Article

Effects of Metacognitive Strategy Training on Chinese Listening Comprehension

Yanmei Liu

Chinese Department, Language Institute Foreign Language Center, Monterey, CA 93940, USA;
yanmliu@gmail.com

Abstract: In an exploration of solutions to improve Chinese second language learners' listening comprehension, this quantitative quasi-experimental study examines the effects of metacognitive strategy training (MST) on learners' metacognitive awareness, listening performance, and proficiency in an intensive language training program. In contrast with the extant research, this study designed a metacognitive learning cycle model, including self-diagnosis, planning, monitoring, evaluation, regulation, and reflection strategies, as the content of the MST. Six classes, comprised of a total of 80 participants, were assigned into three groups: Self-directed, teacher-led, and control groups. The Metacognitive Awareness Listening Questionnaire and a listening comprehension test were administered as pre- and post-tests, in addition to a proficiency test as a post-test only. Results demonstrate no significant differences in metacognitive awareness development, listening performance gains, and proficiency test results among the three groups. The results do indicate that the self-directed MST better enhances development of students' planning and evaluation awareness, and teacher-led MST workshops with special emphasis on the area of monitoring strategy will help students raise awareness. The findings of this study reveal that insufficient training time and MST without the integration of cognitive strategies do not yield significant effects. It is suggested that future MSTs should involve sufficient training time and effective follow-ups to ensure its positive effects. This study proposes that the effectiveness of MST could be improved by combining it with cognitive strategies training.

Keywords: Chinese; listening; metacognition; metacognitive strategy; strategy training

1. Introduction

1.1. Background

Listening is an essential skill for communication and learning in second language acquisition (SLA). However, listening instruction has been neglected in language curricula and practices for a significant period of time (Field 2008; Luo and Gao 2012; Richards and Rodgers 2001; Rubin 1994; Vandergrift 1999). The reason for this neglect is not that listening is unimportant, but most likely because it is difficult to learn and teach. Some researchers (Graham 2002, 2006; Prince 2013; Vandergrift 2004) have stated that listening is the most difficult skill for learners to develop in SLA, since listening comprehension is a complex, unidirectional, and unobservable process. Graham (2006) posits that "given this complexity and perhaps because the process is largely unobservable, it may be difficult for learners to have a clear understanding of how they go about listening in a foreign language, or, more importantly, how they might improve their performance" (p. 166). Listening is a unidirectional process in which listeners cannot interrupt the speaker for clarification or repetition in some contexts, such as radio broadcasts (Graham et al. 2014). Based on these reasons, "listening is probably the least explicit of the four language skills, making it the most difficult skill to learn" (Vandergrift 2004, p. 4).

Many learners have struggled to improve listening comprehension during their second language acquisition (Chang and Read 2006; Goh 2000; Graham 2006; Prince 2013; Vandergrift 2004). In the past decades, the core issues of second language listening research explored the factors attributed to listening skill development and methods to improve listening comprehension. Studies have shown that successful second language listeners used more metacognitive strategies in their listening tasks than less successful listeners (Goh 2000; Smidt and Hegelheimer 2004; Vandergrift 2002). Research has demonstrated that metacognitive awareness and strategies are crucial to learners' listening skill development. Providing metacognitive strategy training (MST) can help learners become more effective listeners (Goh 1999; Graham and Macaro 2008; Graham et al. 2008; Vandergrift 2003).

In order to assess second language listeners' metacognitive awareness and use of strategies, Vandergrift et al. (2006) developed the *Metacognitive Awareness Listening Questionnaire* (MALQ). The MALQ has been used to examine the effectiveness of MST in listening comprehension instruction in numerous studies (Altuwairesh 2013; Chang and Chang 2014; Coskun 2010; Gagen-Lanning 2015; Movahed 2014; O'Bryan and Hegelheimer 2009; Rasouli et al. 2013). Although these studies showed positive results for the use of MALQ, some doubts related to MST instructional methods and effects remain, because these studies mainly adopted teacher-led methods to provide MST; some even directly used the assessment tool, MALQ, to provide MST content (Coskun 2010; Rasouli et al. 2013). This use of the MALQ triggered more questions about MST, such as whether MST can be provided with a student-directed method and whether, with different training materials instead of using the assessment tool as the training material, MST can achieve the same positive results.

The purpose of this study is to examine the effects of MST when applying different methods—specifically self-directed and teacher-led methods—on Chinese as a second language (CSL) learners' metacognitive awareness, listening performance, and proficiency in an intensive language training program.

1.2. Literature Review

1.2.1. Metacognitive Strategy Training (MST)

The concept of metacognition was introduced by Flavell in 1976. Flavell (1976) advocated that metacognition plays an essential role in various cognitive tasks, including language acquisition, comprehension, learning, and self-instruction. He affirmed that it is beneficial and desirable to increase metacognitive knowledge and improve metacognitive skills by providing systematic training for learners (Flavell 1979). Listening comprehension is a complex cognitive process involving attention, perception, memory, information processing, problem-solving, language, and learning; based on Flavell's theory, metacognition may play an important role in listening comprehension.

Recognizing the importance of metacognition, some educators have researched how to incorporate MST into listening comprehension instruction. Goh (2000) proposed four ways to integrate MST into listening instruction: (a) Discussing problems and strategies, (b) encouraging thinking aloud, (c) using listening diaries to reflect, and (d) incorporating metacognitive activities in pre- and post-listening tasks. Vandergrift (2007) stated that the instructional tools capable of raising metacognitive awareness include questionnaires, listening diaries, and discussions, because these tools can activate listening reflection activities for both learners and teachers. Vandergrift (2007) also introduced a pedagogical cycle involving five stages with seven steps for MST in listening instruction.

When providing MST, the first question to be answered is what kind of metacognitive strategies should be taught therein. Flavell (1979) defined metacognition as a cognitive monitoring model and categorized it into four interacted components: Metacognitive knowledge, metacognitive experiences, tasks, and strategies. Metacognitive knowledge determines what strategies and actions a person adopts in a cognitive task, while metacognitive experiences occur to monitor the cognitive course, including planning beforehand, control in the process, and evaluation of the task afterwards. At least three strategies can be drawn from Flavell's description: Planning, controlling, and evaluating.

Wenden (1998) classified metacognitive strategies in SLA as planning, self-monitoring, self-evaluation, and self-reinforcement. Anderson (2002) proposed a model of metacognition that includes five components: "(a) Preparing and planning for learning, (b) selecting and using learning strategies, (c) monitoring strategy use, (d) orchestrating various strategies, and (e) evaluating strategy use and learning" (Anderson 2002, p. 2). Compared with Wenden's (1998) framework, Anderson's (2002) model lacks a component allowing learners to regulate after evaluation. Furthermore, there is an overlap between (b) "selecting and using learning strategies" and another two components: (a) Planning and (b) orchestrating. However, Anderson's inclusion of employed orchestration is noteworthy, as this concept has seldom been mentioned in other metacognition models. As Anderson stated: "The ability to coordinate, organize, and make associations among the various strategies available is a major distinction between strong and weak second language learners" (p. 4). Alongside the MALQ, Wenden's framework and Anderson's model have been widely adopted as training content in MST studies.

1.2.2. Effects of MST

MST has different effects on the learning of different subjects for students of different ages, for example, Diebold's (2011) and Prestwich's (2008) studies showed that MST had no significant effect on fourth grade students' reading, but there are a number of studies that have proven the positive effect of MST. With the greater attention to metacognitive strategies in the SLA literature, there have been increasingly more studies examining the effects of MST on SLA learners' reading and listening development over the past decade. Sterling's (2011) reading research did not find positive effects from MST, but other reading studies and almost all listening studies have found positive effects.

Among the SLA listening studies, participant samples vary in size, some as few as three (Gagen-Lanning 2015), and some around twenty (Altuwairesh 2013; Coskun 2010), and some more than forty (Abdelhafez 2006; Movahed 2014; Nosratinia et al. 2015; Chang and Chang 2014). Rasouli et al. (2013) conducted a MST study in a large-scale experiment. They examined the effectiveness of MST on 120 Iranian English as a second language (ESL) students' listening proficiency, and their results indicated that MST had a positive impact on the participants' English test results. Rasouli et al.'s (2013) study used the MALQ as an assessment. Unfortunately, Rasouli et al. did not mention the MALQ results and did not discuss the participants' metacognitive awareness changes before and after MST.

