienced this kind of listening activity before?
   YES/NO
   - If yes, write about your experience in detail and how it was similar to or different from the current activity.
2. Was it difficult for you to understand what the self-annotation technique is and to apply it during or after listening? Circle a number on the 5-point scale.
   5: VERY MUCH 4: SOMEWHAT 3: UNDECIDED
   2: NOT REALLY 1: NOT AT ALL
   - If so, specify the difficulties.
3. Was the self-annotation technique helpful in enhancing your listening skills?
   5: VERY MUCH 4: SOMEWHAT 3: UNDECIDED
   2: NOT REALLY 1: NOT AT ALL
   - If so, write about your experience in detail.
4. What are the advantages and/or disadvantages of annotating activities?

The analysis of students' responses to the questionnaire unveiled that, despite none reporting prior exposure to similar activities or experiences (Question 1), the majority encountered minimal difficulty in implementing the self-annotation technique during listening sessions (Question 2; $M = 1.16$, $SD = 0.61$, Range = 1-3). Notably, the mean rating garnered from Question 3 stood at 4.20 on the five-point Likert Scale ($SD = 0.71$, Range = 3-5), which signifies that students perceived self-annotation as significantly beneficial in bolstering their listening skills.

Regarding their perspectives on the self-annotation activity in listening classes (Question 4), the prevalent response ($n = 24$) highlighted compensating weaknesses through active analysis (e.g., 'In the past, when studying listening, I focused on solving problems, but meticulous reflection enabled me to identify my deficiencies, which helped me address them'). The second most common response ($n = 17$) was the expression of satisfaction and

appreciation for the class. Students perceived the listening instruction as well-organized, tailored to their actual experiences and needs. For instance, one student remarked, 'I have never experienced this kind of listening class before. It was the best. I appreciate the teacher's comments. This class made English listening feel less daunting and more interesting. I now know how to practice, and all I have to do is *just do it*.' Some students (n = 11) reported heightened concentration and retention (e.g., 'Annotating prompts me to listen more attentively, and I tend to remember words longer when I listen and analyze for this (self-annotation) activity').

In addition, enhanced confidence (n = 8) was noted (e.g., 'Instead of merely contemplating my deficiencies, engaging in (annotating) activities instills confidence in me.'). Others (n = 8) noted heightened motivation (e.g., 'Due to my limited vocabulary and incorrect pronunciation, I am compelled to memorize more English words on a regular basis.'; 'Continuously reflecting on my frequent mistakes directs my focus towards similar types and encourages me to solve them.'). The improved rapport between a teacher and students (n = 7) was also highlighted (e.g., 'I really enjoyed this activity. Teachers' comments made me work harder. […] I feel a little more confident and find English more enjoyable'). This was likely due to the students feeling a stronger rapport with the teacher through the self-annotation and teacher feedback cycle, whether individually or through whole-class sharing. Other students (n = 4) acknowledged the benefit of including self-annotation in listening class because this is not something they would do alone (e.g., 'Including this activity is great because I wouldn't do it voluntarily otherwise.').

However, a few drawbacks of self-annotation were also identified. The most prevalent concern (n = 7) pertained to the time-consuming nature of self-annotation. Nevertheless, even in this case, most students acknowledged its benefits despite the additional time investment (e.g., 'Sometimes, completing a single sentence takes too much time to write down reasons, but other than that, I like how it improves my English'). Other respondents (n = 5) suggested that self-annotation might pose challenges for lower-level students (e.g., 'For students who don't know many words or grammar rules, or can't understand what they're listening to, adding notes might get boring fast and make them tired'). One student explicitly expressed frustration, which lead them to a dislike of self-annotations ('Those who are proficient in English understand why and where they made mistakes, and can analyze and improve. However, for those like me who struggle, it's challenging to identify errors and won't see much improvement in listening skills.'). This comment serves as a reminder of the significance of offering diverse forms of scaffolding, such as strategic, cognitive, or affective support tailored to students' needs, particularly in many Korean high school EFL classrooms where mixed proficiency levels are prevalent.