The MST treatments in these listening studies also differ in the length of time, training content, and instrument. In Abdelhafez's (2006) study, the experimental group received a 12-week training course comprised of three sets of metacognitive strategies. Coskun's (2010) study adopted the CALLA model proposed by Chamot and O'Malley (1994) and the MALQ (Vandergrift et al. 2006) as training materials for a 5-week MST treatment. Movahed's (2014) study used Vandergrift's (2007) pedagogical cycle to deliver MST, and the instruments included an anxiety scale, the MALQ, and the Test of English as a Foreign Language (TOEFL). Nosratinia et al. (2015) adopted Anderson's Model to provide MST in over 18 sessions.

It is worth noting that the MST methods utilized in the listening studies are also quite diverse. Altuwairesh (2013) employed a two-phase treatment: The MST with guided listening diaries to encourage self-reflection and deliberate practice. The study recognized the usefulness of both phases, but emphasized the significance of deliberate practice and claimed that MST was a necessary part of deliberate practice. Chang and Chang (2014) integrated MST with an online videotext listening activity to investigate their combined effects. Their study used self-dictation-generation (SDG) activities on YouTube to emphasize the use of metacognitive strategies in listening processes. In addition to researching the learners' achievement between pre- and post-listening tests, a questionnaire—the Strategy Inventory for Language Learning (SILL), developed by Oxford (1990)—and a focus-group interview were also implemented to examine listening strategy use. The study found that the online video-SDG activity facilitated the participants' development of a reflection strategy, which constitutes

an alternative instructional method to deliver MST. Similarly, Gagen-Lanning (2015) investigated the impact of MST on students' self-directed use of assisted technology in ESL listening. Gagen-Lanning delivered two 60-min sessions, including metacognitive strategies and TED Talk videos, to three participants, and then encouraged them to use self-directed TED Talk videos for their listening improvement. The study utilized the MALQ, a listening worksheet, screen casting software, and a follow-up survey to collect data. The screencast analysis showed what the participants actually did during the listening task. It is important to check what learners actually do after MST instead of only relying on questionnaire responses, but the sample size (three participants) was too small to draw a reliable conclusion from the findings. The study also found that MST could promote self-directed learning, but whether metacognitive strategies can be learned more effectively with the self-directed method in MST remains unanswered.

Although there are many differences in the MST treatments cited above, they have two points in common. One is that MST improved listening performance and metacognitive strategy use. The other is that the participants of these studies were all ESL learners. Goh (2008) posited that it is necessary to examine MST and the MALQ in various target language contexts.

1.2.3. Chinese Listening Research

Compared with ESL, CSL research is still developing, despite the fact that CSL instruction has a long history. There is little listening research in the CSL field. The reason could be either that listening is difficult to learn and research (Vandergrift 1997) or that listening just simply receives less attention. Oxford (1993) commented that language teachers frequently ignored listening as an essential skill in language learning. Although there have been studies (Chang 2010) investigating learners' reading metacognitive strategies in CSL, the number of listening strategies studies is very low, and the study of listening metacognitive strategies is rare.

The status of Chinese listening research is reflected in the following example from a well-known Chinese academic journal. *The Journal of Chinese Language Teachers Association* is an academic platform to exchange CSL studies and teaching ideas in the United States. It publishes three issues each year and has been in operation since 1966. Only seven listening studies were found in this journal from 1966 to 2016, whereas there were more studies in speaking (12), reading (36), and writing (20). Among the seven studies, four articles were reviews of listening textbooks or material, two discussed listening pedagogical issues, and only one, Cai's (2013) study, was a listening research paper.

Cai (2013) investigated four factors affecting Chinese listening proficiency and whether language heritage had an impact on the factors with 51 CSL learners. Cai's (2013) findings revealed that vocabulary and grammar knowledge are more critical for the development of listening proficiency than sound discrimination skill and metacognitive knowledge, the latter assessed by a revised version of the MALQ. Although Cai's study utilized MALQ, the correlation study is descriptive research and only reflects the relationship between metacognitive knowledge and the listening proficiency of Chinese heritage learners. Whether a cause–effect relationship exists between MST and the listening proficiency of Chinese non-heritage learners still needs to be researched. Further metacognitive research in the CSL field is needed.

1.3. Significance of the Study

The absence of studies concentrating on MST in the CSL context highlights the significance of the present study. This study is of importance for several reasons. Firstly, this study may fill gaps in the MST, MALQ, and even metacognition research in CSL listening literature given it examines the effects of MST with a different target language and population from those of existing literature. Chinese listening strategies have not been explored sufficiently in CSL, and further studies are necessary to develop the current knowledge. The results of this study may lead to a deeper understanding of the effects of MST in a CSL context. Secondly, this study presents a new perspective for MST research in the SLA domain. The existing literature on MST mainly focuses on investigating whether

MST positively impacts learners' listening and reading comprehensions. This study moves one step further: Establishing a metacognitive learning cycle (MLC) model to provide MST content, as well as an overview of examined effects of different MST methods. Thirdly, as there is controversy among existing studies surrounding the effects of MST, the findings of this empirical study could provide researchers a more comprehensive understanding of MST and present further evidence for educators and instructional leaders to enable decision making when adopting MST in world languages programs.

1.4. Research Questions

To achieve the purpose of this study, three research questions (RQ) are addressed:

RQ1: Are there significant differences in metacognitive awareness development among two experimental groups receiving MST (self-directed versus teacher-led) and a control group?

RQ2: Are there significant differences in Chinese listening performance gains among two experimental groups receiving MST (self-directed versus teacher-led) and a control group?

RQ3: Are there significant differences in the results of a Chinese listening proficiency test among two experimental groups receiving MST (self-directed versus teacher-led) and a control group?

2. Materials and Methods

2.1. Research Design

This study employed a quantitative quasi-experimental research method, including single factor within-subject and between-subjects designs. The within-subject design tested whether there were differences in metacognitive awareness and Chinese listening performance between the pre- and post-tests of experimental participants, as measured by the MALQ and classroom listening tests, respectively. In order to ensure the three experimental groups comprised of original classes were comparable, the between-subjects design examined whether there were differences in metacognitive awareness development and Chinese listening performance gains, that is, comparing the changes in the MALQ and the classroom listening pre- and post-tests, respectively, before and after the intervention, or MST. For the examination of final listening proficiency, this study used a statistical method to adjust for possible preexisting differences in the three groups, with the classroom listening pre-test as a covariate. The independent variable was the MST method, including three levels: Self-directed training, teacher-led training, and no training. The dependent variables were participants' metacognitive awareness, Chinese listening performance, and Chinese listening proficiency.

2.1.1. MST Content

In contrast with the extant research, this study designed a metacognitive learning cycle (MLC), including self-diagnosis, planning, monitoring, evaluation, regulation, and reflection strategies, to provide MST. In addition to the four essential metacognitive strategies (planning, monitoring, evaluation, and regulation) highlighted by early researchers, this study incorporated self-diagnosis and reflection into the MST. In the MLC, self-diagnosis is the starting point, and reflection the end point. In practice, the two strategies can be applied simultaneously to form a continuous metacognitive cycle, which means the self-diagnosis is a reflection on previous actions, and the reflection is also a self-diagnosis for the following task. This cycle is displayed in Figure 1.

The training material of the MST included an introduction and the MLC strategies content. The introduction presents what metacognition is, along with why and how to learn metacognitive strategies. The content of the self-diagnosis strategy contains how to self-diagnose listening problems and possible reasons based on a self-assessment with the Listening Self-Diagnosis Assistant (LSDA). The planning strategy consists of two parts: How to make plans and how to effectively manage time. The monitoring strategy states how to focus attention and process meaning during a listening activity. The evaluation strategy focuses on how to assess the quality of listening comprehension and the effectiveness of strategy use. The regulation strategy is about how to reinforce and adjust strategy use

after evaluation. The reflection strategy introduces how to analyze and summarize one's own listening performance and problems using answers to the guided questions in the training material. The content of each strategy covers three parts: What, why, and how.
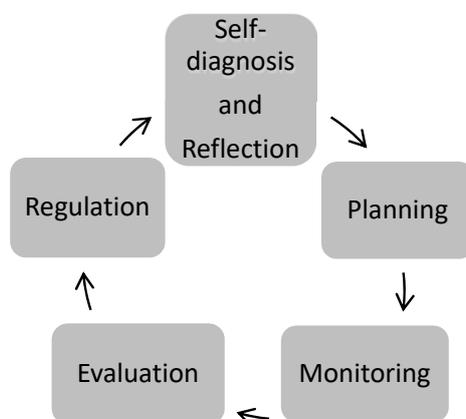
Figure 1. Metacognitive learning cycle.

2.1.2. MST Instruction

In this study, two experimental groups received the same content of the MLC as previously described. The designed learning time of the MLC was the same for both self-directed and teacher-led groups: 90 min. Both groups' participants were encouraged to apply the MLC strategies into daily listening activities and record their strategy use on provided worksheets. There was no preset time length for application activities. The difference between the two experimental groups lies in the instructional method, including training means, procedures, and activities.

The self-directed group attended a self-study session and the training lasted six weeks. The participants spent 15 min on learning one strategy from a handout prepared by the researcher each Monday, and then followed instructions on an application worksheet to practice and record their strategy use every day in that week. The application worksheet provided step-by-step instructions for the strategy use, and varied each week based on the strategy introduced. The application tasks on the worksheet were integrated into the participants' listening homework in this group. A sample page of the self-directed group's worksheets is in Appendix A.

By contrast, the teacher-led group received two 45-min workshops. In the workshops, the researcher introduced the MLC strategies with PowerPoint slides and explained why, when, and how to apply the strategies—the same content as in the handout for the self-directed group. The teacher-led group participants were encouraged to apply the strategies in their daily listening activities whenever and wherever needed after the workshops. They were also asked to record their strategy use on a provided worksheet. However, there were no step-by-step guided instructions for strategy use and no one-by-one focus for each week in the teacher-led group. The worksheet requested the participants to include the date, and to circle the strategy used and the activity applied. A sample of the teacher-led group worksheet is provided in Appendix B.

2.2. Participants and Sampling

The population of this study was the students learning Chinese as a second language in an intensive training program at a world languages center on the west coast in the United States. There were 30 classes, including approximately 360 students in the program when this study started. Among these classes, there were 10 classes in each of the three semesters in the year. As this quasi-experiment was conducted in a real teaching setting, some practical factors had to be considered, such as the appropriate timing of training, available assessment tools, integration with the current curriculum, comparability of different group participants, and so on. These factors led to the decision to adopt a

cluster sampling method for participant selection. According to Babbie (2007), "Cluster sampling is ideal when it is impossible or impractical to compile a list of the elements composing the population" (Creswell 2008, p. 148).

In the first step of cluster sampling, 10 classes of Semester Three became the targeted participating classes. Among the 10 classes, three would graduate in one month and therefore lacked sufficient time to complete the experiment. The remaining seven classes were all invited to voluntarily participate in the study. Six classes accepted the invitation and became participating classes. The students in the same classes were in the same group of the study, in an attempt to mitigate the threat of intervention diffusion among students. There were no heritage learners in the six classes. The students' level of Chinese listening proficiency is Intermediate High or Advanced Low in Semester Three.

For the second step of cluster sampling, the six participating classes were randomly assigned into three groups: Self-directed, teacher-led, and control groups, whereby each group was comprised of two classes. Eighty out of 86 students voluntarily participated in the study, and 72 completed both pre- and post-tests of the MALQ. Four classes, with 49 participants in total, were the experimental groups. Two classes of them, with 21 participants, received the self-directed MST, and the other two, with 28 participants, had the teacher-led MST. Another two classes, with 31 students in total, formed the control group. The participants' demographic information is provided in Table 1.

Table 1. Participants' gender, age, and education frequencies.

		Control Group		Teacher-Led Group		Self-Directed Group	
		N	Percent	N	Percent	N	Percent
Gender	Male	25	80.6	20	71.4	14	66.7
	Female	6	19.4	8	28.6	7	33.3
	Total	31	100.0	28	100.0	21	100.0
Age	19–20	2	15.4	5	41.7	5	23.9
	21–25	8	61.5	5	41.7	9	42.8
	26–30	3	23.1	2	16.6	4	19.0
	31					3	14.3
	Total	13 *	100.0	12 *	100.0	21	100.0
Education	High School	22	71.0	20	71.4	10	47.6
	Associate Degree			4	14.3	2	9.5
	Bachelor's	9	29.0	2	7.1	8	38.1
	Master's			2	7.1	1	4.8
	Total	31	100.0	28	100.0	21	100.0

* Note: One class was not asked to provide ages, but the participants' ages were 19–31 in the class.

2.3. Instrumentation

There were three dependent variables in this quantitative study. The first was listening metacognitive awareness, which was measured by the MALQ. The second, Chinese listening performance, was assessed by the Chinese Listening Comprehension Test (CLCT), a classroom test developed by a test team in the program. The third dependent variable was Chinese listening proficiency, represented by the Defense Language Proficiency Test (DLPT) listening score. The reasons for adopting the two listening test instruments are their convenience and standardization. The CLCT is not a standardized test, but it is convenient for checking changes in participants' listening performances before and after MST. As the DLPT is administered only once at the end of the program at this particular language center, it cannot show the participants' listening performance gains before and after MST; but it is a standardized proficiency test that has been validated.

2.3.1. MALQ

The MALQ (Vandergrift et al. 2006) was designed to assess second language listeners' metacognitive awareness. The reliability of the MALQ ranges from 0.68 to 0.78 according to Cronbach's alpha. During

its validation process, the MALQ developers used the questionnaire to assess French and English learners' listening metacognitive awareness, and a five-factor model emerged based on a confirmatory factor analysis.

The five factors were constituted by 21 items: Factor 1, problem-solving (PS: 5, 7, 9, 13, 17, 19), addresses inference and regulating strategies; Factor 2, planning and evaluation (PE: 1, 10, 14, 20, 21), assesses pre-listening preparation and while or after listening self-judgement; Factor 3, mental translation (MT: 4, 11, 18), contains three obstacles that listeners should overcome to become skilled while listening; Factor 4, directed attention (DA: 2, 6, 12, 16), refers to concentration strategies while listening; and Factor 5, person knowledge (PK: 3, 8, 15), represents listener's self-perception and self-efficacy on task difficulty. The responses to six items (3, 4, 8, 11, 16, 18) must be reverse coded because they are strategies for which lower scores are desirable. In order to avoid a neutral point caused by respondent's hedging, the questionnaire adopted a 6-point Likert scale from "strongly agree" to "strongly disagree". (Vandergrift et al. 2006).

The MALQ has been used for English learners from different countries, such as China, Singapore, Iran, and Saudi Arabia, but it has not been used for CSL non-heritage learners. The researcher received permission to use the MALQ in this study from a leading developer. This study used the original questionnaire without any revisions.

2.3.2. CLCT

The CLCT was developed by an ad hoc test team in the Chinese program at this language center for the purpose of providing a complete test sample to familiarize students with the DLPT format. The CLCT is a traditional paper and pencil instrument and is administered at the beginning and end of semester three in the program. The CLCT includes 40 authentic passages and 60 questions with four answer options. It is graded with raw scores, with a score range from 0 to 60 for the number of questions answered correctly. Each question has only one correct answer. The correct answer keys were provided to each rater, thereby ensuring the inter-rater reliability of different classes.

2.3.3. DLPT

The DLPT is a validated high-stakes proficiency test. It was designed to assess native English speakers' foreign language proficiency as defined by the Interagency Language Roundtable Skill Level Descriptions. The DLPT is used government- and military-wide in the U.S., and is comprised of two separate tests, listening and reading; it is available in different languages. The Chinese DLPT listening test consists of 40 authentic passages with 60 questions related to those passages. The Chinese DLPT is in a multiple-choice format, and each question is followed by four answer choices. The test is delivered on computers at an appointed testing center and graded at the same location. Test scores are in the format of level numbers, including 0+, 1, 1+, 2, 2+, and 3, based on the number of questions answered correctly.