5. DISCUSSION AND CONCLUSION

Guided by theoretical frameworks of metacognitive listening training and empirical studies on self-annotation, this study examined the effectiveness of self-annotation in high school EFL classrooms, particularly in preparation for the CSAT listening test. It also explored the types of self-annotations made by students. This study suggests that active student engagement through self-annotation in a listening classroom could yield greater effectiveness. While conventional listening instruction for test preparation proved beneficial for test performance, with both the experimental and control groups showing score improvements, a more significant increase was observed in the experimental group as compared to the control group. This study empirically corroborates previous findings that EFL students' written reflections in L2 listening can effectively empower them to assume the role of reviewers of their learning process by actively engaging with the listening materials, monitoring their comprehension, and identifying areas for improvement (Chang & Chang, 2014; Lee, 2011; Lee & Cha, 2020; Yang & Kang, 2020).

Among the six primary themes identified in students' self-annotations (i.e., phonetic features, lexical features, resolution, affect, and prior linguistic knowledge), over 60% focused on phonetic and lexical aspects of English listening. Particularly in listening, this emphasis enabled students to actively identify and address their linguistic weaknesses. Some annotations included resolutions through which students could cognitively plan and manage their learning. For some, this led to increased motivation and further action by engaging in additional self-study plans and refining test-taking strategies.

Differing from previous studies which found a more positive impact of reflective practices among lower-level students (Goh & Taib, 2006; Vandergrift & Tagaghodtari, 2010), this study suggests a potential reversal of this relationship: higher listening competence appears to foster deeper reflective practices. This discrepancy may stem from differences in population demographics (university versus high school students), educational contexts (ESL versus EFL), and instructional materials (authentic versus test-preparation), indicating a need for more tailored and context-based research and support. In this study, higher-level students in the experimental group exhibited greater performance gains compared to lower-level students. Higher listening competence appears to correlate with more analytical engagement in self-annotation tasks, as demonstrated by the content, length, and frequency of their annotations. Specifically, higher-level students (n = 12) made 529 self-annotations, whereas lower-level students (n = 14) made a total of 376 self-annotations (see Table 1 for details). This significant difference implies a complex interplay between listening competence and the active engagement of reflective practices.

The only domain where lower-level students commented more frequently was the affective domain, indicating a tendency to express emotions, especially negative ones, more

often. This reliance on affective self-annotations was related to heightened anxiety, limited proficiency, and lack of experience. Concerns raised by students in the post-questionnaire (see Section 4.3.) were also related to the perceived time intensiveness of self-annotation practices and the accompanying cognitive demands, which may pose obstacles for students grappling with foundational listening skills. This indicates that lower-level students may benefit from more tailored scaffolding and support.

Although these concerns exist, students of both proficiency levels generally showed a positive reception towards the self-annotation intervention, as reflected by a mean rating (M = 4.20) on the perception scale. This suggests that self-annotation can be both feasible and effective in high school EFL listening classrooms. Despite lacking prior experience in self-annotation practices, students demonstrated the ability to integrate self-annotation techniques into their listening exercises and articulate their thoughts and emotions, thus showcasing their metacognitive knowledge and practices. The qualitative analysis of students' written feedback also substantiated their favorable attitudes towards self-annotation, highlighting perceived benefits such as ameliorating individual weaknesses, bolstering concentration and retention of English expressions, engendering confidence and motivation, and fostering a deeper rapport with the teacher through feedback. This rapport-building aspect aligns with portfolio assessment research, which demonstrates that ongoing teacher feedback not only enhances student engagement and learning outcomes (Hamp-Lyons, 2006; Hamp-Lyons & Condon, 2000) but also fosters positive emotions by strengthening student-teacher rapport. This improved rapport, in turn, enhances self-assessment and boosts students' self-efficacy (Baturay & Daloğlu, 2010; Lam, 2013; Lam & Lee, 2010; Li, 2016; Romova & Andrew, 2011).