2.3.4. Worksheets

Strategy application worksheets used for the two experimental groups' participants were an additional instrument in the study. Both groups' participants were encouraged to record their strategy applications on the worksheets. Both groups' worksheets were graded based on the same rubric, to quantify the quality and quantity of the experimental participants' actual strategy applications. The scoring rubric included two parts, quantity and quality scores; quantity was the number of uses for a strategy on each day, and quality represented how well the strategy was used each day, including none (0 point), fair (1 point), good (2 points), and excellent (3 points). The worksheet rubric is in Appendix C. The worksheet instrument not only helped the participants apply the strategies learned, but also provided a tool to observe the participants' actual strategy use in MST.

The four instruments formed a continuum of the actual strategy use (worksheets), self-reported strategy use (MALQ), developing listening performance (CLCT), and final listening proficiency (DLPT),

which helped to methodically investigate the effects of MST on the experimental participants' listening comprehension development.

2.4. Procedure

All participants gave their informed consent for inclusion before they participated in this study which was approved by the ethics committees of the researcher's university and affiliation. Before the intervention, the MST, both experimental and control groups had taken the CLCT as a pre-test in the first week of Semester Three. The teachers graded the CLCT. Participants completed the MALQ the second week of Semester Three to assess their listening metacognitive awareness. The researcher administered the MALQ and collected the responses.

The MST started after the completion of the MALQ the third week of Semester Three. The self-directed group participants self-studied one strategy delivered on a handout for 15 min each Monday, and then followed a worksheet to practice and record the strategy use from Tuesday to Friday; the training lasted six weeks, and the total learning time of six MLC strategies was 90 min. The teacher-led group attended two 45-min workshops to learn the six MLC strategies by participating in the trainer's PowerPoint presentations. The first workshop introduced the first three MLC strategies in the third week and the second workshop finished the other three strategies in the fourth week. The participants were encouraged to use a worksheet to record their strategies use on a daily basis after the first workshop. The researcher was the only trainer for the two experimental groups. The researcher developed and provided all learning materials for the self-directed group and conducted MLC workshops for the teacher-led group.

After the MST that lasted 6 weeks, both experimental and control groups completed the MALQ again and the second set of CLCT in their ninth week of Semester Three, and then took the Chinese DLPT in the last week. The researcher collected all data and conducted the analysis.

2.5. Data Analysis

RQ1: Are there significant differences in metacognitive awareness development among the two experimental groups receiving MST (self-directed versus teacher-led) and the control group?

Among 80 participants, 72 took both pre- and post-tests of the MALQ. The data analysis of RQ1 was based on the 72 pairs of MALQ responses. The responses to six items (3, 4, 8, 11, 16, 18) were reverse coded according to the MALQ Scoring and Interpretation Guide. In order to accurately evaluate participants' metacognitive awareness development and reduce the bias of participants' possible differences before the MST, the difference for each item between the pre- and post-tests was calculated first by subtracting the pre-test value from the post-test value. The differences among the two experimental groups and the control group were compared with a one-way ANOVA. In order to look into each group's metacognitive awareness development, each group's pre- and post-test results were compared with a paired-samples *t*-test.

RQ2: Are there significant differences in Chinese listening performance gains among the two experimental groups receiving MST (self-directed versus teacher-led) and the control group?

All 80 participants completed the pre- and post-test of CLCT, and the data analysis was based on their scores in the two tests. A one-way ANOVA was used to examine participants' listening performance gains among three groups, that is, the difference between their pre- and post-test scores. A paired-samples *t*-test was performed, followed by a Pearson correlation analysis, to investigate each group's gains and the relationship with metacognitive awareness development.

RQ3: Are there significant differences in the results of a Chinese listening proficiency test among the two experimental groups receiving MST (self-directed versus teacher-led) and the control group?

All participants completed the Chinese DLPT at the end of the experiment. Their DLPT listening results varied from level 0+ to level 3, specifically 0+, 1, 1+, 2, 2+, and 3. The levels from 0+ to 3 were transformed into numbers from 0 to 5. As participants took the CLCT as a pre-test before the MST, this study made the pre-test scores a covariate when comparing the three groups' DLPT results.

The RQ3 was answered with an analysis of covariance (ANCOVA), a general linear model that blends ANOVA and regression, followed by a Pearson correlation analysis. The ANCOVA was used to reduce the bias from participants' possible different performances before the MST and accurately evaluate the effectiveness of MST. The correlation analysis examined relationships between the DLPT results and the five factors identified in the MALQ.

3. Results

3.1. Metacognitive Awareness Development

After the MST, the two experimental groups demonstrated more awareness than the control group in the metacognitive factors. Table 2 presents the results of the MALQ pre- and post-tests. The overall means of the three participant groups increased from the pre-test to the post-test, with the self-directed group showing the greatest increase (4.48), followed by the teacher-led (3.19) and the control (2.21) groups.

Table 2. Metacognitive Awareness Listening Questionnaire (MALQ) descriptive statistics.

	Group	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Pre-test	Control	24	85.04	7.40	1.51	68.00	98.00
	Teacher-led	27	86.07	9.84	1.86	68.00	107.00
	Self-directed	21	84.57	10.80	2.36	56.00	110.00
	Total	72	85.30	9.31	1.09	56.00	110.00
Post-test	Control	24	87.25	7.32	1.38	72.00	101.00
	Teacher-led	27	89.26	11.81	2.27	70.00	111.00
	Self-directed	21	89.05	8.56	1.87	75.00	103.00
	Total	72	88.46	9.39	1.08	70.00	111.00

There are 21 items constituting five distinct factors/subscales in the MALQ: Problem-solving (PS), planning and evaluation (PE), mental translation (MT), directed attention (DA), and person knowledge (PK). For each of the five factors, the difference between pre- and post-test scores was calculated; this difference represents "metacognitive awareness development." The descriptive statistics of the metacognitive awareness development for the five factors are shown in Table 3. The self-directed group's means of PE, DA, and PK factors were higher than those of the teacher-led and control groups, particularly on PE and PK factors, but lower on PS and MT factors, where the control group showed the greatest development.

Table 3. Metacognitive awareness development on the MALQ factors.

	Group	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
PS	Control	24	0.92	3.361	0.686	−6	10
	Teacher-led	27	0.41	2.707	0.521	−6	5
	Self-directed	21	−0.48	3.696	0.807	−7	8
	Total	72	0.32	3.241	0.382	−7	10
PE	Control	24	1.83	3.102	0.633	−2	8
	Teacher-led	27	2.41	4.236	0.815	−6	11
	Self-directed	21	3.52	6.080	1.327	−4	16
	Total	72	2.54	4.534	0.534	−6	16
MT	Control	24	0.38	2.183	0.446	−4	4
	Teacher-led	27	0.07	2.269	0.437	−5	6
	Self-directed	21	0.10	2.071	0.452	−4	3
	Total	72	0.18	2.158	0.254	−5	6

Table 3. *Cont.*

	Group	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
DA	Control	24	0.58	1.816	0.371	−3	5
	Teacher-led	27	0.93	2.385	0.459	−3	6
	Self-directed	21	1.00	2.720	0.594	−6	6
	Total	72	0.83	2.295	0.270	−6	6
PK	Control	24	−0.67	2.078	0.424	−5	4
	Teacher-led	27	−0.30	2.599	0.500	−6	6
	Self-directed	21	0.33	2.352	0.513	−4	5
	Total	72	−0.24	2.365	0.279	−6	6

Notes: PS = Problem-solving, PE = Planning and Evaluation, MT = Mental Translation, DA = Directed Attention, PK = Person Knowledge.

A one-way ANOVA was conducted to compare the three groups' metacognitive awareness development within the five factors. The results indicate that there were no statistically significant differences in metacognitive awareness development among the two experimental groups (self-directed and teacher-led) and the control group.

In order to examine metacognitive awareness development within groups, each group's pre- and post-scores for the five factors were compared using a paired-samples *t*-test. The *t*-test result presented in Table 4 shows that there were significant differences in PE factor development within the self-directed ($p = 0.015$), teacher-led ($p = 0.007$), and control ($p = 0.008$) groups. The different means indicate that the self-directed (M = 3.524) group showed the greatest difference between pre- and post-tests, followed by the teacher-led (M = 2.407) and control (M = 1.833) groups. Additionally, there was a significant difference in DA factor development within the teacher-led group ($p = 0.05$), although the self-directed group showed the greatest mean difference (M = 1.000). This indicates that all three groups showed a significant increase in metacognitive awareness development on the PE factor, regardless of MST intervention, but only the teacher-led group had a significant increase in the DA factor after MST. Table 4 confirmed the results from Tables 2 and 3. That is, except for PS and MT, two cognitive-focused skills, the other three metacognitive factors clearly demonstrated higher means for the experimental groups than for the control group.

Table 4. Tests of within-subject effects on metacognitive awareness development.

Group		Paired Differences					T	df	Sig. (2-Tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Control	PS	0.917	3.361	0.686	−0.503	2.336	1.336	23	0.195
	PE	1.833	3.102	0.633	0.523	3.143	2.895	23	0.008 *
	MT	0.375	2.183	0.446	−0.547	1.297	0.841	23	0.409
	DA	0.583	1.816	0.371	−0.183	1.350	1.574	23	0.129
	PK	−0.667	2.078	0.424	−1.54	0.211	−1.572	23	0.130
Teacher-led	PS	0.407	2.707	0.521	−0.663	1.478	0.782	26	0.441
	PE	2.407	4.236	0.815	0.732	4.083	2.953	26	0.007 *
	MT	0.074	2.269	0.437	−0.823	0.972	0.170	26	0.867
	DA	0.926	2.385	0.459	−0.017	1.869	2.018	26	0.054 *
	PK	−0.296	2.599	0.500	−1.32	0.732	−0.592	26	0.559
Self-directed	PS	−0.476	3.696	0.807	−2.16	1.206	−0.590	20	0.562
	PE	3.524	6.080	1.327	0.756	6.291	2.656	20	0.015 *
	MT	0.095	2.071	0.452	−0.848	1.038	0.211	20	0.835
	DA	1.000	2.720	0.594	−0.238	2.238	1.685	20	0.108
	PK	0.333	2.352	0.513	−0.737	1.404	0.649	20	0.523

* Correlation is significant at the 0.05 level (2-tailed). Notes: PS = Problem-solving, PE = Planning and Evaluation, MT = Mental Translation, DA = Directed Attention, PK = Person Knowledge.

The participants in the two experimental groups were asked to apply and record their metacognitive strategy use, and, although the participants did not record many uses, their application data was collected and graded according to a scoring rubric. The application scores were analyzed with a Pearson correlation to investigate relationships between strategy application and metacognitive awareness development on the five factors. There were no significant correlations between participants' strategy applications and metacognitive awareness development on the five factors ($p > 0.05$).

3.2. Listening Performance Gains

Table 5 shows descriptive statistics for participants' pre- and post-test results in the CLCT. Table 6 presents three groups' listening performance gains, which are the differences between their post- and pre-test scores. The results in Table 6 show that the control group had the highest gain (M = 9.19; SD = 4.64), followed by the teacher-led (M = 6.82; SD = 4.46) and self-directed (M = 6.00; SD = 5.99) groups.

Table 5. Descriptive statistics for listening performance.

Group		Mean	N	Std. Deviation	Std. Error Mean
Control	Pre	39.19	31	5.606	1.007
	Post	48.39	31	5.258	0.944
Teacher-led	Pre	41.61	28	7.445	1.407
	Post	48.43	28	6.839	1.292
Self-directed	Pre	40.19	21	8.424	1.838
	Post	46.19	21	6.638	1.449

Table 6. Descriptive statistics for listening performance gains.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Control	31	9.1935	4.63623	0.83269	7.4930	10.8941	0.00	16.00
Teacher-led	28	6.8214	4.46429	0.84367	5.0904	8.5525	−4.00	14.00
Self-directed	21	6.0000	5.99166	1.30749	3.2726	8.7274	0.00	24.00
Total	80	7.5250	5.09399	0.56953	6.3914	8.6586	−4.00	24.00

A one-way ANOVA was performed to examine whether the differences among the three groups were statistically significant. The ANOVA results indicate that there were no significant differences (F = 3.018, $p = 0.055$) in listening performance gains among the three groups, although the control group showed the highest mean in the performance gain. Additionally, a Levene's test for the homogeneity of variance was employed, confirming that variances among the three groups were equal (F = 0.700, $p = 0.500$).

Before running the paired-samples *t*-test, the database was split by the MST variable. The paired-samples *t*-test results in Table 7 show that there were significant differences between the pre- and post-test scores in the CLCT for all three groups, including the self-directed ($p = 0.000$), teacher-led ($p = 0.000$), and control ($p = 0.000$) groups.

Table 7. Paired-samples *t*-test on listening performance.

Group		**Paired Differences**					**T**	**df**	**Sig. (2-Tailed)**
		Mean	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**				
					Lower	**Upper**			
Control	Pre-Post	−9.194	4.636	0.833	−10.894	−7.493	−11.041	30	0.000 *
Teacher-led	Pre-Post	−6.821	4.464	0.844	−8.552	−5.090	−8.085	27	0.000 *
Self-directed	Pre-Post	−6.000	5.992	1.307	−8.727	−3.273	−4.589	20	0.000 *

* Correlation is significant at the 0.05 level (2-tailed).

A Pearson correlation analysis was adopted to examine whether the significant listening performance gains in the three groups were related to participants' metacognitive awareness development. The results show that there were no significant correlations between listening performance gains and metacognitive awareness development on the five factors ($p > 0.05$). This means there were likely some other factors that resulted in participants' Chinese listening performance gains across the three groups.

3.3. DLPT Listening Results

The descriptive statistics of the three groups' DLPT listening results show that the teacher-led group's mean (M = 4.07; SD = 0.900) was slightly higher than the control (M = 3.87; SD = 0.846) and self-directed (M = 3.71; SD = 1.102) groups' means. This is similar to the groups' CLCT pre-test sequence, where the teacher-led group (M = 41.61) was slightly higher than the self-directed (M = 40.19) and control (M = 39.19) groups. The Levene's test for equality of variances demonstrates that error variances of the dependent variable were equal across groups (F (77) = 2.055, $p > 0.05$).

An ANCOVA was performed to answer the RQ3, with reduction in the bias from participants' possible different performances before the MST. In the ANCOVA analysis, the DLPT listening results were set as the dependent variable, the MST group as the independent variable, and the CLCT pre-test score as a covariate to compare the three groups' DLPT listening proficiency. The ANCOVA results presented in Table 8 shows that there was no overall statistically significant difference in Chinese DLPT listening results among the three groups after their means were adjusted using the CLCT pre-test scores [p-value (F = 0.733, p = 0.484)].

Table 8. Tests of between-subjects effects on Defense Language Proficiency Test (DLPT) listening proficiency.

Source	**Dependent Variable: DLPT Listening**					
	Type III Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	22.139 [a]	3	7.380	11.917	0.000	0.320
Intercept	2.040	1	2.040	3.294	0.073	0.042
Performance Pretest	20.565	1	20.565	33.211	0.000	0.304
MST	0.907	2	0.454	0.733	0.484	0.019
Error	47.061	76	0.619			
Total	1286.000	80				
Corrected Total	69.200	79				

[a]: R Squared = 0.320 (Adjusted R Squared = 0.293).

After the ANCOVA analysis, a Pearson correlation analysis was performed in order to examine relationships between Chinese DLPT listening scores and the five metacognitive factors measured in the post-test of the MALQ. The results presented in Table 9 show that the DLPT listening results were significantly but negatively correlated with the PE factor (r = −0.235, p = 0.041) and the MT factor (r = −0.238, p = 0.039), and not significantly correlated with the other three factors. The negative correlations means that high DLPT scores were associated with low scores in the PE and MT factors.

Table 9. Correlations between DLPT listening proficiency and the MALQ factors.

		DLPT Listening	PS Post-Test	PE Post-Test	MT Post-Test	DA Post-Test	PK Post-Test
DLPT Listening	Pearson Correlation	1	0.090	−0.235 *	−0.238 *	−0.064	−0.052
	Sig. (2-tailed)		0.441	0.041	0.039	0.581	0.657
	Sum of Squares and Cross-products	69.200	21.658	−69.211	−46.816	−8.961	−8.368
	Covariance	0.876	0.289	−0.923	−0.624	−0.119	−0.112
	N	80	76	76	76	76	76

* Correlation is significant at the 0.05 level (2-tailed). Notes: PS = Problem-solving, PE = Planning and Evaluation, MT = Mental Translation, DA = Directed Attention, PK = Person Knowledge.