For teachers, this study suggests students' self-annotations can offer invaluable opportunities for targeted support based on actual mistakes made in the classroom. This approach was more relevant to students' experiences, thereby increasing their engagement and comprehension. Based on students' error patterns and examples (see Section 4.1.), the teacher could facilitate open discussions to analyze the reasons and strategies together. This seems to effectively foster a sense of "positive interdependence," as teachers can provide their expertise by facilitating discussions and offering feedback on students' self-annotations, while students feel a greater sense of ownership in their learning as active participants (Goh, 2010, p. 186).

For successful implementation of self-annotation interventions in similar test-oriented contexts, this study identifies three significant factors that help students raise metacognitive awareness and engagement. First, equipping students with knowledge of error types before engaging in self-annotation tasks can be effective. Table 1 shows the frequent patterns hindering students' listening comprehension in CSAT mock tests from their self-annotations. If these features are introduced with actual examples beforehand, these analytical lenses can

provide students with concrete ideas to begin with rather than being left on their own. In this study, some students seemed overwhelmed in the beginning by self-annotation tasks as they struggled to know where to begin. Hence, the teacher in this context shared noticeable or common errors from the students' annotations of the previous class at the beginning of the next class, helping those students with difficulties in self-annotation to produce longer and more detailed analyses of their mistakes.

Second, this study highlights the need for tailoring self-annotation activities to specific proficiency levels, potentially through differentiated scaffolding and instructional strategies. Learners with very low listening proficiency and little motivation may become overwhelmed when faced with severe comprehension difficulties. For them, filling in the blanks in a listening cloze test can be a daunting task, often resulting in many empty spaces. This can significantly discourage them from attempting to fill in the blanks, let alone adding self-annotations. In this study, several measures were implemented, such as breaking tasks into manageable segments (e.g., focusing on one question at a time for annotation), sharing student samples for annotation modeling in class, pair/group annotation, providing teacher elaboration on students' simple annotations, and allowing the use of L1 (Korean) to facilitate exploration of English phonetics without worrying about spelling. This practice encouraged students to annotate and thus actively engage in their learning process.

Third, a teacher's acknowledgment, whether small or large, is critical for building rapport and enhancing students' motivation. High school teachers are often very busy meeting institutional needs and handling administrative work. Spending a lot of time on feedback is not always feasible. In this test-focused context, peer feedback or group work was not always possible either because of time limitations. Instead, various short and time-saving feedback methods were useful and effective such as sharing some students' self-annotations in class, providing praise stamps, leaving short comments and questions next to student annotations, or verbally commenting on their progress and impressions in person. These simple acknowledgments saved time but were powerful because they gave students the impression that their work was being monitored, appreciated, and utilized in class. Indeed, when their self-annotations were introduced and discussed in class, students appeared excited and engaged, as one student explicitly mentioned in the post-questionnaire: 'My annotations were introduced twice in class, and when I found similar mistakes being made among my friends too, and my analysis was complemented by the teacher, I felt awesome'. Consequently, as observed in Sections 4.1. and 4.3., the iterative cycle of student self-annotation and teacher feedback seems to foster a sense of engagement and drive for improvement among the students while increasing rapport. This observation aligns with Vandergrift's (2003) research, which emphasized the motivational benefits of post-reflection activities in listening classrooms. Through collaborative problem-solving and the recognition of perceived progress based on students' self-annotation intervention, teachers

can enhance rapport, increase class satisfaction, bolster their credibility and perceived expertise, and thus significantly improve the learning environment as "effective teachers," as Paris and Winograd (1990) put forth:

> "Effective teachers display both empathy and expertise; they guide students' learning with sensitivity. Classroom practices should allow teachers and students to discuss their thoughts and feelings about learning in order to promote metacognition and motivation." (p. 31)

Despite the undeniable benefits of self-annotation across all levels in the experimental group, the present empirical study is not without limitations. The study's small sample size and restriction to a single high school setting for a relatively brief duration (comprising 10 sessions over two months) may limit the generalizability of its findings. Conducting a long-term investigation involving larger and more diverse participants could yield additional insights applicable to a broader range of educational contexts. Also, improvements in performance on specific test types, such as the CSAT mock listening test utilized in this study, may not consistently indicate an overall enhancement in listening competence. Moreover, authentic materials instead of the CSAT test materials can shed light on alternative facets of self-annotations in listening. Finally, as previous studies warned, students' self-annotation tasks, like any other interventions, should be used judiciously and not overused. Repetitive reflection tasks "run the risk of being boring and tedious to learners after a while" (Goh, 2008, p. 200) and learners may "get easily bored with teaching methods that are used repeatedly throughout a semester" (Lee & Cha, 2020, p. 62). Further research endeavors in not only self-annotations but also other metacognitive interventions (e.g., Cross, 2015) will contribute to a deeper understanding of test-oriented listening classrooms in high school EFL contexts or similar settings with various institutional constraints.

Applicable levels: Secondary

REFERENCES

Al-Khresheh, M. H., & Alruwaili, S. F. (2024). Metacognition in listening comprehension: Analyzing strategies and gender differences among Saudi EFL University students. *Cogent Social Sciences, 10*(1), 1-20.

Bassetti B., Cerni T., & Masterson J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics, 43*(3), 683-705.

Baturay, M. H., & Daloğlu, A. (2010). E-portfolio assessment in an online English language course. *Computer Assisted Language Learning, 23*(5), 413-428.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Chang, C., & Chang, C. (2014). Developing students' listening metacognitive strategies using online videotext self-dictation-generation learning activity. *The EUROCALL Review, 22*(1), 3-19.

Charles, M. (1990). Responding to problems in written English using a student self-monitoring technique. *ELT Journal, 44*(4), 286-293.

Cotterall, S. (2000). Promoting learner autonomy through the curriculum: Principles for designing language courses. *ELT Journal, 54*(2), 109-117.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Los Angeles: SAGE Publications.

Cross, J. D. (2009). Effects of listening strategy instruction on news videotext comprehension. *Language Teaching Research, 13*, 151-176.

Cross, J. (2015). Metacognition in L2 listening: Clarifying instructional theory and practice. *TESOL Quarterly, 49*(4), 883-892.

Dahunsi, T. N., & Ewata, T. O. (2022). An exploration of the structural and colligational characteristics of lexical bundles in L1-L2 corpora for English language teaching. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688211066572

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906- 911.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive development* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Flowerdew, J. & Miller. L. (1992). Student perceptions, problems and strategies in L2 lectures. *RELC Journal, 23*(2), 60-80.

Gilliland, B. (2015). Listening logs for extensive listening practice. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 13-22). New York: Routledge.

Goh, C. C. M. (1997). Metacognitive awareness and second language listeners. *ELT Journal*, *51*(4), 361-369.

Goh, C. C. M. (2002). Exploring listening comprehension tactics and their interaction patterns. *System, 30*(2), 185-206.

Goh, C. C. M. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implication. *RELC Journal, 39*(2), 188-213.

Goh, C. C. M. (2010). Listening as process: Learning activities for self-appraisal and self-regulation. In N. Harwood (Ed.), *English language teaching materials: Theory and practice* (pp. 179-206). Cambridge, England: Cambridge University Press.

Goh, C. C. M. (2018). Listening instruction and assessment: Activities and options. In J. M. Newton, D. R. Ferris, C. C. M. Goh, W. Grabe, F. L. Stoller, & L. Vandergrift (Eds.), *Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking* (pp. 163-179). New York: Routledge.

Goh, C. C. M., & Taib, Y. (2006). Metacognitive instruction in listening for young learners. *ELT Journal, 60*(3), 222-232.

Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower intermediate learners of French. *Language Learning, 58*(4), 747-783.

Hamp-Lyons, L. (2006). Feedback in portfolio-based writing courses. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 140-161). Cambridge, England: Cambridge University Press.

Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. New York: Hampton Press.

Kemp, J. (2010). The listening log: Motivating autonomous learning. *ELT Journal, 64*(4), 385-395.