3.4. Results Summary

Overall results showed that there were no statistically significant differences in metacognitive awareness development, listening performance gains, and DLPT listening proficiency among the three groups. However, there were significant differences in PE factor development within the three groups, with the self-directed group showing the highest difference between pre- and post-tests, followed by the teacher-led and control groups. There was also a significant difference in DA factor development within the teacher-led group. Furthermore, the two experimental groups demonstrated more awareness than the control group in most metacognitive factors except for two cognitive-focused factors, PS and MT, in the MALQ descriptive data. The self-directed group showed higher scores on DA, and particularly on PE and PK, than the other two groups. The results indicate that the self-directed MST can bring the highest significant difference on the PE factor and the teacher-led MST will significantly improve DA awareness. Both MST training methods promoted the awareness of metacognitive factors except for the cognitive-focused factors. The different MST methods did not show significant differences in participants' Chinese listening performance gains nor in proficiency test results.

4. Discussion

4.1. RQ1 Findings

It was anticipated that there would be significant differences in metacognitive awareness development among the two experimental groups receiving MST (self-directed and teacher-led) and the control group. The differences were not supported by the statistics found in this study. In other words, there was no apparent development of metacognitive awareness after participating in MST. However, the *t*-test results indicate that the self-directed MST can bring the highest significant difference on the PE factor and the teacher-led MST can significantly improve DA awareness.

One possible reason for the non-significant overall results could be the differences between the instrument (MALQ) and the training content of MST (MLC). The MALQ is a questionnaire using self-report to assess metacognitive awareness represented by five factors: Problem-solving, planning and evaluation, mental translation, directed attention, and person knowledge. This study did not directly teach the five factors, but used the MLC model, consisting of self-diagnosis, planning, monitoring, evaluation, regulation, and reflection strategies, to provide MST. Although there were some overlaps (planning and evaluation) and similarities (DA factor and monitoring strategy, PK factor and self-diagnosis strategy) between the MALQ and the MLC, there were differences on other factors, such as problem-solving and mental translation in the MALQ, and regulation and reflection in the MLC. Some previous studies used the MALQ to provide MST content (Coskun 2010; Rasouli et al. 2013), which may then lead to a significant positive effect when answering the questionnaire. The MALQ is likely not the best instrument to assess what was taught in an experiment with different MST materials, such as the MLC model in this study.

The second possible explanation for the non-significant results could be attributed to insufficient training time and the teaching quality of the MST in this study. Although 90 min was sufficient for

introducing metacognitive strategies to the experiment participants, it may have been insufficient for a positive impact on the change of their metacognitive awareness. The participants might need more time to apply and internalize the strategies learnt, and the positive effects of MST might be reflected after more practice. In previous studies, treatment time was 45 to 50 min per week for five or more weeks (Birjandi and Rahimi 2012; Coskun 2010; Rasouli et al. 2013). For example, Birjandi and Rahimi's (2012) MST lasted six weeks and took 45 min once a week. Aside from the time insufficiency, the intensity of training might not have been enough; the researcher had concerns that the experiment participants would opt to withdraw from the MST if the training was too intense, because they already had a heavy learning workload and other duties on a daily basis. In addition, because participation was completely voluntary, the training lacked accountability, which means there were no follow-up actions if the participants did not practice, record, and submit their strategies application worksheets. The lack of accountability may have influenced the quality of the MST. As a result, the MST appears not to have had an overall impact on the experiment participants' listening metacognitive awareness development.

The *t*-test results showed that there were significant differences in planning and evaluation factor development for all three groups, including the control group. Planning before listening and evaluating after listening are two strategies commonly used in listening activities and may be taught in listening instruction or improved by learners themselves as they use metacognitive strategies. If participants did not spontaneously improve their planning and evaluation strategies as they developed their listening comprehension skill, then the two strategies were very likely taught to the control group during their routine listening instruction, which may have resulted in the significant gains in this group. The experimental groups' significant improvement could be caused by the MST interventions as well as by routine listening instruction. The two experimental groups, self-directed and teacher-led, both showed higher means than the control group, and the self-directed group showed the highest development on the planning and evaluation factor. Additionally, the planning and evaluation strategies were directly taught in the MLC model, and were measured by the MALQ as a merged factor. This may illustrate that the matching of training content and assessment tools could lead to significant differences between the participants' pre- and post-tests in the MALQ.

Furthermore, the *t*-test result showed that there was a significant difference on the development of the direct attention factor in the teacher-led group only. A possible explanation for this might be that the importance of the monitoring strategy was particularly emphasized in the teacher-led MST workshops. The monitoring strategy in the MLC is focused on two parts of the listening process: Attention focusing and meaning processing. The first, attention focusing, is very close to the DA factor assessed in the MALQ. When introducing the monitoring strategy, the researcher highlighted that this is the most important but most difficult metacognitive strategy for the listening skill. The special emphasis on the importance of this strategy might have attracted the teacher-led participants' attention to it. This also further speaks to the positive effects on MST that the matching of MST content and assessment tool could bring. On the other hand, compared with the PE factor, the difference on the DA factor between the teacher-led group and the other two groups may indicate that students cannot develop the monitoring strategy spontaneously or through self-directed learning, but teacher-led workshops with special emphasis on the strategy will help students raise awareness in this area.

The significant differences on PE and DA reflected in the *t*-test show that the interventions with the two MST methods, self-directed and teacher-led, have a certain effect on the metacognitive awareness development, but not on cognitive skills, such as the problem-solving factor in the MALQ, because it was not included in the MLC training content. Literature reviews indicate that promoting metacognitive awareness can improve listening performance and proficiency. However, since there were no significant differences in metacognitive awareness among the three participant groups, it may very well be that the three groups were actually not different with respect to the other dependent variables in this study, and, therefore, we would not expect any significant effects in RQ2 and RQ3.

4.2. RQ2 Findings

The ANOVA results of RQ2 indicated that there were no significant differences in listening performance gains among the three participant groups, although there were significant differences between pre- and post-tests across the three groups. A Pearson correlation analysis showed that there was no significant correlation between participants' listening performance gains and their metacognitive awareness development. This finding is consistent with those previous studies that did not discover significant effects of MST (Diebold 2011; Leary 1999; Prestwich 2008; Sterling 2011), but differs from the research showing the significant effects of MST (Coskun 2010; Fan 2009; Movahed 2014; Wang 2009).

Fan's (2009) and Wang's (2009) studies, for example, revealed significant effects when using MST combined with cognitive strategy training. Fan (2009) explored the impact of MST on ESL learners' reading comprehension with 143 first-year university students. Her study showed that the participants receiving MST performed better on a reading comprehension test than the students without MST. The three metacognitive strategies taught in her MST were think-aloud, text structure, and summarization. Rigorously speaking, text structure and summarization are not metacognitive strategies, but reading cognitive strategies. Similarly, Wang (2009) employed a sequential mixed-method research design to investigate the effects of MST on ESL high school students' reading comprehension, strategies awareness, and motivation. Wang's (2009) dissertation presented a strong research design, but the core content of her experimental treatment were basic reading strategies rather than the metacognitive strategies she originally proposed and presented in her title. Fan's (2009) and Wang's (2009) studies reminded the researcher that MST with a stress on cognitive strategies could result in more positive significant effects on participants' cognitive skills. However, the MST intervention in this present study focused on metacognitive strategies without integrating any cognitive strategies. This could reveal that a MST focused exclusively on metacognitive strategies without integrating cognitive strategies may have a limited impact on listening performance, because listening comprehension is a cognitive task, and therefore requires cognitive strategy application.

The provision of explicit instruction combining metacognitive and cognitive strategies for MST has been advocated by a number of scholars (Schraw 1998; Veenman et al. 2006; Schneider 2008). Leopold and Leutner (2015) conducted three experiments to compare the effects of cognitive-only strategy training and cognitive and metacognitive combined strategy training on students' learning using scientific texts. Their results showed that the combined training helped students improve performance; the cognitive-only strategy training was not effective. This combined MST effect was proven across their three experiments. Based on the previous studies and the findings of RQ2, this present study makes a new proposal for future research, that effectiveness in MST may be improved by combining it with cognitive strategy training. The cognitive strategies could include pre-listening prediction, skipping unknown parts while listening, and post-listening information compensation strategies. These cognitive strategies can be combined with corresponding metacognitive strategies, for example, prediction can be introduced into the planning strategy, the skipping strategy integrated with the monitoring strategy, and the compensation strategy included in the evaluation strategy.