Kissling, E. M. (2018). Pronunciation instruction can improve L2 learners' bottom-up processing for listening. *The Modern Language Journal, 102*(4), 653-675.

Lam, R. (2013). Two portfolio systems: EFL students' perceptions of writing ability, text improvement, and feedback. *Assessing Writing, 18*(2), 132-153.

Lam, R., & Lee, I. (2010). Balancing dual functions of portfolio assessment. *ELT Journal, 64*(1), 54-64.

Lee, J. (2011). Comment on comment: The effectiveness of student annotations and teacher feedback in L2 business writing. *Korean Journal of English Language and Linguistics, 11*(3), 547-575.

Lee, Y., & Cha, K. (2020). Writing listening logs and its effect on improving L2 students' metacognitive awareness and listening proficiency. *International Journal of Contents, 16*(4), 50-67.

Li, Q. (2016). Chinese EFL learners' perceptions of their experience with portfolio-based writing assessment: A case study. *Chinese Journal of Applied Linguistics, 39*(2), 215-247.

Matsumoto, K. (1996). Helping L2 learners reflect on classroom learning. *ELT Journal, 50*(2), 143-149.

Milliner, B., & Dimoski, B. (2024). The effects of a metacognitive intervention on lower proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research, 28*(2), 679-713.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* New York: Cambridge University Press.

Papaleontiou-Louca, E. (2000). *Metacognition and theory of mind*. Newcastle, England: Cambridge Scholars Publishing.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rahimi, M., & Katal, M. (2012). The role of metacognitive listening strategies awareness and podcast-use readiness in using podcasting for learning English as a foreign language. *Computers in Human Behavior, 28*(4), 1153-1161.

Reid, J. (1994) Responding to ESL students' texts: The myth of appropriation. *TESOL Quarterly*, *28*(2), 273-322.

Renandya, W. (2012). Five reasons why listening strategy instruction might not work with lower proficiency learners. *English Language Teaching World Online, 3.* Retrieved September 20, 2024, from https://repository.nie.edu.sg/entities/publication/dcdfd7e0-32f3-4bdf-826e-c6750afed467/details

Renandya, W. A., & Farrell, T. S. C. (2011). 'Teacher the tape is too fast!' Extensive listening in ELT. *ELT Journal, 65*(1), 52-59.

Romova, Z., & Andrew, M. (2011). Teaching and assessing academic writing via the portfolio: Benefits for learners of English as an additional language. *Assessing Writing, 16*(2), 111-122.

Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning, 68*(3), 709-743.

Schmidt, R. (1990). The role of consciousness in language learning. *Applied Linguistics, 11*(2), 129-158.

Schraw, G., Crippen, K. J. & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.

Song, B., & Kim, T. (2017). The dynamics of demotivation and remotivation among Korean high school EFL students. *System, 65*, 90-103.

Storch, N., & Tapper, J. (1996). Patterns of NNS student annotations when identifying areas of concern in their writing. *System, 24*(3), 323-336.

Uffmann, C. (2007). Intrusive [r] and optimal epenthetic consonants. *Language Sciences, 29,* 451-476.

Vandergrift, L. (1997). The comprehension strategies of second language listeners: A descriptive study. *Foreign Language Annals, 30,* 387-409.

Vandergrift, L. (1998). Successful and less successful listeners in French: What are the strategy differences? *The French Review, 71*(3), 370-394.

Vandergrift, L. (2003). Orchestrating strategy use: Towards a model of the skilled L2 listener. *Language Learning, 53(3)*, 463-496.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* New York: Routledge.

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning, 56*(3), 431-462.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning, 60*(2), 470-497.

Yang, J., & Kang, M. (2020). Application of a self-annotated transcription task in an EFL listening context. *Modern English Education, 21*(3), 37-47.

Yeh, S., Lo, J., & Chu, H. (2014). Application of online annotations to develop a web-based Error Correction Practice System for English writing instruction. *System, 47*, 39-52.