Regarding the results of non-significant differences between the self-directed and teacher-led groups, this present study is different from Ball (1998) and Manning (2003), but echoes Leary's (1999), Prestwich's (2008), and Sterling's (2011). Sterling (2011) examined the effects of teacher-led and peer-led MST on community college students' achievement in English, finding that there was no statistical difference between the two methods.

4.3. RQ3 Findings

The RQ3 results showed that there was no significant difference in DLPT listening proficiency among the three participant groups. This indicates that the different MST methods did not produce significant differences in listening proficiency in this study.

As with the RQ1 and RQ2 findings, the results for RQ3 might be attributed to insufficient training time and the ineffectiveness of the MST. The RQ1 results showed that there were no significant

differences in metacognitive awareness development among the three participant groups. This means the three groups were not different when it comes to metacognitive awareness, and, therefore, one might expect no significant differences in the related dependent variables, including listening proficiency, among the three groups. As discussed in the RQ2 findings, the MST curriculum focused only on metacognitive strategy training and did not provide cognitive strategy training. The MST alone, without connecting corresponding cognitive strategy in training, could have a very limited impact on listening proficiency.

The DLPT, a measure of language proficiency, is not closely aligned with the MST in this study—the training of generic metacognitive strategies. Language proficiency is a comprehensive result of language learning and is influenced by a combination of factors. This quasi-experiment was conducted in the existing classes that had different students, teachers, teaching materials, and daily teaching schedules. These factors might be relevant to the participants' listening proficiency. The participants in the three groups might have different motivations for improving their listening proficiency and adopting varied listening practices. The teachers in the three groups may have different knowledge and teaching skills about listening instruction and metacognition. The participating classes had different teacher to student ratios and utilized different teaching materials and schedules during the experiment. These uncontrollable factors may have potentially influenced participants' listening performance gains and proficiency results. The MST was not necessarily the only factor that impacted the listening performance and proficiency results.

Interestingly, the correlation analysis of RQ3 showed that participants' DLPT listening proficiency negatively correlated with the PE and MT factors. This means that the participants who use the PE less or the MT more could obtain higher results in the DLPT listening test. Within the MALQ, the MT factor represents the translation strategies that skillful listeners must avoid (Vandergrift et al. 2006). In this study, all three items (4, 11, 18) representing the MT factor were reverse coded before data analysis; therefore, the negative correlation indicates that the more MT is used, the higher the listening proficiency, which is consistent with Goh and Hu (2014) findings. In their multiple regression analysis with 113 participants, they also found that there was a significant negative correlation between the MT factor and listening proficiency test results. Goh and Hu explained that this correlation might relate to their participants' lack of vocabulary and inability to recognize the sounds of words while listening. The repeated negative correlation results in this present study may suggest that the relationship between translation strategy and listening proficiency is not as simple as commonly thought when speaking of metacognition, and needs further research. This suggestion can be supported by Liu's (2008) study results, where the translation strategy showed variations between the most and least efficient listeners among advanced, intermediate, and elementary levels.

5. Conclusions

In this study, three research questions were not upheld by statistical results. The results indicate that MST does not necessarily show statistically significant effects on metacognitive awareness development, listening performance gains, and proficiency test results. The non-significant results may be attributed to the instruments, MST duration and intensity, and some uncontrollable factors related to the participating classes in a routine language course.

Despite the fact that the results were not as expected, this study answered some unresolved questions and may reveal important principles for future MST design. This study demonstrates that the MST using training materials different from the assessment tool, such as the MLC model in this study, may not necessarily achieve the same significant effects as those with the MALQ as training materials. The results do indicate that the self-directed MST enhances development of students' planning and evaluation awareness more than the teacher-led MST and non-training. It was also found that students could not develop their monitoring strategy spontaneously or through self-directed learning, but teacher-led MST workshops with special emphasis on the importance of monitoring strategy will help students evolve in this area. With respect to previous MST studies, the findings

of this study reveal that insufficient training time and the MST without the integration of cognitive strategies do not yield significant effects. Therefore, it is suggested that future MSTs should last for a longer period, with sufficient training time and effective follow-ups to ensure its positive effects. In addition, this study proposes a hypothesis for further research: MST effectiveness could be improved by combining it with cognitive strategy training.

Conflicts of Interest: The author declares no conflict of interest.

Disclaimer: The content of this article is the sole responsibility of the author and is not necessarily the official views of, or endorsed by the author's affiliation.

Appendix A

Table A1. A sample of application worksheets for the self-directed group.

Week 1: Self-diagnosis Learning Objectives: 1. Identify your listening problems. 2. Find possible reasons of the problems. Tasks: Follow instructions of each day, complete assigned activities with 15 min daily.			
Instructions	Problems Identified	Possible Reasons	Notes
Day 1: Review the LSDA. Identify and summarize your problems and possible reasons with the LSDA. Fill out the right columns.			
Day 2: Listen to a text in homework. After listening, fill out the right columns. Pay attention to linguistic problems.			
Day 3: Listen to a text in homework. After listening, fill out the right columns. Pay attention to your listening strategies.			
Day 4: Listen to a text in homework. After listening, fill out the right columns. Pay attention to your management strategies, anxieties, and nervousness.			
Day 5: Listen to a text in homework. Check the problems identified and summarize patterns.			

Appendix B

Table A2. Application worksheet for the teacher-led group.

Date Length	What Activity (circle one or more as actual use)	Strategies Used (circle one or more as actual use)	Notes
MM/DD mins	Homework Self-study Preview Review Others	Self-diagnosis Planning Time-management Monitoring Evaluation Regulation	

Appendix C

This rubric was used for quantifying experimental participants' strategy applications recorded on their worksheets in the MST. The strategy applications were scored based on the quantity and quality of strategies used on each day. Quantity is the number of the strategies used on each day. Quality represents how well the strategy is used. Since each day is represented by one row on the recording worksheets, the following rubric is the scoring standard for each row.

Table A3. Scoring rubric for strategy use on the application worksheets.

	None (0 point)	Use (1 point)			Max. points for each row
Quantity	No application record in any column on each row.	Shows one application record in any one column on each row.			1
	None (0 point)	Fair (1 point)	Good (2 points)	Excellent (3 points)	Max. points for each row
Quality	No application record in any column on each row.	Shows an application record in any one column on each row.	Shows application records in any two columns on each row. Or shows an application record in any one column on each row with detailed description or multiple circles.	Shows application records in any three columns on each row. Or shows records in any two columns on each row with detailed description or multiple circles.	3
Total points	0	2	3	4	4

References

Abdelhafez, Ahmed M. M. 2006. The effect of a suggested training program in some metacognitive language learning strategies on developing listening and reading comprehension of university EFL students. In *ERIC Online Submission*. Available online: https://eric.ed.gov/?id=ED498262 (accessed on 16 April 2016).

Altuwairesh, Nasrin. 2013. Expertise in L2 Listening: Metacognitive Instruction and Deliberate Practice in a Saudi University Context. Ph.D. dissertation, University of Leeds, Leeds, UK.

Anderson, Neil J. 2002. The role of metacognition in second language teaching and learning. In *ERIC Digest*; ERIC Identifier: ED463659. Available online: files.eric.ed.gov/fulltext/ED463659.pdf (accessed on 15 May 2020).

Babbie, Earl. 2007. *The Practice of Social Research*, 11th ed. Belmont: Wadsworth/Thomson.

Ball, Marjann K. 1998. The Effects of Thinking Maps on Reading Scores of Traditional and Nontraditional College Students. Ph.D. dissertation, University of Southern Mississippi, Hattiesburg, MP, USA.

Birjandi, Parviz, and Amir Hossein Rahimi. 2012. The effect of metacognitive strategy instruction on the listening performance of EFL students. *International Journal of Linguistics* 4: 495–17. [CrossRef]

Cai, Wei. 2013. Investigating second language listening: Factors affecting Chinese listening and the effect of language heritage. *Journal of Chinese Language Teachers Association* 48: 67–97.

Chamot, Anna Uhl, and J. Michael O'Malley. 1994. *The CALLA Handbook: Implementing the Cognitive Academic Language Learning Approach*. New York: Addison-Wesley Publishing Company.

Chang, Cecilia. 2010. See how they read: An investigation into the cognitive and metacognitive strategies of nonnative readers of Chinese. In *Research among Learners of Chinese as a Foreign Language*. Edited by Michael Everson and Helen Shen. Honolulu: University of Hawaii, National Foreign Language Resource Center (NFLRC), pp. 93–116.

Chang, Ching, and Chih-Kai Chang. 2014. Developing students' listening metacognitive strategies using online videotext self-dictation-generation learning activity. *The EuroCALL Review* 22: 3–19. [CrossRef]

Chang, Anna C., and John Read. 2006. The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly* 40: 375–97. [CrossRef]

Coskun, Abdullah. 2010. The effect of metacognitive strategy training on the listening performance of beginner students. *Novitas-ROYAL (Research on Youth and Language)* 4: 35–50.

Creswell, John W. 2008. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed. Newbury Park: Sage publications, p. 148.

Diebold, Tamara W. 2011. Relationship between Metacognitive Strategy Instruction and Reading Comprehension in At-risk Fourth Grade Students. Master's thesis, Walden University, Minneapolis, MN, USA.

Fan, Hsiu-Chiao S. 2009. The Effectiveness of Metacognitive Strategies in Facilitating Taiwanese University Learners in EFL Reading Comprehension. Ph.D. dissertation, University of Kansas, Lawrence, KS, USA.

Field, John. 2008. *Listening in the Language Classroom*. Cambridge: Cambridge University Press.

Flavell, John H. 1976. Metacognitive aspects of problem-solving. In *The Nature of Intelligence*. Edited by Lauren B Resnick. Hillsdale: Erlbaum, pp. 231–36.

Flavell, John H. 1979. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34: 906–11. [CrossRef]

Gagen-Lanning, Kelsey. 2015. The Effects of Metacognitive Strategy Training on ESL Learners' Self-directed Use of TED Talk Videos for Second Language Listening. Master's thesis, Iowa State University, Ames, IA, USA.

Goh, Christine. 1999. How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics* 4: 17–40.

Goh, Christine. 2000. A cognitive perspective on language learners' listening comprehension problems. *System* 28: 55–75. [CrossRef]

Goh, Christine. 2008. Metacognitive instruction for second language listening development: Theory, practice, and research implications. *RELC Journal* 39: 188–13. [CrossRef]

Goh, Christine, and Guangwei Hu. 2014. Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language Awareness* 23: 255–74. [CrossRef]

Graham, Suzanne. 2002. Experiences of learning French: A snapshot at Years 11, 12 and 13. *Language Learning Journal* 1: 15–20. [CrossRef]

Graham, Suzanne. 2006. Listening comprehension: The learners' perspective. *System* 34: 165–82. [CrossRef]

Graham, Suzanne, and Ernesto Macaro. 2008. Strategy instruction in listening for lower-intermediate learners of French. *Language Learning* 58: 747–83. [CrossRef]

Graham, Suzanne, Denise Santos, and Robert Vanderplank. 2008. Listening comprehension and strategy use: A longitudinal exploration. *System* 36: 52–68. [CrossRef]

Graham, Suzanne, Denise Santos, and Ellie Francis-Brophy. 2014. Teacher beliefs about listening in a foreign language. *Teaching and Teacher Education* 40: 44–60. [CrossRef]

Leary, Samuel Ferebee, Jr. 1999. The Effect of Thinking Maps Instruction on the Achievement of Fourth-grade Student. Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

Leopold, Claudia, and Detlev Leutner. 2015. Improving students' science text comprehension through metacognitive self-regulation when applying learning strategies. *Metacognition and Learning* 10: 313–46. [CrossRef]

Liu, Hsueh-Jui. 2008. A study of the interrelationship between listening strategy use, listening proficiency levels, and learning style. *Annual Review of Education, Communication & Language Sciences* 5: 84–104.

Luo, Xiaorong, and Jian Gao. 2012. On the existing status in listening teaching and some suggestions for it. *Theory and Practice in Language Studies* 6: 1270. [CrossRef]

Manning, Cynthia. 2003. Improving Reading Comprehension through Visual Tools. Master's thesis, Eastern Nazarene College, Quincy, MA, USA.

Movahed, Roya. 2014. The effect of metacognitive strategy instruction on listening performance, metaconitive awareness and listening anxiety of beginner Iranian EFL students. *International Journal of English Linguistics* 4: 88–99. [CrossRef]

Nosratinia, Mania, Samira Ghavidel, and Alireza Zaker. 2015. Teaching metacognitive strategies through Anderson's model: Does it affect EFL learners' listening comprehension? *Theory and Practice in Language Studies* 6: 1233–43. [CrossRef]

O'Bryan, Anne, and Volker Hegelheimer. 2009. Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée* 12: 9–38.

Oxford, Rebecca L. 1990. *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House.

Oxford, Rebecca L. 1993. Research update on teaching L2 listening. *System* 21: 205–11. [CrossRef]

Prestwich, Dorothy L. 2008. Effects of Linguistic or Non-linguistic Cognitive Maps on Fourth Grade Students' Reading Comprehension. Ph.D. dissertation, University of Mississippi, Mississippi, MS, USA.

Prince, Peter. 2013. Listening, remembering, writing: Exploring the dictogloss task. *Language Teaching Research* 17: 486–500. [CrossRef]

Rasouli, Maliheh, Kambiz Mollakhan, and Alireza Karbalaei. 2013. The effect of metacognitive listening strategy training on listening comprehension in Iranian EFL context. *European Online Journal of Natural and Social Sciences* 2: 115–28.

Richards, Jack C., and Theodore S. Rodgers. 2001. *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Rubin, Joan. 1994. A review of second language listening comprehension research. *The Modem Language Journal* 2: 199–221. [CrossRef]

Schneider, Wolfgang. 2008. The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education* 2: 114–21. [CrossRef]

Schraw, Gregory. 1998. Promoting general metacognitive awareness. *Instructional Science* 26: 113–25. [CrossRef]

Smidt, Esther, and Volker Hegelheimer. 2004. Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning* 17: 517–56. [CrossRef]

Sterling, Ra Shaunda Vernee. 2011. The Effect of Metacognitive Strategy Instruction on Student Achievement in a Hybrid Developmental English Course. Ph.D. dissertation, University of South Alabama, Mobile, AL, USA.

Vandergrift, Larry. 1997. The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals* 30: 387–409. [CrossRef]

Vandergrift, Larry. 1999. Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal* 53: 168–76. [CrossRef]

Vandergrift, Larry. 2002. 'It was nice to see that our predictions were right': Developing metacognition in L2 listening comprehension. *Canadian Modern Language Review* 58: 555–75. [CrossRef]

Vandergrift, Larry. 2003. Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning* 53: 463–96. [CrossRef]

Vandergrift, Larry. 2004. Learning to listen or listening to learn? *Annual Review of Applied Linguistics* 24: 3–25. [CrossRef]

Vandergrift, Larry. 2007. Recent developments in second and foreign language listening comprehension research. *Language Teaching* 40: 191–210. [CrossRef]

Vandergrift, Larry, Christine C. M. Goh, Catherine J. Mareschal, and Marzieh H. Tafaghodtari. 2006. The metacognitive awareness listening questionnaire: Development and validation. *Language Learning* 56: 431–62. [CrossRef]

Veenman, Marcel V. J., Bernadette HAM Van Hout-Wolters, and Peter Afflerbach. 2006. Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning* 1: 3–14. [CrossRef]

Wang, Min-Tzu. 2009. Effects of Metacognitive Reading Strategy Instruction on EFL High School Students' Reading Comprehension, Reading Strategies Awareness, and Reading Motivation. Ph.D. dissertation, University of Florida, Gainesville, FL, USA.

Wenden, Anita L. 1998. Metacognitive knowledge and language learning. *Applied Linguistics* 19: 515–37. [CrossRef]